CSCD: a database for cancer-specific circular RNAs

Siyu Xia[1,2,3,†], Jing Feng[4,†], Ke Chen[5,†], Yanbing Ma[1], Jing Gong[6], Fangfang Cai[1], Yuxuan Jin[1], Yang Gao[1], Linjian Xia[1], Hong Chang[1], Lei Wei[1], Leng Han[6,*] and Chunjiang He[1,2,3,*]

[1]School of Basic Medical Sciences, Wuhan University, Wuhan 430071, Hubei, China, [2]Hubei Province Key Laboratory of Allergy and Immunology, Wuhan 430071, Hubei, China, [3]Hubei Provincial Key Laboratory of Developmentally Originated Disease, Wuhan 430071, Hubei, China, [4]International School of Software, Wuhan University, Wuhan 430072, Hubei, China, [5]Department of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China and [6]Department of Biochemistry and Molecular Biology, McGovern Medical School at The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

ABSTRACT

Circular RNA (circRNA) is a large group of RNA family extensively existed in cells and tissues. High-throughput sequencing provides a way to view circRNAs across different samples, especially in various diseases. However, there is still no comprehensive database for exploring the cancer-specific circRNAs. We collected 228 total RNA or polyA(-) RNA-seq samples from both cancer and normal cell lines, and identified 272 152 cancer-specific circRNAs. A total of 950 962 circRNAs were identified in normal samples only, and 170 909 circRNAs were identified in both tumor and normal samples, which could be further used as non-tumor background. We constructed a cancer-specific circRNA database (CSCD, http://gb.whu.edu.cn/CSCD). To understand the functional effects of circRNAs, we predicted the microRNA response element sites and RNA binding protein sites for each circRNA. We further predicted potential open reading frames to highlight translatable circRNAs. To understand the association between the linear splicing and the back-splicing, we also predicted the splicing events in linear transcripts of each circRNA. As the first comprehensive cancer-specific circRNA database, we believe CSCD could significantly contribute to the research for the function and regulation of cancer-associated circRNAs.

INTRODUCTION

Circular RNA (circRNA) is largely discovered by high-throughput sequencing (1), including many tissue-specific and cell-specific circRNAs (2,3). A few circRNAs were functional characterized in human diseases and other biological processes (4). For example, circRNA CDR1as (anti-sense to the cerebellar degeneration-related protein 1 transcript) was reported as miR-7 sponge and can inhibit the function of miR-7 in various cancers (5,6). CircCCDC66 serves as miRNA sponge and regulates colon cancer growth and metastasis (7). CircMTO1 acts as a sponge of oncogenic miR-9 to promote p21 (cyclin-dependent kinase inhibitor 1) expression (8). CircHIPK3 regulates cell growth by sponging 9 miRNAs (9). CircRNAs also act as RNA binding protein sponges (10,11). For example, CircFoxo3 interacts with anti-senescent protein ID-1 and transcription factor E2F1 to increase cellular senescence (12). A fusion-circRNA derived from fusion genes exerts important functions in leukemogenesis by interacting with the fusion protein (13). RNA binding proteins are also involved in the back-splicing (14). For example, RNA binding protein FUS regulates circRNA biogenesis in mouse motor neurons (15).

Recent studies reported the functional impact of alternative splicing on biogenesis of circRNA (10). Modulation of splicing factor muscleblind (MBL) levels affects circMBL biosynthesis, revealing the competition between regular splicing and circularization (16). Inhibition of canonical spliceosome by mRNA splicing inhibitor reduces both the levels of circRNA and the parent linear transcript, indicating the role of mRNA spliceosome in circRNA biogenesis (17). Several circRNAs can be translated into proteins. For example, Circ-ZNF609 contains an open reading frame and is translated into a protein in murine and human myoblasts (18). CircMBL encodes a protein in fly head

Table 1. Prediction of cancer-specific and normal circRNAs

	Number of circRNAs	exonic	intronic	intergenic	mRNA	lncRNA	MRE	RBP	ORF
Cancer specific	272 152	119 887	105 398	31 575	213 882	11 403	14 921 788	15 719 824	564 047
Normal	950 962	505 705	310 872	80 882	789 166	27 411	52 417 822	66 182 210	2 287 210
Common	170 909	133 447	20 398	10 349	150 494	3351	9 100 345	22 025 003	610 840
Total	1 394 023	759 039	436 668	122 806	1 153 542	42 165	76 439 955	103 927 037	3 462 097

(19). However, the functional features of most circRNAs remained to be characterized.

CircRNAs were potentially utilized as diagnostic markers (6,20–22). Recent studies identified large amount of circRNAs and constructed several databases for circRNAs. For example, Circ2Traits is compiled to link circRNAs and human diseases, including diabetes and asthma (23). CircBase integrates several circRNAs datasets into a standardized database, which allows users to explore circRNAs or download customized python scripts to identify circRNAs from their own RNA-seq data (24). CircNet provides circRNA expression profiles across hundreds of samples and illustrates circRNA-miRNA-gene regulatory networks (25). circRNADb provides the protein-coding annotations for human exonic circRNAs (26). CircInteractome explores the binding of microRNA and RNA binding proteins in circRNAs from circBase (27). However, there was no database focusing on cancer-specific circRNAs. Therefore, we collected RNA sequencing data from 87 cancer cell line samples across 19 cancers types and 141 normal cell samples from ENCODE, and constructed a cancer-specific circRNA database (CSCD, http://gb.whu.edu.cn/CSCD).

DATA COLLECTION AND DATABASE CONTENT

Cell line samples in CSCD

Previous studies showed that RNA-seq with libraries prepared by total RNA with rRNA depleted or polyA(-) enriched method (enriched for RNAs with no polyA tails (28–30)) are appropriate and efficient to characterize circRNAs (31). Therefore, we collected these RNA-seq samples from ENCODE (https://www.encodeproject.org/). In total, we collected 87 cancer cell line samples across 19 cancer types and 141 normal cell line samples (Supplementary Table S1).

Identification of cancer-specific circRNAs

To identify cancer-specific circRNAs (CS-circRNAs), four popular algorithms with high performance (32): CIRI2 (31,33), find_circ (5), circRNA_finder (34) and Circexplorer (35,36) were utilized for detecting the back-splice junction sites of circRNAs. Identification of potential splicing exons of circRNA was performed by CIRI2 (31). We included all circRNAs identified by either one of four algorithms with at least one back-splice junction read, so that the users could select circRNAs for the following experiments by their own criteria (e.g. number of junction reads). Genome assembly GRCh37 and GENCODE (version 19) gene annotation were used. We also converted circRNAs coordinates from GRCh37 to GRCh38 for users to browse. We identified 443 061 circRNAs from cancer cell lines. We then compared these circRNAs with 1 121 871 circRNAs identified from normal cell lines, and defined 272 152 CS-circRNAs (Supplementary Figure S1A). Those circRNAs identified in normal samples could be further used as non-tumor background. Interestingly, we identified more circRNAs in normal samples than tumor samples even after adjusting the samples size. This is likely due to the larger number of mapped reads in normal samples (Supplementary Figure S1B). Among these CS-circRNAs, 119 887, 105 398 and 31 575 of CS-circRNAs are located in exonic, intronic and intergenic regions, respectively. We also identified 213 882 and 11 403 of CS-circRNAs located in mRNA and lncRNA, respectively (Table 1). We identified many circRNAs which were not identified by other databases. For example, we identified 17 circRNAs for CDR1-AS, which only one was observed in circBase. This is largely due to the reason that CSCD included different samples, especially cancer samples, further suggesting the importance of a cancer-specific database.

Prediction of cellular localization

Previous work reported that most circRNAs derived from exon were identified in cytosolic (37), while those circRNAs consisting of intron or exon–intron were mainly identified in nucleus (38). To comprehensively view the cellular localization of CS-circRNAs, we extracted the cellular localization data if applicable. There are 19 228, 2107, 7020, 35 734, 37 453, 37 141 and 16 976 CS-circRNAs localized in cytosolic, insoluble cytoplasmic, membrane, chromatin, nuclear, nucleoplasmic and nucleolus, respectively.

Prediction of MRE, RBP and ORF

CircRNAs were reported to act as microRNA sponge and regulating gene expression through microRNA response elements (MREs) (39). To understand the potential regulatory functions of CS-circRNAs, 100 bp window (± 50 bp) (27) surrounding the back-splicing site of each CS-circRNAs was selected to scan the potential MREs by TargetScan (40). By scanning the junction region for miRNA seeds (7mer-m8, 7mer-1a and 8mer), we identified 14 921 788 MREs in CS-circRNAs, 52 417 822 MREs in normal circRNAs and 9 100 345 MREs in common circRNAs.

Another potential function of circRNAs is that circRNA serves as sponges for RNA-binding proteins (RBPs) (41). Utilizing CLIP-Seq data for protein binding sites of 37 RBPs from STARBASE (42), we identified potential RBP binding events in circRNAs. We identified 15 719 824 RBPs in CS-circRNAs, 66 182 210 in normal circRNAs and 22 025 003 in common circRNAs. Comparing with the previous database, CircInteractome, CSCD collected more cir-
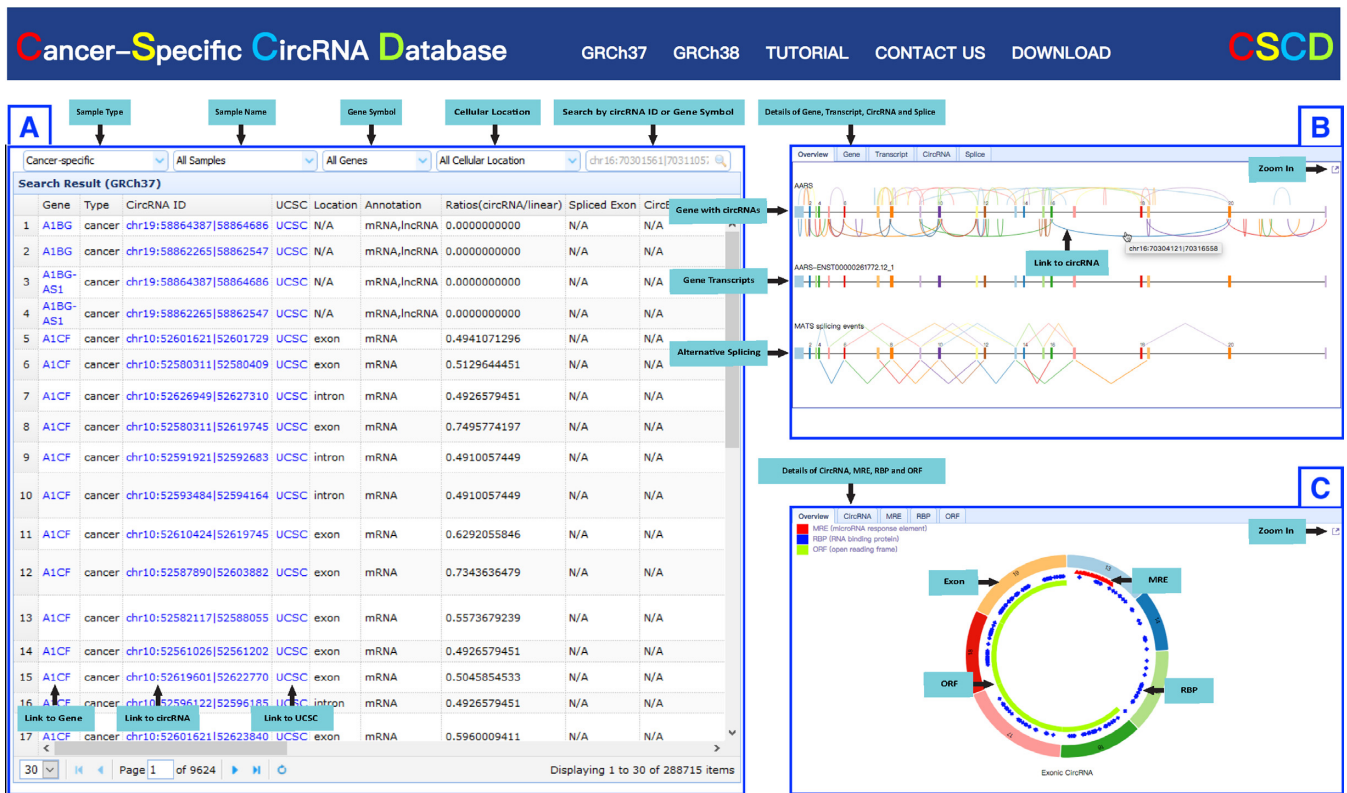
Figure 1. Overview of CSCD. (**A**) Query panel of the circRNA. CircRNA can be viewed and searched by sample name, gene symbol and circRNA ID, etc. (**B**) Gene panel. Image and information of gene, transcript, alternative splicing and circRNAs are displayed on this panel. (**C**) CircRNA panel. Image and information of circRNA and related location of MRE, RBP and ORF are displayed on this panel.

cRNAs, MREs and RBPs (Supplementary Table S2), as we included more samples, especially many cancer samples.

Recent studies showed the protein-encoding ability of circRNAs, which were considered as non-coding RNA before (18,19). To examine the translational potential of circRNAs, the open reading frames (ORFs) were predicted using full-length sequence of circRNA by ORF Finder. The minimal length of ORF length was set as 300nt according to a previous work (26). We identified a total of 564 047, 2 287 210 and 610 840 ORFs in cancer-specific, normal and common circRNAs, respectively.

Detection of alternative splicing events of parent genes

Alternative splicing (AS), the most frequent events in transcriptional process, may affect the biogenesis of circRNAs (31,36). To understand the relationship between alternative splicing in linear gene and biogenesis of circRNAs, we predicted alternative splicing events across all samples. RNA-seq read alignment was performed by STAR (43). Potential alternative splicing events of a linear gene, including skipped exons, alternative 5' splice site, alternative 3' splice site, mutually exclusive exons and retained introns, were detected by rMATS (44) with default parameters.

DATABASE ORGANIZATION AND WEB INTERFACE

All the data, including gene annotation, circRNAs, MRE, RBP, ORF, AS associated with circRNAs were organized

into a set of interactive MySQL tables. ThinkPHP, an open-source web framework based on PHP (https://github.com/top-think) and JavaScript library were used to construct the CSCD database. To make the data query convenient and efficient, we organized our database into three sub-databases by data type (cancer-specific, normal and common). The web interface of CSCD is summarized in Figure 1. The main page of CSCD is composed of three panels.

Query Panel to search/browse circRNAs

In this panel, users can browse circRNAs by selecting sample type, sample name, gene symbol, cellular localization and search circRNA ID (e.g. chrX:18928998|18938303, which represents the donor and acceptor site of each circRNA) or gene symbol in searching box. All information, including the parent gene, sample type, circRNA ID, UCSC genome browser link (45), sample source, genomic coordinates, lncRNA/mRNA annotation, ratio of circRNA/linear RNA, spliced exons, circBase ID, cellular localization, identification algorithm, number of junction reads and log2 SRPTM (number of circular reads/number of mapped reads (units in trillion)/read length) (2) for each circRNA are displayed in the table (Figure 1A). Gene symbol links to the Gene Panel with all circRNAs across different samples (Figure 1B). CircRNA ID links to the circRNA Panel with a circRNA in a specific sample (Figure 1C).

Gene panel to view all circRNAs for selected gene

In this panel, users can view the circRNAs and their linear parent gene in Overview tab. Linear gene structures are displayed with different colored rectangles for exons and black lines for introns, while circRNAs are shown as colorful curves. All transcripts and potential alternative splicing events of this linear parent gene are also displayed below. Users can zoom in for a high-resolution image by clicking the top right corner of the panel. All the detailed information is listed in gene tab, transcript tab, circRNA tab and splicing tab (Figure 1B). The circRNA curve links to specific circRNA in circRNA panel (Figure 1C).

CircRNA panel to view specific circRNA

In this panel, users can view selected circRNAs with consisting exons in a colored circle (Figure 1C). Each arc with number ID depicts one exon, while introns are displayed in black lines. Users can also view the number and position of MRE (red triangle), RBP (blue rectangle) and ORF (green arc) elements located in circRNA and check the detailed information through the circRNA, MRE, RBP and ORF tabs, respectively. Users can zoom in for a high-resolution image by clicking the top right corner of the panel.

SUMMARY AND FUTURE DIRECTIONS

CSCD collects available RNASeq datasets in cancer and normal samples, and provides an integrated circRNA database to benefit functional studies of cancer-specific circRNAs. CSCD also collects normal and common circRNAs, which allow users for other studies. For example, users can compare circRNAs in their cancer samples with CSCD to examine whether those circRNAs are cancer-specific or not. Users can view the potential functional regulation and translation on these circRNAs through prediction of MRE, RNA binding protein and open reading frame. Users can also link the back-splicing and alternative splicing of linear genes through splicing events prediction. Due to the limited RNA-seq datasets for total RNA with rRNA depleted or polyA (-) in primary tumor samples, we did not include any primary tumor samples. We will update CSCD when more sequencing data from primary samples are available.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Jeck,W.R. and Sharpless,N.E. (2014) Detecting and characterizing circular RNAs. *Nat. Biotechnol.*, **32**, 453–461.
2. Xia,S., Feng,J., Lei,L., Hu,J., Xia,L., Wang,J., Xiang,Y., Liu,L., Zhong,S., Han,L. *et al.* (2016) Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief. Bioinformatics.*, bbw081.
3. Salzman,J., Chen,R.E., Olsen,M.N., Wang,P.L. and Brown,P.O. (2013) Cell-type specific features of circular RNA expression. *PLoS Genet.*, **9**, e1003777.
4. Zhao,Z.J. and Shen,J. (2015) Circular RNA participates in the carcinogenesis and the malignant behavior of cancer. *RNA Biol.*, **14**, 514–521.
5. Memczak,S., Jens,M., Elefsinioti,A., Torti,F., Krueger,J., Rybak,A., Maier,L., Mackowiak,S.D., Gregersen,L.H., Munschauer,M. *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333–338.
6. Weng,W., Wei,Q., Toden,S., Yoshida,K., Nagasaka,T., Fujiwara,T., Cai,S., Qin,H., Ma,Y. and Goel,A. (2017) Circular RNA ciRS-7-A promising prognostic biomarker and a potential therapeutic target in colorectal cancer. *Clin. Cancer Res.*, **23**, 3918–3928.
7. Hsiao,K.Y., Lin,Y.C., Gupta,S.K., Chang,N., Yen,L., Sun,H.S. and Tsai,S.J. (2017) Noncoding effects of circular RNA CCDC66 promote colon cancer growth and metastasis. *Cancer Res.*, **77**, 2339–2350.
8. Han,D., Li,J., Wang,H., Su,X., Hou,J., Gu,Y., Qian,C., Lin,Y., Liu,X., Huang,M. *et al.* (2017) Circular RNA MTO1 acts as the sponge of miR-9 to suppress hepatocellular carcinoma progression. *Hepatology*, **66**, 1151–1164.
9. Zheng,Q., Bao,C., Guo,W., Li,S., Chen,J., Chen,B., Luo,Y., Lyu,D., Li,Y., Shi,G. *et al.* (2016) Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.*, **7**, 11215.
10. Qu,S., Yang,X., Li,X., Wang,J., Gao,Y., Shang,R., Sun,W., Dou,K. and Li,H. (2015) Circular RNA: a new star of noncoding RNAs. *Cancer Lett.*, **365**, 141–148.
11. Ebbesen,K.K., Kjems,J. and Hansen,T.B. (2016) Circular RNAs: Identification, biogenesis and function. *Biochim. Biophys. Acta*, **1859**, 163–168.
12. Du,W.W., Yang,W., Chen,Y., Wu,Z.K., Foster,F.S., Yang,Z., Li,X. and Yang,B.B. (2017) Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur. Heart J.*, **38**, 1402–1412.
13. Guarnerio,J., Bezzi,M., Jeong,J.C., Paffenholz,S.V., Berry,K., Naldini,M.M., Lo-Coco,F., Tay,Y., Beck,A.H. and Pandolfi,P.P. (2016) Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. *Cell*, **165**, 289–302.
14. Conn,S.J., Pillman,K.A., Toubia,J., Conn,V.M., Salmanidis,M., Phillips,C.A., Roslan,S., Schreiber,A.W., Gregory,P.A. and Goodall,G.J. (2015) The RNA binding protein quaking regulates formation of circRNAs. *Cell*, **160**, 1125–1134.
15. Errichelli,L., Dini Modigliani,S., Laneve,P., Colantoni,A., Legnini,I., Capauto,D., Rosa,A., De Santis,R., Scarfo,R., Peruzzi,G. *et al.* (2017) FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.*, **8**, 14741.
16. Ashwal-Fluss,R., Meyer,M., Pamudurti,N.R., Ivanov,A., Bartok,O., Hanan,M., Evantal,N., Memczak,S., Rajewsky,N. and Kadener,S. (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell*, **56**, 55–66.
17. Starke,S., Jost,I., Rossbach,O., Schneider,T., Schreiner,S., Hung,L.H. and Bindereif,A. (2015) Exon circularization requires canonical splice signals. *Cell Rep.*, **10**, 103–111.

18. Legnini,I., Di Timoteo,G., Rossi,F., Morlando,M., Briganti,F., Sthandier,O., Fatica,A., Santini,T., Andronache,A., Wade,M. *et al.* (2017) Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell*, **66**, 22–37.

19. Pamudurti,N.R., Bartok,O., Jens,M., Ashwal-Fluss,R., Stottmeister,C., Ruhe,L., Hanan,M., Wyler,E., Perez-Hernandez,D., Ramberger,E. *et al.* (2017) Translation of CircRNAs. *Mol. Cell*, **66**, 9–21.

20. Meng,S., Zhou,H., Feng,Z., Xu,Z., Tang,Y., Li,P. and Wu,M. (2017) CircRNA: functions and properties of a novel potential biomarker for cancer. *Mol. Cancer*, **16**, 94.

21. Zhang,H.D., Jiang,L.H., Sun,D.W., Hou,J.C. and Ji,Z.L. (2017) CircRNA: a novel type of biomarker for cancer. *Breast Cancer*, doi:10.1007/s12282-017-0793-9.

22. Xuan,L., Qu,L., Zhou,H., Wang,P., Yu,H., Wu,T., Wang,X., Li,Q., Tian,L., Liu,M. *et al.* (2016) Circular RNA: a novel biomarker for progressive laryngeal cancer. *Am. J. Transl. Res.*, **8**, 932–939.

23. Ghosal,S., Das,S., Sen,R., Basak,P. and Chakrabarti,J. (2013) Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front. Genet.*, **4**, 283.

24. Glazar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.

25. Liu,Y.C., Li,J.R., Sun,C.H., Andrews,E., Chao,R.F., Lin,F.M., Weng,S.L., Hsu,S.D., Huang,C.C., Cheng,C. *et al.* (2016) CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res.*, **44**, D209–D215.

26. Chen,X., Han,P., Zhou,T., Guo,X., Song,X. and Li,Y. (2016) circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.*, **6**, 34985.

27. Dudekula,D.B., Panda,A.C., Grammatikakis,I., De,S., Abdelmohsen,K. and Gorospe,M. (2016) CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.*, **13**, 34–42.

28. Wu,Q., Kim,Y.C., Lu,J., Xuan,Z., Chen,J., Zheng,Y., Zhou,T., Zhang,M.Q., Wu,C.I. and Wang,S.M. (2008) Poly A- transcripts expressed in HeLa cells. *PLoS One*, **3**, e2803.

29. Lai,F., Gardini,A., Zhang,A. and Shiekhattar,R. (2015) Integrator mediates the biogenesis of enhancer RNAs. *Nature*, **525**, 399–403.

30. Yang,L., Duff,M.O., Graveley,B.R., Carmichael,G.G. and Chen,L.L. (2011) Genomewide characterization of non-polyadenylated RNAs. *Genome Biol.*, **12**, R16.

31. Gao,Y., Wang,J., Zheng,Y., Zhang,J., Chen,S. and Zhao,F. (2016) Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.*, **7**, 12060.

32. Hansen,T.B., Veno,M.T., Damgaard,C.K. and Kjems,J. (2016) Comparison of circular RNA prediction tools. *Nucleic Acids Res.*, **44**, e58.

33. Gao,Y., Wang,J. and Zhao,F. (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.*, **16**, 4.

34. Westholm,J.O., Miura,P., Olson,S., Shenker,S., Joseph,B., Sanfilippo,P., Celniker,S.E., Graveley,B.R. and Lai,E.C. (2014) Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.*, **9**, 1966–1980.

35. Zhang,X.O., Wang,H.B., Zhang,Y., Lu,X., Chen,L.L. and Yang,L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134–147.

36. Zhang,X.O., Dong,R., Zhang,Y., Zhang,J.L., Luo,Z., Zhang,J., Chen,L.L. and Yang,L. (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.*, **26**, 1277–1287.

37. Salzman,J., Gawad,C., Wang,P.L., Lacayo,N. and Brown,P.O. (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One*, **7**, e30733.

38. Li,Z., Huang,C., Bao,C., Chen,L., Lin,M., Wang,X., Zhong,G., Yu,B., Hu,W., Dai,L. *et al.* (2015) Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.*, **22**, 256–264.

39. Hansen,T.B., Jensen,T.I., Clausen,B.H., Bramsen,J.B., Finsen,B., Damgaard,C.K. and Kjems,J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384–388.

40. Agarwal,V., Bell,G.W., Nam,J.W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.

41. Du,W.W., Yang,W., Liu,E., Yang,Z., Dhaliwal,P. and Yang,B.B. (2016) Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.*, **44**, 2846–2858.

42. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (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.

43. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

44. Shen,S., Park,J.W., Lu,Z.X., Lin,L., Henry,M.D., Wu,Y.N., Zhou,Q. and Xing,Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5593–E5601.

45. Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W., Thomas,D.J. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.