

circAtlas 3.0: a gateway to 3 million curated vertebrate circular RNAs based on a standardized nomenclature scheme

Wanying Wu¹, Fangqing Zhao^{1,2,3,*} and Jinyang Zhang^{1,*}

¹Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou, China

*To whom correspondence should be addressed. Tel: +86 10 84504372; Fax: +86 10 84504372; Email: zhangjinyang@biols.ac.cn

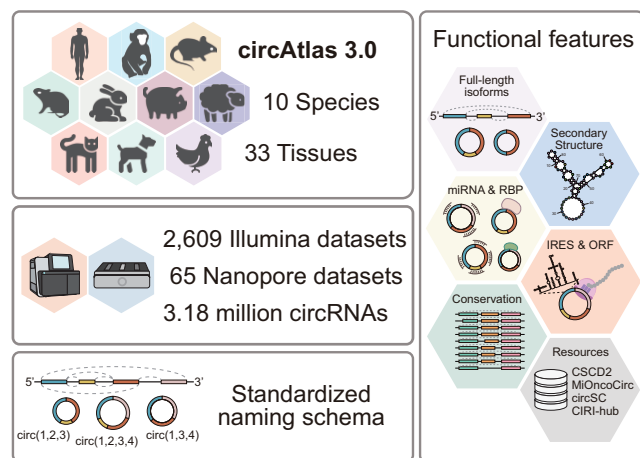
Correspondence may also be addressed to Fangqing Zhao. Tel: +86 10 84504172; Fax: +86 10 84504172; Email: zhfq@biols.ac.cn

Present address: Wanying Wu, Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany.

Abstract

Recent studies have demonstrated the important regulatory role of circRNAs, but an in-depth understanding of the comprehensive landscape of circRNAs across various species still remains unexplored. The current circRNA databases are often species-restricted or based on outdated datasets. To address this challenge, we have developed the circAtlas 3.0 database, which contains a rich collection of 2674 circRNA sequencing datasets, curated to delineate the landscape of circRNAs within 33 distinct tissues spanning 10 vertebrate species. Notably, circAtlas 3.0 represents a substantial advancement over its precursor, circAtlas 2.0, with the number of cataloged circRNAs escalating from 1 007 087 to 3 179 560, with 2 527 528 of them being reconstructed into full-length isoforms. circAtlas 3.0 also introduces several notable enhancements, including: (i) integration of both Illumina and Nanopore sequencing datasets to detect circRNAs of extended lengths; (ii) employment of a standardized nomenclature scheme for circRNAs, providing information of the host gene and full-length circular exons; (iii) inclusion of clinical cancer samples to explore the biological function of circRNAs within the context of cancer and (iv) links to other useful resources to enable user-friendly analysis of target circRNAs. The updated circAtlas 3.0 provides an important platform for exploring the evolution and biological implications of vertebrate circRNAs, and is freely available at <http://circatlas.biols.ac.cn> and <https://ngdc.cncb.ac.cn/circatlas>.

Graphical abstract



Introduction

Circular RNAs (circRNAs) are a large class of covalently closed circular transcripts that are widely existed in eukaryotes. Recent studies have demonstrated the important regulatory role of circRNAs, including miRNA/RBP sponge (1,2), encoding peptide (3,4), regulating transcription (5,6)

and translation (7,8). Despite the increasing studies of circRNAs over the past decade, the overall landscape of circRNAs within vertebrates still remains significantly underexplored. Therefore, a comprehensive exploration of circRNAs across different species is essential to understand their functions in different biological processes.

Received: August 12, 2023. Revised: August 31, 2023. Editorial Decision: September 4, 2023. Accepted: September 8, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

The growing number of circRNA sequencing datasets has facilitated the emergence of multiple databases aimed at elucidating the circRNA landscape in different species (9–12). However, most of these databases are still limited to specific species or contain a relatively small number of datasets. In 2020, we developed the circAtlas 2.0 (13), which included a repository of one million circRNAs from six distinct species. Given the substantial expansion of circRNA sequencing datasets in recent years, a more comprehensive database of the vertebrate circRNA landscape and their functional profiles has become an imperative requirement. This urgency arises from the necessity to establish a robust platform that not only facilitates insights into circRNA evolution but also serves as a means to prioritize functionally significant circRNA candidates.

The high sequence similarity of circRNA and its cognate linear counterparts makes it difficult to reconstruct full-length circRNA isoforms using short-read Illumina RNA-seq (14). Although several computational tools have been developed to reconstruct circRNAs from these short-read sequencing data (15,16), the reconstruction of long circRNA sequences remains constrained (17). Recent advances in long-read sequencing technologies have promoted the development of several long-read based circRNA sequencing strategies (18–22). Using the rolling circle reverse transcription or RCA assay, these methods can reconstruct the full-length circRNA sequences, enabling the robust profiling of full-length circRNAs across extended genomic spans. Thus, these long-read sequencing datasets offer a distinct advantage in discerning circRNA internal structures and enabling effective functional prioritization.

To this end, we constructed the circAtlas 3.0 database, which comprises a comprehensive compendium of 2609 Illumina and 65 nanopore RNA-seq datasets from 33 diverse tissues within 10 distinct species. Compared to other resources, circAtlas 3.0 provides the largest collection of over 3 million vertebrate circRNAs, with their full-length sequences and comprehensive functional analyses including orthologous circRNAs, secondary structure, miRNA binding sites, RBP binding sites and IRES/ORF prediction. In addition, circAtlas 3.0 also incorporated a standardized nomenclature scheme for circRNAs, which will bridge the gap arising from inconsistent nomenclature practices across various circRNA resources.

Materials and methods

Data collection

To obtain a comprehensive cohort for profiling circRNA across different species, a total of 2609 Illumina and 65 nanopore RNA-seq datasets from 33 different tissues in 10 species were downloaded from the SRA (<https://www.ncbi.nlm.nih.gov/sra>) (23) and GSA (<https://ngdc.cncb.ac.cn/gsa/>) (24) databases. The hg38 and mm10 reference genomes and GTF annotations were downloaded from the GENCODE project (25), and other genomes of rheMac8, rn6, oryCun2, susScr11, oviAri3, felCat9, canFam3 and galGal4 were downloaded from the UCSC Genome Browser database (26). The summary of the newly collected samples compared to circAtlas 2.0 can be found in Supplementary Table S1.

Identification and reconstruction of full-length circRNAs

For Illumina RNA-seq data, the cleaned reads were aligned to the reference genome using bwa (v0.7.17) (27), and circRNAs were subsequently identified and quantified using the CIRI2 (v2.0.6) (28,29) and CIRIquant (v1.1.2) (30) pipeline. To reconstruct the full-length sequence of circRNA isoforms, the CIRI-full (v2.1.1) (15,16,31) and CIRI-vis (v1.4.1) (32) pipeline were employed to assemble full-length circRNAs using the reverse overlap feature. To further reduce the number of false positives (33), three other software, CIRCexplorer2 (v2.3.8) (34), DCC (v0.5.0) (35) and find_circ (v1.2) (36), were also used to identify circRNAs, and circRNAs that are supported by at least one of the other tools were retained for further analysis.

For long-read sequencing data, the raw sequencing data from the CIRI-long protocol were trimmed using Porechop (<https://github.com/rwick/Porechop>), and CIRI-long (v1.1.0) (18,37) was performed to identify circRNAs from nanopore sequencing reads. For other long-read circRNA sequencing protocols including isoCirc (19), circFL-seq (20) and an adapted version of CIRI-long (38,39), the processed data were obtained from their original manuscripts. Finally, the reconstructed full-length sequences were used to annotate the spliced isoforms of the identified circRNAs (Figure 1).

Standardized naming of circRNAs

To achieve a standardized nomenclature scheme between different circRNA resources, we employed the circRNA naming criteria that recently proposed by the circRNA community (40). In brief, circRNAs were first divided into different types, and then named according to the order of their circular exons. The final unified ID of each circRNA consisted of the circRNA type, the name of their gene, the annotation of each circular exon, and an appended numerical ID to distinguish circRNAs with the same exon annotation, which provides nomenclature information on the internal structure of circRNAs.

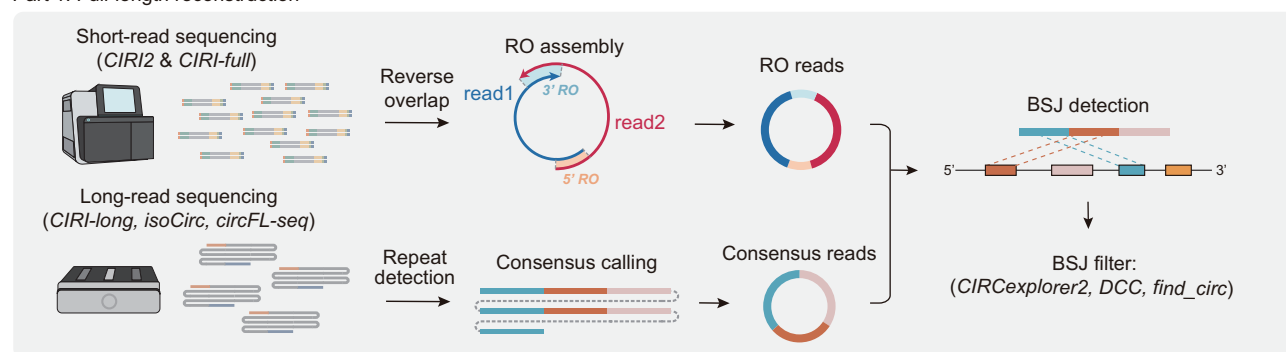
Orthology analysis of circRNAs

Orthologous circRNAs were predicted using a customized assay similar to circAtlas 2.0 (13). Briefly, the candidate orthologous genes were first obtained from the OMA orthology database (41). Then, the 50 bp fragments on both sides of the circRNA back-splicing junction (BSJ) were extracted to represent the BSJ sequence. Both full-length and BSJ sequences of circRNAs in in orthologous gene pairs were then aligned using BLAT (42) and a threshold of 50% identity of full-length or BSJ sequences. Finally, orthologous circRNA groups were determined using MultiMSOAR v2.0 software (43).

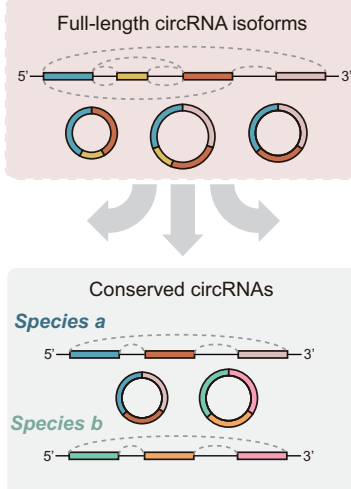
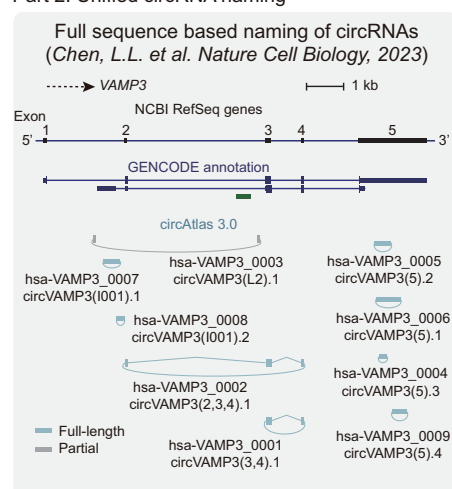
Conversion of circRNA ID and position in different genome assemblies

circAtlas 3.0 provides user-friendly conversion of circAtlas ID to various databases including circBase (9), circRNADb (44), deepBase v2.0 (45), CIRCpedia v2 (10). The conversion of circRNA BSJ between circAtlas 3.0 and several commonly used assemblies (hg19, mm9, rheMac10, rn7, susScr3, oviAri4, felCat8, canFam6, galGal6) was performed using LiftOver.

Part 1. Full-length reconstruction



Part 2. Unified circRNA naming



Part 3. Functional annotation

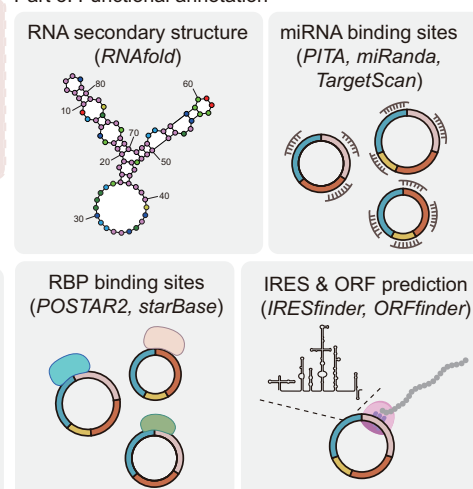


Figure 1. Schematic of the data processing pipeline in the circAtlas 3.0 database. The circAtlas 3.0 integrated both Illumina and nanopore sequencing datasets to identify full-length circRNA isoforms, and three additional circRNA identification tools were used to filter high-confidence back-splicing events. Based on the identified circRNA isoforms, the circRNAs were named according to the host gene and circular exon annotation. Then, orthology analysis and functional annotation including RNA secondary structure, miRNA/RBP binding sites and IRES/ORF prediction were performed as described in the Materials and Methods section.

RNA structure and functional annotation

Based on the assembled full-length circRNA sequences, the secondary structure was predicted using the RNAfold algorithm of the Vienna RNA 2.0 package (46). To predict the circRNA-miRNA interaction, mature miRNA sequences of different species were downloaded from miRBase (47). Then, three independent miRNA prediction software, miRanda (v3.3.a) (48), TargetScan (v7.0) (49) and PITA (v2.1.2) (50), were used to predict the circRNA-microRNA binding position. The predicted miRNA response elements on the circRNAs were integrated to represent the possible circRNA-miRNA interactions. To predict the circRNA-RBP interaction, the high-confidence RBP binding sites were downloaded from the starBase (v2.0) (51) and POSTAR2 (52) database. The RBP binding sites located within the 1 kb flanking region of BSJs or within the circular exons were retained for further analysis. To explore the coding potential of circRNAs, IRESfinder (v1.1.0) (53) and ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder>) were performed to predict IRES and coding peptides in the circRNA sequence.

Database deployment

The circAtlas 3.0 database was deployed in a virtual ma-

chine running CentOS 7.9. The backend of the circAtlas 3.0 database was based on MySQL (v14.14), and the web server was deployed using Apache (v2.4.6) and PHP (v7.4.33). The interactive visualization of the analysis results was implemented using jQuery (3.6.4), Apache ECharts (5.4.0) (54) and D3.js (v3.5.17). All online analysis was performed using Python (v3.10.4). The circAtlas 3.0 database is publicly available at <http://circatlas.biols.ac.cn> and <https://ngdc.cncb.ac.cn/circatlas>, and no login is required. Users can easily download the detected circRNAs and full-length circRNA sequences. All software involved in the circAtlas 3.0 is also listed on the download page.

Results

A comprehensive atlas of circRNAs across 10 vertebrate species

The circAtlas 3.0 contains an updated compendium of 2536 total RNA and 73 RNase R treated Illumina RNA-seq libraries from 10 vertebrate species. In order to achieve reliable full-length circRNA isoform reconstruction, we have also included 65 long-read circRNA sequencing data from human and mouse samples. In total, 21.02 terabytes of Illumina and 651.62 gigabytes of Nanopore sequencing data

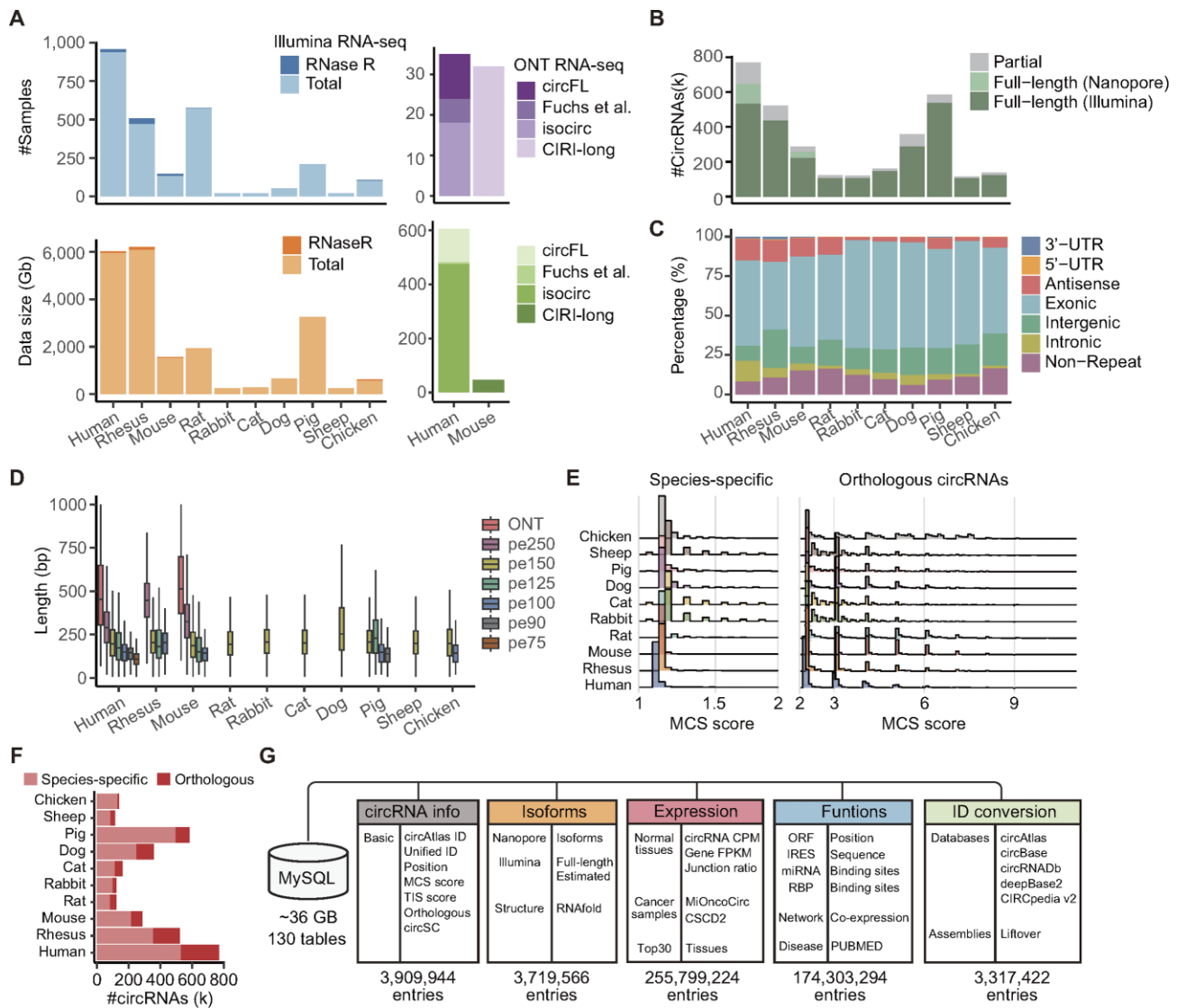


Figure 2. Data statistics and summary of identified circRNAs. **(A)** Number of collected samples and total library size in each species. The nanopore sequencing datasets were divided according to their sequencing protocol. **(B)** Number of partially assembled and reconstructed full-length circRNAs in each species. **(C)** Proportion of circRNAs in different genomic regions. **(D)** Length of reconstructed circRNA in each species. Colors represent circRNAs detected by different sequencing strategies. **(E)** The distribution of MCS scores of the detected circRNAs. All circRNAs were divided into species-specific circRNAs (left) and orthologous circRNAs that are conservatively expressed in multiple species (right). **(F)** Number of species-specific and orthologous circRNAs in each species. **(G)** Schematic view of the MySQL database structure of circAtlas 3.0.

were collected, forming a rich resource for profiling circRNAs across different species (Figure 2A). Through stringent filters, we identified over 3.18 million circRNAs across 10 species, with a remarkable 79.49% of these circRNAs assembled into full-length sequences. Notably, 15.06% of human and 13.23% of mouse circRNAs in our dataset were supported by nanopore long reads, significantly enhancing the foundational support for downstream functional analyses (Figure 2B).

To better demonstrate the diversity of circRNAs, all circRNAs were divided into different groups according to the region of their back-splicing junctions (Figure 2C). Interestingly, we observed a significantly high increase in the proportion of antisense circRNAs from carnivores to primates, which is consistent with the pervasive antisense transcription of mammalian genes (55). It should be noted that ciRS-7, the most famous circRNA that was first found to be a miRNA sponge

(56), also originates from the antisense strand of the CDR1 gene. Considering that these antisense circRNAs may have a potential function to bind to cognate precursor mRNAs, these findings underscore the utility of circAtlas 3.0 to explore the regulatory roles of vertebrate circRNAs. As shown in Figure 2D, the average length of the reconstructed circRNAs ranged from 200 to 500 nucleotides across all 10 species. Since different read lengths can affect the ability to reconstruct longer circRNAs (16), the length of circRNAs detected by different sequencing strategies was calculated separately. Here, a dramatic increase in circRNA length was observed with the increasing sequencing length. The incorporation of nanopore sequencing datasets substantially broadens the scope of circRNA detection, magnifying the range from the 200–300 nt observed in Illumina RNA-seq to an approximate of 500 nt. This underscored the indispensability of integrating long-read sequencing technology to expand the ability for reconstruct-

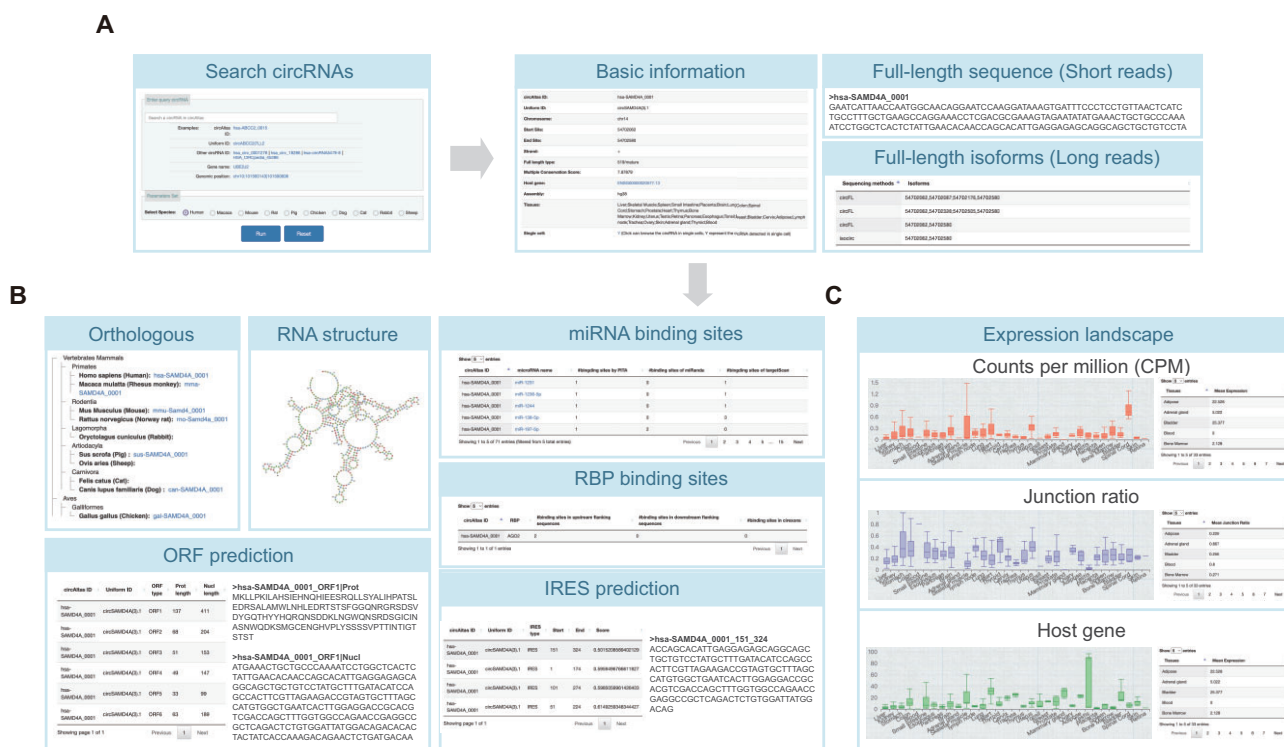


Figure 3. Basic search and analysis functions in circAtlas 3.0. **(A)** Summary of the search function and basic information of circRNAs. **(B)** Summary of the orthologous circRNAs and functional analysis results based on full-length sequences. **(C)** Visualization of the expression levels of circRNAs and their host genes. The distribution of junction ratio in different species is also shown to represent the strength of back-splicing events.

ing longer circRNAs, thereby enriching our understanding of circRNA landscapes.

Based on the reconstructed full-length circRNA sequences, we further calculated the orthologous circRNA across diverse species. Here, the MCS score metric (13) was employed to measure the conservation of circRNAs. Similar to circAtlas 2.0, most of the circRNAs were only detected in limited species and tissues (Figure 2E), which is consistent with the high species and tissue specificity of circRNAs (12). We further divided all circRNAs into two categories: species-specific circRNAs (only detected in one species) and orthologous circRNAs (consistently expressed across multiple species). While as expected, species-specific circRNAs dominated most species (Figure 2F), a significant proportion of orthologous circRNAs was observed between human and rhesus, which could be explained by the close relationship between human and rhesus. Taken together, these results demonstrate that circAtlas 3.0 can be applied for exploring the evolution of circRNAs in different vertebrate species.

Database structure

The circAtlas 3.0 database consisted of 3.18 million consistently named circRNAs, which constituted a comprehensive relationship database for circRNA expression and functional annotation (Figure 2F). For each circRNA, the basic information including circAtlas ID, unified ID, BSJ position, and other attributes was stored in the circRNA table, while supplementary details including full-length isoforms, expression values, functional annotations and conversion to other genome assemblies or circRNA resources were stored as separate tables. In summary, a total of 130 MySQL tables with 441 049 450 entries were generated, resulting in ~36 GB of MySQL data.

Here, this huge dataset provides an important basis for the exploration of circRNA landscape using our updated circAtlas 3.0 database.

Web interface and applications

The circAtlas 3.0 database provides user-friendly search and download functionalities, designed to explore the circRNA expression landscape and their biological functions across the 10 collected species. As shown in Figure 3A, users can (i) search for specific circRNAs by circAtlas ID, unified ID, genomic position and host gene name or ID; (ii) browse all circRNAs in specific species and optionally sorted by MCS scores or circRNA length; (iii) select the top 30 highly expressed circRNAs per tissue. After selecting a target circRNA, the basic information will be displayed, including the full-length isoform sequences in both short-read and long-read sequencing datasets.

Based on the full-length sequences, circAtlas 3.0 provides a variety of analysis results (Figure 3B). First, circAtlas 3.0 provides the orthologous circRNAs across various species, the user can click the links of orthologous circRNA, which provides important criteria for selecting species-specific or highly conserved circRNAs. Furthermore, circAtlas 3.0 provides useful functional analysis results, including (i) RNA secondary structure prediction, which provides useful information for linking RNA structure with their biological functions; (ii) miRNA binding sites using three different miRNA prediction software. Each row represents the number of binding sites of specific microRNA; (iii) RBP binding sites, where each row represents the number of binding sites in the circular exons and upstream/downstream flanking regions; (iv) IRES prediction results. Each row indicates the position and

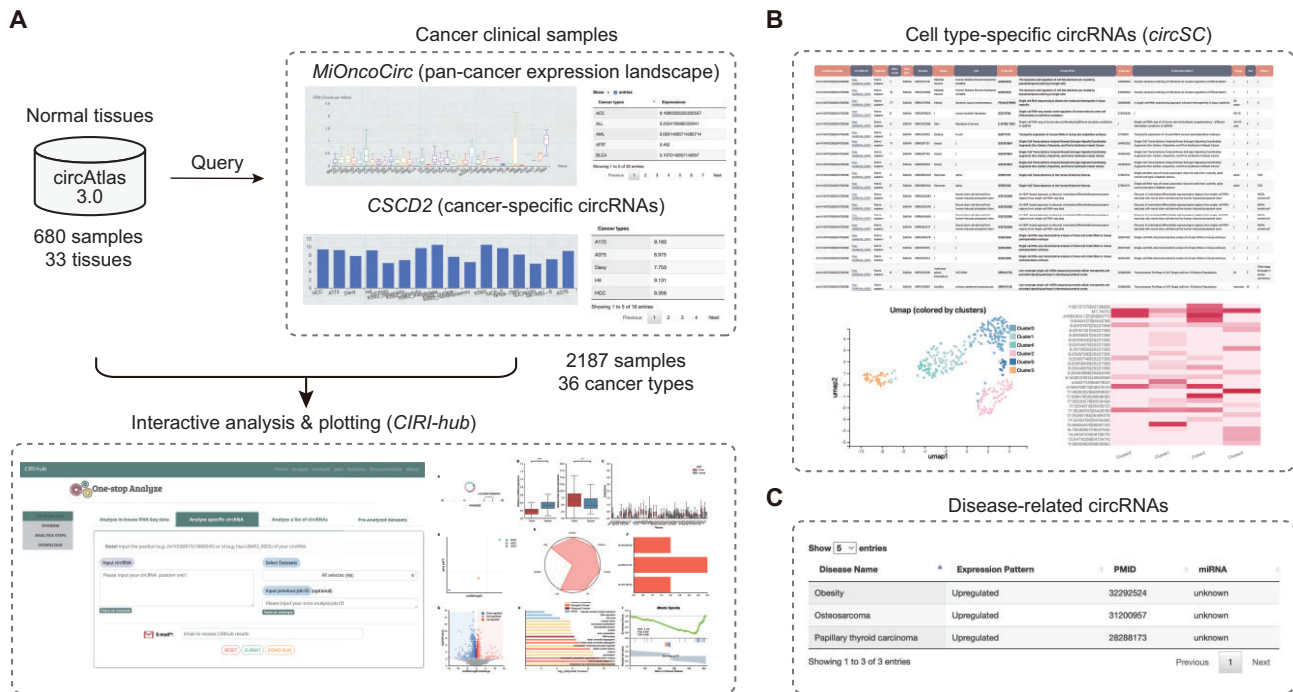


Figure 4. Overview of the integrated links to other resources. **(A)** circAtlas 3.0 provides the circRNA expression landscape in clinical cancer samples (MiOncoCirc & CSCD2), and links to CIRI-hub for interactive analysis and plotting of circRNAs in cancer. **(B)** circAtlas 3.0 provides links of circRNAs detected in full-length single-cell sequencing datasets (circSC). **(C)** circAtlas 3.0 integrates the circad, circR2Disease and circRNADisease databases to provide curated functions of disease-related circRNAs.

sequence of the predicted IRES element and (v) ORF prediction results. The predicted nucleotide and peptide sequences of the ORF are shown. The IRES and ORF prediction results can be retrieved using the circRNA ID on a separate page. Collectively, these analysis results provide insight into the common function of circRNAs as miRNA/RBP sponges or encoding short peptides. Finally, circAtlas 3.0 provides the expression landscape of the target circRNA and its host gene within 33 different tissues (Figure 3C). To explore the splicing strength of back-splicing events, the junction ratio of circRNA is also shown. The change of expression level and junction ratio provides the basis for exploring linear-circular switching and circular transcript usage switching events (30). In summary, the rich information of circAtlas 3.0 provides useful resources for functional prioritization of circRNAs across 10 vertebrate species.

Integration of other useful resources

The previous circAtlas 2.0 consisted mainly of normal tissues samples. However, since circRNAs have important regulatory roles in disease and cancer, we also integrated the expression landscape of circRNAs in clinical samples across 36 cancer types from the MiOncoCirc (57) and CSCD2 (58) databases (Figure 4A). The expression levels of circRNAs in different cancer types are also presented in the results page. Moreover, we provided a link to CIRI-hub (59), an online interactive analysis platform based on human samples collected in the circAtlas, MiOncoCirc and CSCD2. Additionally, circAtlas 3.0 also included the link to circRNAs in circSC (60), allowing user-friendly analysis of the single-cell level expression landscape of target circRNAs (Figure 4B). Finally, circAtlas 3.0 integrates 2269 latest curated disease-associated circRNAs from the circad (61), circR2Disease (62) and circR-

NADisease (63) databases (Figure 4C). Taken together, the integration of these online resources provides a useful platform for exploring the biological function of target circRNAs.

Comparison with other resources

Recent studies have developed several databases to investigate circRNAs in different tissues and species (Supplementary Table S2). For instance, circBase (9) contains 96891 circRNAs from 78 samples across five species, and CIRCpedia v2 (10) consists of 262782 circRNAs from 185 samples from six species. TSCD (64) contains 284 296 circRNAs from 16 adult and 15 fetal human tissues. Notably, most of these databases are limited by the small and outdated number of samples and species, making them difficult to get a comprehensive overview of circRNAs in vertebrates. Among them, our circAtlas 2.0 database (13) provides the largest collection of 1 million circRNAs from 1070 samples, but it is still limited to model species such as humans, rhesus, mouse and rat. In addition, MiOncoCirc (57), CSCD2 (58) and CircRic (65) focus on circRNAs in clinical cancer samples, but lack normal tissues as a control. In this study, we presented the most comprehensive set of circRNAs and their expression and functional profiles from 2674 RNA-seq datasets across 10 species. The new features of circAtlas 3.0 are as follows: (i) the largest collection of over 3.1 million circRNAs from 10 species; (ii) integration of accurate full-length sequences from nanopore long reads; (iii) embedded expression landscape of circRNAs in both normal tissues and cancer samples; (iv) the employment of a standardized nomenclature scheme recently proposed by the circRNA community. Taken together, circAtlas 3.0 provides convenient features for studying the evolution and biological functions of vertebrate circRNAs.

Discussion

The circAtlas 3.0 provides the largest collection of more than 3.2 million consistently named circRNAs across 10 species and 33 tissue types. The circAtlas 3.0 has two unique features: (i) it is the first database to integrate Illumina and Nanopore circRNA sequencing datasets, providing the full-length sequences and downstream functional analysis results of over 2.5 million circRNAs. The accurate full-length sequence is essential for exploring the biological function of circRNAs; (ii) it is the first database to employ the standardized nomenclature scheme proposed by the circRNA community. The unified naming can ensure clarity and reproducibility between different resources, which eliminates the ambiguous naming and reduces the meaningless effort in converting circRNA IDs between different resources. In addition, circAtlas 3.0 incorporates circRNA expression profiles from clinical cancer samples and provides links to other useful resources such as CIRI-hub, circSC and diverse disease-related databases. These extensive datasets provide an important foundation for in-depth exploration into the roles of circRNAs in cancer and disease.

With the increasing number of long-read circRNA sequencing data, the future version of circAtlas could be enhanced by including assimilating additional long-read sequencing datasets, which will greatly facilitate the identification of longer circRNA isoforms. Furthermore, the inclusion of diverse vertebrate classes will broaden the circRNA landscape in the current databases and help us to understand the evolution of circRNAs. Overall, circAtlas 3.0 provides a useful resource of circRNAs across a variety of species and tissues as well as clinical samples, to improve our understanding of the significance of circRNAs in evolution and disease.

Data availability

The circAtlas 3.0 database is freely available at <http://circatlas.biols.ac.cn> and <https://ngdc.cncb.ac.cn/circatlas>. The RNA-seq datasets in circAtlas 3.0 can be accessed in Supplementary Table S1.

Supplementary data

[Supplementary Data](#) are available at NAR Online.

Acknowledgements

We thank NGDC for its support in providing computing resources in database construction.

Funding

National Natural Science Foundation of China (NSFC) [32025009, 32130020, 32200530]; National Key R&D Project [2021YFA1300500, 2021YFA1302000]; China National Postdoctoral Program for Innovative Talents [BX20220296]; China Postdoctoral Science Foundation (CPSF) [2022M723134]; Special Research Assistant of the Chinese Academy of Science. Funding for open access charge: National Key R&D Project [2021YFA1302000].

Conflict of interest statement

None declared.

References

- Xu,H., Guo,S., Li,W. and Yu,P. (2015) The circular RNA Cdr1as, via miR-7 and its targets, regulates insulin transcription and secretion in islet cells. *Sci. Rep.*, **5**, 12453.
- Liu,C.X., Li,X., Nan,F., Jiang,S., Gao,X., Guo,S.K., Xue,W., Cui,Y., Dong,K., Ding,H., *et al.* (2019) Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell*, **177**, 865–880.
- Lin,H.H., Chang,C.Y., Huang,Y.R., Shen,C.H., Wu,Y.C., Chang,K.L., Lee,Y.C., Lin,Y.C., Ting,W.C., Chien,H.J., *et al.* (2023) Exon junction complex mediates the cap-independent translation of circular RNA. *Mol. Cancer Res.* <https://doi.org/10.1158/1541-7786.MCR-22-0877>.
- Zhang,M., Zhao,K., Xu,X., Yang,Y., Yan,S., Wei,P., Liu,H., Xu,J., Xiao,F., Zhou,H., *et al.* (2018) A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat. Commun.*, **9**, 4475.
- Xu,X., Zhang,J., Tian,Y., Gao,Y., Dong,X., Chen,W., Yuan,X., Yin,W., Xu,J., Chen,K., *et al.* (2020) CircRNA inhibits DNA damage repair by interacting with host gene. *Mol. Cancer*, **19**, 128.
- 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.
- Chen,S., Cao,X., Zhang,J., Wu,W., Zhang,B. and Zhao,F. (2022) circVAMP3 drives CAPRIN1 phase separation and inhibits hepatocellular carcinoma by suppressing c-Myc translation. *Adv. Sci. (Weinh)*, **9**, e2103817.
- Chen,S., Zhang,J. and Zhao,F. (2022) Screening linear and circular RNA transcripts from stress granules. *Genom. Proteom. Bioinform.* <https://doi.org/10.1016/j.gpb.2022.01.003>.
- Glazar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
- Dong,R., Ma,X.K., Li,G.W. and Yang,L. (2018) CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genom. Proteom. Bioinform.*, **16**, 226–233.
- Liu,M., Wang,Q., Shen,J., Yang,B.B. and Ding,X. (2019) Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol.*, **16**, 899–905.
- Ji,P., Wu,W., Chen,S., Zheng,Y., Zhou,L., Zhang,J., Cheng,H., Yan,J., Zhang,S., Yang,P., *et al.* (2019) Expanded expression landscape and prioritization of circular RNAs in mammals. *Cell Rep.*, **26**, 3444–3460.
- Wu,W., Ji,P. and Zhao,F. (2020) CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.*, **21**, 101.
- Gao,Y. and Zhao,F. (2018) Computational strategies for exploring circular RNAs. *Trends Genet.*, **34**, 389–400.
- Zheng,Y. and Zhao,F. (2018) In: Dieterich,C. and Papantonis,A. (eds.) *Circular RNAs: Methods and Protocols*. Springer New York, NY, pp. 1–8.
- Zheng,Y., Ji,P., Chen,S., Hou,L. and Zhao,F. (2019) Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Med.*, **11**, 2.
- Zhang,J. and Zhao,F. (2021) Reconstruction of circular RNAs using Illumina and Nanopore RNA-seq datasets. *Methods*, **196**, 17–22.
- Zhang,J., Hou,L., Zuo,Z., Ji,P., Zhang,X., Xue,Y. and Zhao,F. (2021) Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat. Biotechnol.*, **39**, 836–845.
- Xin,R., Gao,Y., Gao,Y., Wang,R., Kadash-Edmondson,K.E., Liu,B., Wang,Y., Lin,L. and Xing,Y. (2021) isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat. Commun.*, **12**, 266.
- Liu,Z., Tao,C., Li,S., Du,M., Bai,Y., Hu,X., Li,Y., Chen,J. and Yang,E. (2021) circFL-seq reveals full-length circular RNAs with rolling circular reverse transcription and nanopore sequencing. *Elife*, **10**, e69457.

21. Rahimi,K., Veno,M.T., Dupont,D.M. and Kjems,J. (2021) Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nat. Commun.*, **12**, 4825.
22. Zhang,J. and Zhao,F. (2021) Characterizing circular RNAs using nanopore sequencing. *Trends Biochem. Sci.*, **46**, 785–786.
23. Leinonen,R., Sugawara,H., Shumway,M. and International Nucleotide Sequence Database,C. (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
24. Chen,T., Chen,X., Zhang,S., Zhu,J., Tang,B., Wang,A., Dong,L., Zhang,Z., Yu,C., Sun,Y., *et al.* (2021) The genome sequence archive family: toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.*, **19**, 578–583.
25. Frankish,A., Diekhans,M., Jungreis,I., Lagarde,J., Loveland,J.E., Mudge,J.M., Sisu,C., Wright,J.C., Armstrong,J., Barnes,I., *et al.* (2021) Gencode 2021. *Nucleic Acids Res.*, **49**, D916–D923.
26. Nassar,L.R., Barber,G.P., Benet-Pages,A., Casper,J., Clawson,H., Diekhans,M., Fischer,C., Gonzalez,J.N., Hinrichs,A.S., Lee,B.T., *et al.* (2023) The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.*, **51**, D1188–D1195.
27. Li,H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv doi: <https://arXiv.org/abs/10.48550/1303.3997>, 26 May 2013, preprint: not peer reviewed.
28. Gao,Y., Zhang,J. and Zhao,F. (2018) Circular RNA identification based on multiple seed matching. *Brief Bioinform.*, **19**, 803–810.
29. Gao,Y., Wang,J. and Zhao,F. (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.*, **16**, 4.
30. Zhang,J., Chen,S., Yang,J. and Zhao,F. (2020) Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat. Commun.*, **11**, 90.
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. Zheng,Y. and Zhao,F. (2020) Visualization of circular RNAs and their internal splicing events from transcriptomic data. *Bioinformatics*, **36**, 2934–2935.
33. Vromman,M., Anckaert,J., Bortoluzzi,S., Buratin,A., Chen,C.Y., Chu,Q., Chuang,T.J., Dehghannasiri,R., Dieterich,C., Dong,X., *et al.* (2023) Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nat. Methods*, **20**, 1159–1169.
34. 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.
35. Cheng,J., Metge,F. and Dieterich,C. (2016) Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics*, **32**, 1094–1096.
36. 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.
37. Hou,L., Zhang,J. and Zhao,F. (2023) Full-length circular RNA profiling by nanopore sequencing with CIRI-long. *Nat. Protoc.*, **18**, 1795–1813.
38. Fuchs,S., Babin,L., Andraos,E., Bessiere,C., Willier,S., Schulte,J.H., Gaspin,C. and Meggetto,F. (2022) Generation of full-length circular RNA libraries for Oxford Nanopore long-read sequencing. *PLoS One*, **17**, e0273253.
39. Fuchs,S., Danzmann,C., Klironomos,F., Winkler,A., Fallmann,J., Krutzfeldt,L.M., Szymanski,A., Naderi,J., Bernhart,S.H., Grunewald,L., *et al.* (2023) Defining the landscape of circular RNAs in neuroblastoma unveils a global suppressive function of MYCN. *Nat. Commun.*, **14**, 3936.
40. Chen,L.L., Bindereif,A., Bozzoni,I., Chang,H.Y., Matera,A.G., Gorospe,M., Hansen,T.B., Kjems,J., Ma,X.K., Pek,J.W., *et al.* (2023) A guide to naming eukaryotic circular RNAs. *Nat. Cell Biol.*, **25**, 1–5.
41. Altenhoff,A.M., Train,C.M., Gilbert,K.J., Mediratta,I., Mendes de Farias,T., Moi,D., Nevers,Y., Radoykova,H.S., Rossier,V., Warwick Vesztrocy,A., *et al.* (2021) OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res.*, **49**, D373–D379.
42. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
43. Shi,G., Peng,M.C. and Jiang,T. (2011) MultiMSOAR 2.0: an accurate tool to identify ortholog groups among multiple genomes. *PLoS One*, **6**, e20892.
44. 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.
45. Zheng,L.L., Li,J.H., Wu,J., Sun,W.J., Liu,S., Wang,Z.L., Zhou,H., Yang,J.H. and Qu,L.H. (2016) deepBase v2.0: identification, expression, evolution and function of small RNAs, lncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res.*, **44**, D196–D202.
46. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
47. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
48. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in *Drosophila*. *Genome Biol.*, **5**, R1.
49. Agarwal,V., Bell,G.W., Nam,J.W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
50. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
51. 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.
52. Zhu,Y., Xu,G., Yang,Y.T., Xu,Z., Chen,X., Shi,B., Xie,D., Lu,Z.J. and Wang,P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
53. Zhao,J., Wu,J., Xu,T., Yang,Q., He,J. and Song,X. (2018) IRESfinder: identifying RNA internal ribosome entry site in eukaryotic cell using framed k-mer features. *J. Genet. Genomics*, **45**, 403–406.
54. Li,D., Mei,H., Shen,Y., Su,S., Zhang,W., Wang,J., Zu,M. and Chen,W. (2018) ECharts: a declarative framework for rapid construction of web-based visualization. *Visual Informatics*, **2**, 136–146.
55. He,Y., Vogelstein,B., Velculescu,V.E., Papadopoulos,N. and Kinzler,K.W. (2008) The antisense transcriptomes of human cells. *Science*, **322**, 1855–1857.
56. 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.
57. Vo,J.N., Cieslik,M., Zhang,Y., Shukla,S., Xiao,L., Zhang,Y., Wu,Y.M., Dhanasekaran,S.M., Engelke,C.G., Cao,X., *et al.* (2019) The landscape of circular RNA in cancer. *Cell*, **176**, 869–881.
58. Feng,J., Chen,W., Dong,X., Wang,J., Mei,X., Deng,J., Yang,S., Zhuo,C., Huang,X., Shao,L., *et al.* (2022) CSCD2: an integrated interactional database of cancer-specific circular RNAs. *Nucleic Acids Res.*, **50**, D1179–D1183.
59. Wu,W., Zhang,J. and Zhao,F. (2023) CIRI-hub: an integrated and visual analytics platform for circular RNAs in cancers. bioRxiv doi: <https://doi.org/10.1101/2023.08.09.552604>, 12 August 2023, preprint: not peer reviewed.

60. Wu,W., Zhang,J., Cao,X., Cai,Z. and Zhao,F. (2022) Exploring the cellular landscape of circular RNAs using full-length single-cell RNA sequencing. *Nat. Commun.*, **13**, 3242.
61. Rophina,M., Sharma,D., Poojary,M. and Scaria,V. (2020) Circad: a comprehensive manually curated resource of circular RNA associated with diseases. *Database (Oxford)*, **2020**, baaa019.
62. Fan,C., Lei,X., Fang,Z., Jiang,Q. and Wu,F.X. (2018) CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database (Oxford)*, **2018**, bay044.
63. Zhao,Z., Wang,K., Wu,F., Wang,W., Zhang,K., Hu,H., Liu,Y. and Jiang,T. (2018) circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. *Cell Death. Dis.*, **9**, 475.
64. Xia,S., Feng,J., Lei,L., Hu,J., Xia,L., Wang,J., Xiang,Y., Liu,L., Zhong,S., Han,L., *et al.* (2017) Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform.*, **18**, 984–992.
65. Ruan,H., Xiang,Y., Ko,J., Li,S., Jing,Y., Zhu,X., Ye,Y., Zhang,Z., Mills,T., Feng,J., *et al.* (2019) Comprehensive characterization of circular RNAs in ~ 1000 human cancer cell lines. *Genome Med.*, **11**, 55.