



DATABASE

CIRCpedia v2: An Updated Database for Comprehensive Circular RNA Annotation and Expression Comparison



Rui Dong^{1,#,a}, Xu-Kai Ma^{1,#,b}, Guo-Wei Li^{1,c}, Li Yang^{1,2,*,d}

¹ CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

² School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

Received 19 May 2018; revised 17 July 2018; accepted 10 August 2018

Available online 29 August 2018

Handled by Fangqing Zhao

KEYWORDS

Circular RNA;
CircRNA;
Back-splicing;
Database;
CIRCpedia

Abstract Circular RNAs (circRNAs) from back-splicing of exon(s) have been recently identified to be broadly expressed in eukaryotes, in tissue- and species-specific manners. Although functions of most circRNAs remain elusive, some circRNAs are shown to be functional in gene expression regulation and potentially relate to diseases. Due to their stability, circRNAs can also be used as biomarkers for diagnosis. Profiling circRNAs by integrating their expression among different samples thus provides molecular basis for further functional study of circRNAs and their potential application in clinic. Here, we report **CIRCpedia v2**, an updated **database** for comprehensive circRNA annotation from over 180 RNA-seq datasets across six different species. This atlas allows users to search, browse, and download circRNAs with expression features in various cell types/tissues, including disease samples. In addition, the updated database incorporates conservation analysis of circRNAs between humans and mice. Finally, the web interface also contains computational tools to compare circRNA expression among samples. CIRCpedia v2 is accessible at <http://www.picb.ac.cn/rnomics/circpedia>.

Introduction

Eukaryotic mRNA precursors (pre-mRNAs) usually undergo splicing to remove introns and join exons by ligating an upstream 5' splice site (5' ss) with a downstream 3' ss to produce linear RNAs with 5' to 3' polarity. Splicing can also happen in a reversed order to ligate a downstream 5' ss with an upstream 3' ss, referred to as back-splicing. Back-splicing of

* Corresponding author.

E-mail: liyong@picb.ac.cn (Yang L).

Equal contribution.

^a ORCID: 0000-0003-0985-6211.

^b ORCID: 0000-0003-3779-4070.

^c ORCID: 0000-0002-2972-3398.

^d ORCID: 0000-0001-8833-7473.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2018.08.001>

1672-0229 © 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

exon(s) leads to the formation of circular RNAs (circRNAs), featured by covalently close circle structure in nature [1–5]. Although discovered more than 25 years ago in higher eukaryotes, only a few circRNAs were found at that time, which were thought as byproducts from mis-splicing with little functional potential [6–10]. However, recent genome-wide profiling of nonpolyadenylated RNA transcriptomes with specific computational approaches that aim to identify reads mapped to back-splicing junctions, has identified a large number of circRNAs in various cell lines/tissues and across different species [11–19].

Recent research progresses have shown that back-splicing is catalyzed by canonical spliceosomal machinery [20–22], but in a relatively low efficiency [23], possibly due to the unfavorable spliceosome assembly at back-splicing junction sites. The formation of circRNAs can be regulated by both *cis*-elements and *trans*-factors. On the one hand, orientation-opposite complementary sequences, when located pairwise in introns flanking circRNA-producing exons, can facilitate back-splicing for circRNA formation. On the other hand, RNA binding proteins can associate with intronic *cis*-elements/sequences to further regulate circRNA biogenesis [16,24–27]. Recent studies have shown that some circRNAs get involved in gene expression networks through a variety of molecular mechanisms, such as competing with canonical splicing, sequestering microRNAs/proteins or interfering with RNA polymerase II-mediated transcription [13,24,27–30]. These new findings thus suggest that circRNAs, at least some of them, can act as regulatory RNA molecules. Therefore, comprehensive annotation of circRNAs will likely provide molecular basis for understanding the function of circRNAs and their potential applications in clinic.

Although generally expressed at low levels, circRNAs are highly enriched in brains/neurons [18,23,31]. This phenomenon can be partially explained by the resistance of circRNAs to exonucleolytic degradation, which allows these RNA circles to gradually accumulate to relatively high levels in cells with slow division rates, such as neurons [23]. Interestingly, significantly more circRNAs are detected in humans than in mice, fruitflies, and worms [32]. As produced from annotated protein-coding genes, sequence conservation of circRNA-producing exons is similar to that of their neighboring linear RNA-producing exons across humans and mice [18,33]. However, only a small portion (10%–20%) of human circRNAs could be observed in mouse samples [18,32,33], which is likely due to the different composition of orientation-opposite complementary sequences across species [32]. It has been shown that short interspersed nuclear elements (*SINEs*) and primate-specific *Alus*, when juxtaposed in introns flanking circRNA-forming exons, contribute the most to the circRNA formation. The complementarity of *Alu* elements in humans is much higher than those of *SINEs* or other complementary sequences in mice, fruitflies, and worms, which may lead to the observed species-specific circRNA expression [18,32].

Despite the diversified circRNA expression across cell lines/tissues and species, multiple circRNAs can be produced by different selections of alternative back-splicing and/or canonical alternative splicing within circRNAs [34,35]. Competition of putative RNA pairs formed by orientation-opposite complementary sequences across different sets of introns that bracket alternative back-splice sites can lead to alternative back-splicing selection [35], which is largely

diverse among the cell lines/tissues examined. In this case, identification of different alternative back-splicing events serves as a basis to study their cell type-specific and tissue-specific functions.

Here, we report CIRCpedia v2, an updated database for comprehensive circRNA annotation from over 180 RNA-seq datasets across six different species. This atlas allows users to search, browse, and download circRNAs and alternative back-splicing events with expression characteristics/features in various cell types/tissues, including disease samples. The updated database also provides conservation information of circRNAs between humans and mice after LiftOver analysis. Finally, the web interface also contains computational tools for simple comparison of circRNA expression between samples. CIRCpedia v2 is accessible at <http://www.picb.ac.cn/rnomics/circpedia>.

Implementation

CIRCpedia v2 is implemented with Flask (a lightweight framework designed to build web applications; <http://flask.pocoo.org/>), Bootstrap (an open source front-end framework for developing web pages; <https://getbootstrap.com/>), jQuery (a library designed to simplify the client-side javascript programming; <https://jquery.com/>), Bokeh (a python library for interactive data visualization; <https://bokeh.pydata.org/en/latest/>), and JBrowse (a genome browser for visualizing track files; <https://jbrowse.org/>). SQLite (a structured query language database engine; <https://www.sqlite.org/index.html>) is used to store metadata information. Codes are developed by using Visual Studio Code (a flexible and customizable source code editor that supports a series of programming languages; <https://code.visualstudio.com/>). Finally, CIRCpedia v2 is hosted on a Red Hat Enterprise Linux system with Apache HTTP Server to provide a stable web access.

Database content and usage

Summary of new features

CIRCpedia v2 significantly expands the data contents by collecting various RNA-seq datasets from the Gene Expression Omnibus (GEO), Encyclopedia of DNA Elements (ENCODE) project, and the EMBL-European Bioinformatics Institute (EMBL-EBI) database. Compared to the previous version [35], CIRCpedia v2 includes a comprehensive annotation of circRNAs and alternative back-splicing events by analyzing additional cell lines/tissues and species (Figure 1A). We have applied fragments mapped to back-spliced junctions per million mapped fragments (FPM) value to quantitate circRNA expression, which is suitable for both single-read and paired-end sequencing datasets. CIRCpedia v2 incorporates conservation information of circRNAs between humans and mice by LiftOver analysis [32]. In addition, a new version of JBrowse is embedded for circRNA visualization. Finally, a new web interface for direct circRNA expression comparison among different samples is incorporated into CIRCpedia v2, labeled as “Tool”. In summary, CIRCpedia v2 is a comprehensive, user-friendly, and powerful database with circRNA search, visualization, downloading, and online analysis.

A

Welcome to CIRCpedia v2

CIRCpedia v2 Search Browse Download Tool About CIRCpedia v1

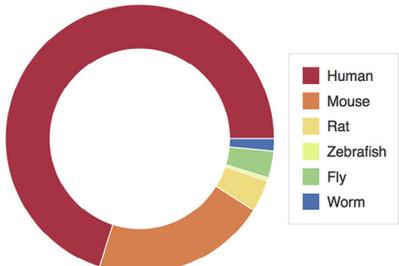
Welcome to CIRCpedia v2

CIRCpedia v2 is an updated comprehensive database containing circRNA annotations from over 180 RNA-seq datasets across six different species. This atlas allows users to search, browse and download circRNAs with expression characteristics/features in various cell types/tissues, including disease samples. In addition, the updated database incorporates conservation analysis of circRNAs between humans and mice. Finally, the web interface also contains computational tools for the comparison of circRNA expression between samples. The updated database is accessible at <http://www.picb.ac.cn/rnomics/circpedia>.

Statistics

Summary of circRNA numbers

No. of total circRNAs: 262,782



Species	No. of circRNAs
Human	183,943
Mouse	55,312
Rat	10,197
Zebrafish	911
Fly	8,560
Worm	3,859

News

07/07/18 circRNA identification based on human reference genome hg19 was added into the CIRCpedia.

06/29/18 Another 17 human samples and 27 mouse samples have been added into CIRCpedia.

05/18/18 Another 22 human samples and 12 fly samples have been added into CIRCpedia. In addition, 2 zebrafish samples and 6 rat samples have been added into CIRCpedia.

05/17/18 FPM was used to represent circRNA

B

Species	No. of samples			No. of circRNAs
	Ribo ⁻	poly(A) ⁻	RNase R	
Human (hg38)	54	13	3	183,943
Mouse (mm10)	51	1	1	55,312
Rat (rn6)	6	0	0	10,197
Zebrafish (danRer10)	1	0	1	911
Fly (dm6)	33	0	9	8560
Worm (ce10)	6	0	6	3859

Figure 1 Homepage of CIRCpedia v2

A. Summary of annotated circRNAs in CIRCpedia v2. B. Datasets used in CIRCpedia v2.

Data statistics of CIRCpedia v2

Current version of CIRCpedia v2 contains 185 datasets (86 new ones compared with those in CIRCpedia) from ribo⁻, poly(A)⁻, ribo⁻/RNase R, or poly(A)⁻/RNase R RNA-seq (Figure 1B). Among these 86 new datasets, 39 samples are from humans, 27 are from mice, and 12 are from fruitflies, respectively. In addition to humans, mice, fruitflies, and worms that have been analyzed in early version, datasets from two

additional species are also incorporated in CIRCpedia v2, including six ribo⁻ RNA-seq datasets from rats, as well as one ribo⁻ RNA-seq and one ribo⁻/RNase R RNA-seq datasets from zebrafish. After identifying RNA-seq reads that are mapped to back-splicing junction sites by CIRCexplorer2 [35] (<http://circexplorer2.readthedocs.io/en/latest/>), a circRNA annotation track file (BED), an alternative 5' back-splicing (A5BS) annotation track file (BED) and an alternative 3' back-splicing (A3BS) annotation track file (BED) are obtained

for each RNA-seq dataset. In total, about 262,782 back-splicing events, which represent at least 262,782 circRNAs, are annotated in CIRCpedia v2 in the samples examined (Figure 1). About 39,225 A5BS and 34,747 A3BS events are also included in CIRCpedia v2. Expression levels of all back-splicing events (including A5BS and A3BS) are indicated using FPMs.

Web interface for search

CIRCpedia v2 provides a “Search” function by querying genomic locations, gene symbols, or CIRCpedia IDs (circIDs) (Figure 2A). Search using a specific gene symbol (genomic location) can retrieve all circRNA and related alternative back-splicing events in this locus (/genomic location). Output table files include circID, species, gene, isoform, location, strand, FPM, ExonStart-ExonEnd, seq type, cell line, conservation between humans and mice, information on annotation by MapSplice alignment, and enrichment fold change after RNase R treatment. Here, the conservation information is obtained by using LiftOver tool to identify conserved circRNAs between humans and mice (Methods). Notably, other than CIRCexplorer2 [16,35], another computational pipeline, MapSplice [36], is also used for human circRNA annotation. In addition, when applicable, enrichment fold change is also listed in ribo⁻ and poly(A)⁻ samples with corresponding RNase R treatment. Meanwhile, output files for A5BS and

A3BS by the same queries are available, listed with the information of species, gene, isoform, location, strand, percent circularized-site usage (PCU), sequencing type, and cell line. Users can limit their queries to specific species or cell line/tissue using different setting options (Figure 2A). The output table files can be exported in different formats, such as JSON, XML, CSV, or TXT. Useful links are also available for circRNA visualization (see below) and host gene identification at GeneCards websites (for humans only).

Web interfaces for browse and downloading

A “Browse” function is also available in CIRCpedia v2. After “Search”, the identified circRNAs can be directly linked to the website-embedded JBrowse for visualization (Figure 2B). Alternatively, users can visualize circRNAs by selecting different options on the page of “Browse”. Different types of tracks, including gene annotation, RNA-seq dataset, circRNA annotation, and alternative back-splicing event, are available in the left panel of JBrowse, with the selected tracks shown in the right panel. Users can jump between different species by clicking “Genome” button at the top left corner of JBrowse. Furthermore, by clicking the track of a given circRNA in JBrowse, its expression value (FPM) pops up for visualization.

With the “Download” function, tables of circRNAs and alternative back-splicing events from the selected cell lines/tissues and species can be retrieved from CIRCpedia v2

A

Search

Species: Human (hg38) Cell Line: PA1

Gene/Genomic Location (Chr:Start-End): SMARCA5 (e.g. CAMSAP1 or Chr9:1350)

(Optional) List of circIDs:

Search

circRNA A5BS A3BS

circID	Species	Gene	Isoform	Location	Strand	FPM	ExonStart-ExonEnd	Seq Type	Cell Line	Conservation	Mapsplice	Enrichment
HSA_CIRCpedia_47541	Human (hg38)	SMARCA5	ENST00000283131.3	Chr4:143543508-143543972	+	2.717	15-16	RNase R	PA1	Species-specific	Yes	N.A.
HSA_CIRCpedia_47538	Human (hg38)	SMARCA5	ENST00000283131.3	Chr4:143527867-143530526	+	0.076	7-9	RNase R	PA1	Species-specific	Yes	N.A.
HSA_CIRCpedia_47541	Human (hg38)	SMARCA5	ENST00000283131.3	Chr4:143543508-143543972	+	0.57	15-16	poly(A) ⁻	PA1	Species-specific	Yes	4.757

B

Browse

Available Tracks

- filter tracks
- knownGene
- refGene
- A5BS
- A3BS
- circRNA
- Reference sequence
- Reference sequence
- Wiggle Track
- A5R_poly(A)
- A5M50_poly(A)
- S1_poly(A)
- Cerebellum_Ribo
- Cortex_Ribo
- Pituitary_Ribo
- Forebrain_Ribo
- Gli378_poly(A)
- H1_poly(A)
- HP_Ribo
- HP_Ribo
- HP_poly(A)
- HP_poly(A)

Genome Track View Help

50,000,000 100,000,000 150,000,000

chr4 | chr4:143543512..143543978 (466 bp) | Go

143,543,635 143,543,750

Reference sequence

Wiggle Track

knownGene

SMARCA5

PA1_RNaseR_circ

PA1_RNaseR

C

Download

Table

circRNA

Species

Human (hg38)

Cell Line

All

Download

Figure 2 Screenshots of web interface in CIRCpedia v2

A. Search circRNAs with query. The result table shows the annotation of circRNAs originating from human *SMARCA5* locus. B. Browse circRNAs by JBrowse visualization. C. Download circRNAs for further analysis.

(Figure 2C). The output files are in CSV format, listed with the information of circID, species, gene, isoform, location, strand, FPM, ExonStart-ExonEnd, sequencing type, cell line, conservation, information on annotation by MapSplice alignment, and enrichment fold change after RNase R treatment.

Web interface for circRNA expression analysis

A new “Tool” function is now available in the web interface of CIRCpedia v2. A list of circRNAs with their genomic locations can be directly copied and pasted for expression comparison analysis. Currently, CIRCpedia v2 enables two types of data visualization, heatmap and point plot (Figure 3). As shown in Figure 3A, a heatmap diagram is generated to show differential circRNA expression (FPM values) in selected cell lines. Alternatively, users can choose a different method, such as point plot, to illustrate differential circRNA expression among various samples (Figure 3B).

Discussion and perspectives

Without 3′ polyadenylated tails, circRNAs had been largely missed by the enrichment of polyadenylated RNAs for deep sequencing (referred to as poly(A)⁺ RNA-seq), but were successfully captured by two nonpolyadenylated RNA isolation strategies prior to deep sequencing (referred to as ribo⁻ or poly(A)⁻ RNA-seq, respectively) [16,35]. The ribo⁻ RNA-seq profiles both nonpolyadenylated and polyadenylated transcripts after depleting rRNAs, and the poly(A)⁻ RNA-seq examines relatively pure nonpolyadenylated transcripts by getting rid of both rRNAs and polyadenylated RNAs. With the advent of deep sequencing for nonpolyadenylated RNAs, a large number of circRNAs from back-splicing of exons have been widely identified in various cell lines/tissues and across distinct species. A comprehensive annotation of circRNAs is necessary for further circRNA research.

Here, we update CIRCpedia database to version 2 by providing comprehensive circRNA annotation and online tool for circRNA expression analysis. We integrate “Search”, “Browse”, “Download”, and “Tool” functions into the web interface. Users can search circRNAs by querying specific genomic regions or gene names. Meanwhile, tracks of circRNAs, alternative back-splicing events, and RNA-seq files can be visualized by embedded JBrowse directly in the web interface. With the “Download” function, information of circRNAs and alternative back-splicing events from selected cell lines/tissues and species can be retrieved for additional analysis. Finally, users with limited bioinformatics background can choose a new online “Tool” function for direct comparison of circRNA expression among different cell lines/tissues. Compared to other circRNA databases, such as circBase [37], CSCD [38], or circRNAdb [39], CIRCpedia v2 provides a more comprehensive annotation of circRNAs with some unique characteristics. For instance, it contains 185 RNA-seq datasets, covering a wide range of cell lines/tissues and species, and leads to the identification of over 260,000 circRNAs from the transcriptomes examined. In addition, conservation analysis of circRNAs expressed across humans and mice is also included in CIRCpedia v2. Finally, CIRCpedia v2 provides useful bioinformatic tools for direct online comparison of circRNA expression.

We will continue to improve and update this circRNA atlas for the research community. When additional RNA-seq datasets are available, they will be incorporated into this database. Given combination of different circRNA prediction algorithms in transcriptome-wide profiling can reduce false positives [40], we also use other pipelines, such as MapSplice [36], to provide additional information for circRNA prediction. Currently, all datasets have been screened using CIRCexplorer2, and all human RNA-seq datasets have also been analyzed using MapSplice. circRNAs predicted using at least two methods will be further highlighted in CIRCexplorer2. Other useful online tools, such as calculating pairing capacity of orientation-opposite RNA complementary sequences [32], will also be introduced to CIRCpedia v2.

Taken together, we report here CIRCpedia v2, an updated circRNA database with integrated information of circRNA annotation, expression, and conservation. The CIRCpedia v2 database is freely available at <http://www.picb.ac.cn/nomics/circpedia>.

Methods

Database design

CIRCpedia v2 contains circRNA tables, gene annotation tables, alternative back-splicing tables, and conservation tables. All the tables were stored in SQLite database. The pages for “Search” and “Download” are more convenient for cell line selection in specific species. JBrowse (version: 1.14.0) is embedded into CIRCpedia v2. In addition, a useful tool to compare circRNA expression is incorporated in the “Tool” page. Expression levels of circRNAs can be drawn by heatmap and point plot.

Annotation and alternative back-splicing analysis of circRNAs

In total, 185 RNA-seq datasets were retrieved from GEO (<https://www.ncbi.nlm.nih.gov/geo/>), ENCODE (<https://www.encodeproject.org/>) project, and EMBL-EBI (<https://www.ebi.ac.uk/>) database across six species for analysis in this study. CIRCexplorer2 [16,35] was used to retrieve RNA-seq reads that are mapped to back-splicing junction sites for circRNA prediction. Briefly, RNA-seq reads were mapped to reference genomes (human: GRCh38/hg38 and GRCh37/hg19; mouse: GRCm38/mm10; rat: RGSC 6.0/rn6; zebrafish: GRCz10/danRer10; fly: BDGP Release 6 + ISO1 MT/dm6; and worm: WS220/ce10) using HISAT2 (2.1.0, parameter: --no-softclip --score-min L, -16,0 --mp 7,7 --rfg 0,7 --rdg 0,7 --dta-cufflinks -k 1 --max-seeds 20). The unmapped reads were then further mapped to relevant genomes using TopHat-Fusion (version: tophat-2.0.12, parameter: --fusion-search --keep-fastq-order --bowtie1 --no-coverage-search). Reads mapped to back-splicing junction sites identified using TopHat-Fusion were further retrieved using CIRCexplorer2 pipeline based on known gene annotations (human (hg38): gencode.v27.annotation.gtf and refFlat.txt updated at 2017/08/23; human (hg19): refFlat.txt updated at 2016/01/04, knownGene.txt updated at 2013/01/30 and ensGene.txt updated at 2014/04/06; mouse (mm10): refFlat.txt 2016/06/04, knownGene.txt 2015/06/01 and ensGene.txt updated at 2014/04/06; rat: refFlat.txt updated at 2018/01/01

A**Tool**

To analyze circRNAs:

1. Provide circRNA locations in "List of circRNAs";
2. Select "Species";
3. Select "Cell Line". Select multiple items quickly by "Shift + Click";
4. Select "Plot Type";
5. Click "Draw" button.

Options

1. List of circRNAs

Chr3:107710451-107755678
 Chr18:49272177-49333853
 Chr4:2176265-2229243
 Chr21:33581301-33586181
 Chr2:127053421-127057601

2. Species

Human (hg38)

3. Cell Line

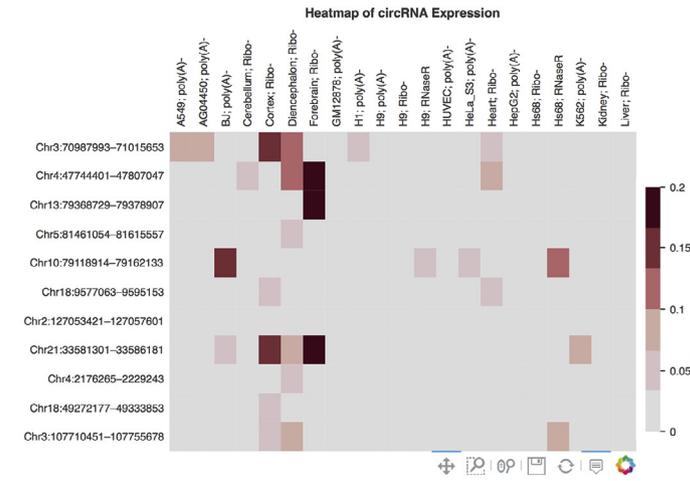
A549
 AG04450
 BJ
 Cerebellum

4. Plot Type

heatmap

Draw

Output

**B****Tool**

To analyze circRNAs:

1. Provide circRNA locations in "List of circRNAs";
2. Select "Species";
3. Select "Cell Line". Select multiple items quickly by "Shift + Click";
4. Select "Plot Type";
5. Click "Draw" button.

Options

1. List of circRNAs

Chr3:107710451-107755678
 Chr18:49272177-49333853
 Chr4:2176265-2229243
 Chr21:33581301-33586181
 Chr2:127053421-127057601

2. Species

Human (hg38)

3. Cell Line

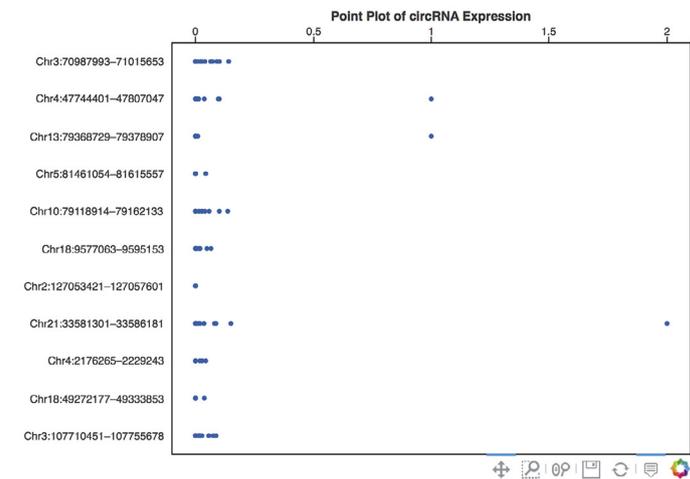
A549
 AG04450
 BJ
 Cerebellum

4. Plot Type

point

Draw

Output

**Figure 3** Screenshots of a new "Tool" interface for direct online analysis

A. Heatmap to show differential circRNA expression among selected samples. **B.** Point plot to show differential circRNA expression among selected samples.

and ensGene.txt updated at 2017/09/03; zebrafish (danRer10): refFlat.txt updated at 2018/02/25 and ensGene.txt 2018/02/04; fly (dm6): refFlat.txt updated at 2017/05/28 and ensGene.txt updated at 2018/02/04; worm (ce10): refFlat.txt updated at 2013/03/18 and ensGene.txt updated at 2012/06/05) for circRNA annotation. Alternative back-splicing was analyzed using CIRCexplorer2 with default parameters [35]. For each of the RNA-seq datasets analyzed, a circRNA annotation track file (BED), an A5BS annotation track file (BED), and an A3BS annotation track file (BED) were generated after CIRCexplorer2 analysis. Another circRNA prediction pipeline, MapSplice (version: 2.1.8, parameter: --fusion --min-fusion-distance 200) [36], was also used for human circRNA annotation. The expression level of circRNAs was evaluated with FPM.

Conservation analysis

In order to obtain the circRNAs conserved between humans and mice, LiftOver tool (parameters: -bedPlus=3 -tab -minMatch=0.1 -minBlocks=1) was used to identify orthologous back-splicing sites between humans and mice. For each human circRNA, the circRNA is defined as a conserved circRNA between humans and mice, if it is expressed in mouse orthologous locus within 5-nt difference; otherwise, the circRNA would be considered as human-specific, if no mouse circRNA ortholog is found. Similar pipeline was also applied to mouse circRNAs. It is worth noting that more conserved circRNAs are expected to be identified if additional datasets are available for this analysis in the future.

Authors' contributions

LY conceived and designed the project. RD, XKM, and GWL collected the data. RD analyzed the data with the help from XKM and GWL. XKM designed the website with the help from RD. LY wrote the paper with inputs from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Acknowledgments

We are grateful to LL Chen and all lab members for discussion. This work is supported by the National Natural Science Foundation of China (Grant Nos. 31730111, 31471241, and 91540115).

References

- [1] Chen LL, Yang L. Regulation of circRNA biogenesis. *RNA Biol* 2015;12:381–8.
- [2] Chen LL. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol* 2016;17:205–11.
- [3] Yang L. Splicing noncoding RNAs from the inside out. *Wiley Interdiscip Rev RNA* 2015;6:651–60.
- [4] Li X, Yang L, Chen LL. The biogenesis, functions, and challenges of circular RNAs. *Mol Cell* 2018;71:428–42.
- [5] Wilusz JE. A 360 degrees view of circular RNAs: from biogenesis to functions. *Wiley Interdiscip Rev RNA* 2018;9:e1478.
- [6] Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, et al. Scrambled exons. *Cell* 1991;64:607–13.
- [7] Cocquerelle C, Daubersies P, Majerus MA, Kerckaert JP, Bailleul B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J* 1992;11:1095–8.
- [8] Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, et al. Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell* 1993;73:1019–30.
- [9] Cocquerelle C, Mascrez B, Hetuin D, Bailleul B. Mis-splicing yields circular RNA molecules. *FASEB J* 1993;7:155–60.
- [10] Pasman Z, Been MD, Garcia-Blanco MA. Exon circularization in mammalian nuclear extracts. *RNA* 1996;2:603–10.
- [11] Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 2012;7:e30733.
- [12] Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;19:141–57.
- [13] Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;495:333–8.
- [14] Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet* 2013;9:e1003777.
- [15] Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, et al. Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep* 2014;9:1966–80.
- [16] Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell* 2014;159:134–47.
- [17] Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* 2015;10:170–7.
- [18] Rybak-Wolf A, Stottmeister C, Glazar P, Jens M, Pino N, Giusti S, et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 2015;58:870–85.
- [19] Gao Y, Zhao F. Computational strategies for exploring circular RNAs. *Trends Genet* 2018;34:389–400.
- [20] Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, et al. Exon circularization requires canonical splice signals. *Cell Rep* 2015;10:103–11.
- [21] Wang Y, Wang Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* 2015;21:172–9.
- [22] Liang DM, Tatomer DC, Luo Z, Wu H, Yang L, Chen LL, et al. The output of protein-coding genes shifts to circular RNAs when the pre-mRNA processing machinery is limiting. *Mol Cell* 2017;68:940–954.e3.
- [23] Zhang Y, Xue W, Li X, Zhang J, Chen S, Zhang JL, et al. The biogenesis of nascent circular RNAs. *Cell Rep* 2016;15:611–24.
- [24] Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014;56:55–66.
- [25] Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* 2014;28:2233–47.
- [26] Conn SJ, Pillman KA, Toubia J, Conn VM, Salamanidis M, Phillips CA, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell* 2015;160:1125–34.

- [27] Li X, Liu CX, Xue W, Zhang Y, Jiang S, Yin QF, et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol Cell* 2017;67:214–227.e7.
- [28] Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature* 2013;495:384–8.
- [29] Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 2015;22:256–64.
- [30] Chen YG, Kim MV, Chen X, Batista PJ, Aoyama S, Wilusz JE, et al. Sensing self and foreign circular RNAs by intron identity. *Mol Cell* 2017;67:228–238.e5.
- [31] You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* 2015;18:603–10.
- [32] Dong R, Ma XK, Chen LL, Yang L. Increased complexity of circRNA expression during species evolution. *RNA Biol* 2017;14:1064–74.
- [33] Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* 2014;15:409.
- [34] Gao Y, Wang J, Zheng Y, Zhang J, Chen S, Zhao F. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat Commun* 2016;7:12060.
- [35] Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* 2016;26:1277–87.
- [36] Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* 2010;38:e178.
- [37] Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;20:1666–70.
- [38] Xia S, Feng J, Chen K, Ma Y, Gong J, Cai F, et al. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res* 2018;46:D925–9.
- [39] Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* 2016;6:34985.
- [40] Hansen TB, Venø MT, Damgaard CK, Kjems J. Comparison of circular RNA prediction tools. *Nucleic Acids Res* 2016;44:e58.