

Computational approaches for circRNAs prediction and *in silico* characterization

Camilo Rebolledo[†], Juan Pablo Silva[†], Nicolás Saavedra and Vinicius Maracaja-Coutinho

Corresponding authors. Nicolás Saavedra, Center of Molecular Biology & Pharmacogenetics, Department of Basic Sciences, Scientific and Technological Resources, Universidad de La Frontera, Temuco 4810296, Chile. Email: nicolas.saavedra@ufrofrontera.cl; Vinicius Maracaja-Coutinho, Advanced Center for Chronic Diseases - ACCDiS, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago 8320000, Chile. Email: vinicius.maracaja@uchile.cl

[†]Camilo Rebolledo and Juan Pablo Silva contributed equally to this work.

Abstract

Circular RNAs (circRNAs) are single-stranded and covalently closed non-coding RNA molecules originated from RNA splicing. Their functions include regulatory potential over other RNA species, such as microRNAs, messenger RNAs and RNA binding proteins. For circRNA identification, several algorithms are available and can be classified in two major types: pseudo-reference-based and split-alignment-based approaches. In general, the data generated from circRNA transcriptome initiatives is deposited on public specific databases, which provide a large amount of information on different species and functional annotations. In this review, we describe the main computational resources for the identification and characterization of circRNAs, covering the algorithms and predictive tools to evaluate its potential role in a particular transcriptomics project, including the public repositories containing relevant data and information for circRNAs, recapitulating their characteristics, reliability and amount of data reported.

Keywords: circRNA, circRNA regulation, circRNA-miRNA prediction, bioinformatics

INTRODUCTION

The past decade has seen growing interest in the role of circular RNAs (circRNAs) in various diseases [1–5], highlighting their regulatory potential compared with other RNA species [6–8]. circRNAs are single-stranded non-coding RNA molecules that are covalently connected by their 3' and 5' ends, generating circRNA molecules rather than the linear structure found in typical RNAs. They can be composed of single or multiple exons, derived from intronic sequences and generated from both long non-coding RNAs (lncRNAs) or messenger RNAs (mRNAs) [1]. These molecules are resistant to exonucleases, leading to a long lifespan of more than 48 hours, which is almost twice the lifespan of mRNAs [9, 10]. Studies have shown that circRNAs are also important endogenous RNA competitors, acting as microRNA (miRNA) sponges, regulating gene expression [11].

circRNAs are present in different subcellular locations [12, 13] and their expression levels change in several tissues during development and cell differentiation [14, 15]. Although circRNAs are poorly conserved across species, apparently arising through convergent evolution, they seem to be more species-specific compared with other RNAs [14]. Literature suggests that transposable elements inserted during evolution have stabilized them and enabled their production [16].

Research has also pointed circRNAs as promising biomarkers for several diseases, including cardiovascular infirmities and

cancers. For example, *circANRIL* has been linked to atherosclerosis [17], and *circMICRA* was described as a predictive molecule potentially associated to left ventricular dysfunction [18]. In cancer, *circSMARCA5* has been correlated with tumor differentiation and cancer diameter [19], whereas *circCDYL* has been shown to act as a predictive and prognostic marker of breast cancer [20].

There are a multitude of algorithms for identifying back splicing junctions (BSJ) and circRNA formation. These algorithms are implemented in computational tools with high capacity for predicting these molecules, usually by extracting information from RNA sequencing (RNA-seq) data [21]. Predicted circRNAs are often deposited in databases and repositories that annotate the identified molecules based on various features, such as their genomic coordinates, strand of origin, mature nucleotide sequence, among other characteristics [22]; such as their interactions with other molecules like small RNAs and RNA-binding proteins (RBP). Normally, these circRNA-small RNA or circRNA-RBP relationships are also obtained through computational approaches, which are useful to obtain novel layers of information in their *in silico* characterization and, consequently, obtain further insights related to their functional roles in the cells [23, 24].

In this review, we will first provide an overview of what circRNAs are and how they are originated at the cellular level. This information is crucial for the development of predictive tools for its identification by exploring RNA-seq data or even using

Camilo Rebolledo is a PhD student in Applied Cellular and Molecular Biology at the Universidad de La Frontera, Chile. His research interests are bioinformatics, transcriptomics and molecular biology.

Juan Pablo Silva is a PhD student in Biochemistry at the Universidad de Chile, Chile. His main research interests are transcriptomics, circRNAs functional characterization and RNA epitranscriptomics.

Nicolás Saavedra is an associate professor and the President of the Scientific Ethics Committee of Universidad de La Frontera, Chile. His research interests are genetics, cardiovascular medicine and microbiome associations with chronic diseases.

Vinicius Maracaja-Coutinho is an associate professor at the Universidad de Chile. He leads the Laboratory of Integrative Bioinformatics and is the Director of the Graduate Diploma in Bioinformatics and Computational Biology. His interests are RNA bioinformatics, long non-coding RNAs and RNA epitranscriptomics.

Received: November 22, 2022. **Revised:** March 20, 2023. **Accepted:** March 30, 2023

© The Author(s) 2023. Published by Oxford University Press.

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

the transcript nucleotide sequences features. We will explore the different tools available for circRNA prediction, discussing the positive and negative aspects of them. We will then review the approaches to identify splicing variants and to perform a quantification, normalization and differential expression analysis of these RNAs, as well as the various algorithms for predicting their interactions with other molecules and to obtain insights of their functional roles in the cells. Finally, we provide a brief overview of the databases and repositories containing relevant data and information for circRNAs.

The biogenesis of circRNAs and key features to be considered in its computational prediction

The formation of mature mRNAs or lncRNAs is preceded by splicing and rearrangement processes which take place on primary RNAs (Figure 1A). Two different forms of mature RNAs can arise depending mainly on splicing: linear RNAs or circRNAs. The linear mature RNA is generated by the canonical splicing event, in which introns are removed from pre-mRNA and exons are joined together to form a mature mRNA. The spliceosome, a complex of proteins and RNA molecules, recognizes the intron-exon boundaries and removes the introns through a series of enzymatic reactions. Canonical splicing results in a single linear mRNA molecule from each pre-mRNA transcript. Alternative splicing events can also happen in this process, by which different combinations of exons can be included in the mature mRNA, resulting in multiple mature mRNA transcripts from a single pre-mRNA transcript. Alternative splicing is regulated by a combination of cis-acting elements in the pre-mRNA and trans-acting splicing factors (Figure 1B).

The circRNAs are formed when the cell spliceosome machinery joins the 3' and 5' ends of the canonical spliced RNA through a back splicing reaction, which results in the circularization of the RNA molecule (Figure 1C). Aufiero et al. (2019) [25] wrote an extensive review describing the mechanisms of circularization, pointing out three main processes: intron pairing-driven circularization, RNA-binding-protein-driven circularization and lariat-driven circularization. (i) The intron pairing-driven circularization is based on the complementarity of cis elements present in flanking introns, which normally are short non-coding regulatory sequences or short interspersed nuclear elements, like Alu sequences and GU-rich sequences, possessing sequence complementarities that guides the covalently bonded structure conformation that facilitates the circularization and production of the mature circRNAs. (ii) RNA-binding protein-driven circularization depends on trans elements guided by RBPs that interact with cis elements to conform the circularization. These proteins recognize and bind to specific motifs in the intronic flanking regions. In this way, these proteins can dimerize and achieve structures that allow the back-splicing conformation and, consequently, the RNA circularization. Liu et al. (2019) [26] used a neural network model to classify cis elements in flanking regions of back-splicing sites of 21,472 human circRNAs, identifying RNA binding and protein binding sites as a main factor for circRNA biogenesis. Huang et al. (2020) [27] summarized the functions of RNA-binding proteins, such as ADAR, QKI and FUS, and nuclear factors like NF90/NF110 as facilitators of this back-splicing process. (iii) In lariat-driven circularization, the intronic or exonic sequences containing a looped structure called 'lariats' are removed from the primary RNA during linear splicing, and the lasso-like structure is confirmed by sequence complementarity, giving rise to a circRNA that can be of exonic or intronic origin.

Distinct categories of circRNAs are recognized based on the origin of the genomic region that conforms the final circRNA sequence. Normally, intron pairing-driven and RNA-binding-protein-driven circularizations can generate three main categories, which can include the combination of multiple exons or intronic regions in the final circRNAs (Figure 1C). (i) The exonic circular RNAs (EcircRNAs) includes all exons part of the circularization event; (ii) the alternative EcircRNAs, which includes the exons processed after an alternative splicing event; and (iii) the exonic-intronic circular RNAs (EicircRNAs), in which an intronic region is retained in the final circRNA together with the exons part of its sequence. Lariat-driven circularization normally generates single EcircRNAs or intronic circular RNAs (IcicRNAs), which involves the base-pairing of splicing sites motif sequences of exonic and intronic sequences, respectively.

These characteristics are important features to be considered in the prediction of circRNAs, and different approaches are being developed based on evaluation of the mapping structure of RNA sequencing reads against a reference genome or even through machine learning approaches considering characteristics present in the sequence of a protein-coding or non-coding transcript.

Algorithms and computational tools for circRNA prediction

RNA-seq and bioinformatic analysis provide a comprehensive understanding of the eukaryotic transcriptome and a deep exploration of the various types of RNA molecules available in the cell. For circRNA analysis, RNA-seq of total RNA with ribosomal RNA depletion is widely used, often followed by an RNase R step - which efficiently removes linear RNAs - to increase the necessary depth of sequencing. Despite the effectiveness of this approach, results can vary depending on circRNA expression levels, sample quality and sequencing quality [21, 22].

The software normally developed to explore RNA-seq data can be classified into two main categories, associated with the mapping of sequencing reads against a reference genome and/or gene annotation file: split-alignment-based (segmented-based) and pseudo-reference-based (candidate-based) [28]. Segmented-based tools identify BSJ by aligning reads to the reference genome and are considered *de novo* tools, whereas pseudo-reference tools rely on gene annotation files to identify junction reads (Figure 2). To give more details on how these algorithms are implemented in predictive tools, we chose to briefly describe three of them: CIRCEplorer3 [29], CIRIquant [30] and find_circ [31].

CIRCEplorer3 [30] requires trimmed paired-end FASTQ files from RNA-seq samples, a genome reference and annotation files as input. It first uses HISAT2 [32], a versatile aligner, to map reads to the genome reference, with a focus on finding reads spanning splice junctions in coding transcripts. Unmapped reads are then remapped to the same reference genome. The use of STAR [33], a fast and user-friendly aligner, with an annotation of the reference genome allows the identification of candidate circRNA reads. The precise start and end positions (donor and acceptor sites) of each BSJ event are then determined, and the BSJ is remapped to the annotation file. The final output file contains circRNA annotation information such as the circRNA ID, chromosome, start and end coordinates, strand, read number part of the circRNA, flanking introns and other features interesting for downstream analyses.

CIRIquant [30], the latest version of CIRI2, aligns reads using HISAT2 [32] as the first step. It then makes a *de novo* identification of circRNAs from unmapped reads, which are mapped to the reference genome using BWA [34]. Then, the SAM files are scanned twice. During the first scanning, paired chiasmic clipping

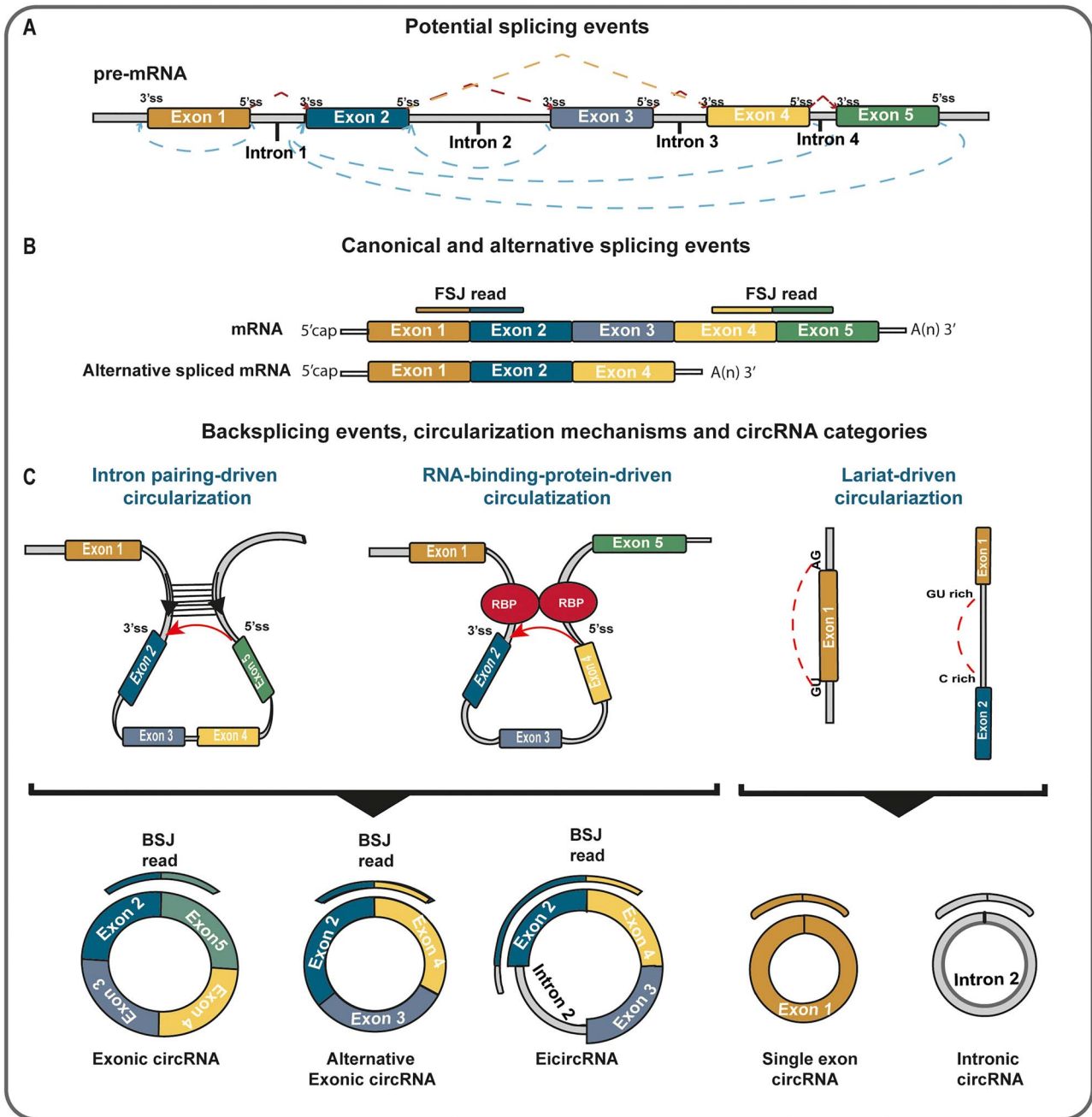


Figure 1. General back-splicing mechanism of circRNAs. (A) Transcription mediated by RNA polymerase II gives rise to a pre-mRNA that is subjected to canonical or alternative splicing, becoming mature mRNA accessible to protein translation and detectable as Forward splice junction reads (FSJ reads). (B) Intron-free mature mRNA have special well-defined characteristics, including defined 5' - and 3' sites, capping and poly(A) tail. (C) In circRNAs, alternative splicing is replaced by back-splicing reaction generating a diversity of circRNAs through different mechanisms of circularization such as: Intron pairing-driven circularization facilitated by proximity of donor (5') and acceptor (3') sites by inverse complementary sequences near to BSJ splice site, RNA-binding-protein-driven circularization, mediated by proteins that forms dimers allowing to circularize RNA sequence, and lariat-driven circularization facilitated by specific motifs (GU-AG; GC rich zone and others) on flanking sequences. In this manner, different factors bring about a diversity of circRNAs.

(PCC) signals are used to detect BSJ and candidate circRNAs; which are then filtered using paired-end mapping (PEM) and splicing junction signals (GT-AG). After grouping BSJ reads and register circRNA candidates, the SAM file is scanned again to detect additional circRNA candidates, and an additional filter is applied to eliminate false positives that are incorrectly mapped to homologous genes and repetitive sequences. The final output file also contains circRNA annotation information such as circRNA ID, chromosome, start and end coordinates, strand, read number, flanking introns and others.

Find_circ [31] is a *de novo* detection algorithm for circRNAs prediction. The input files need to be trimmed to the best possible Phred Quality Score. The first step involves alignment to the reference genome using Bowtie2 [35], with SAM format files resulting from the unmapped reads stored. The second step involves the detection of splicing by extracting 20-mers from both ends of the unmapped reads and aligning them to the reference genome to find unique anchor positions. The output file contains all detected splice junctions in BED format, along with additional columns of information and statistics. In 2018, Hansen [36] published a

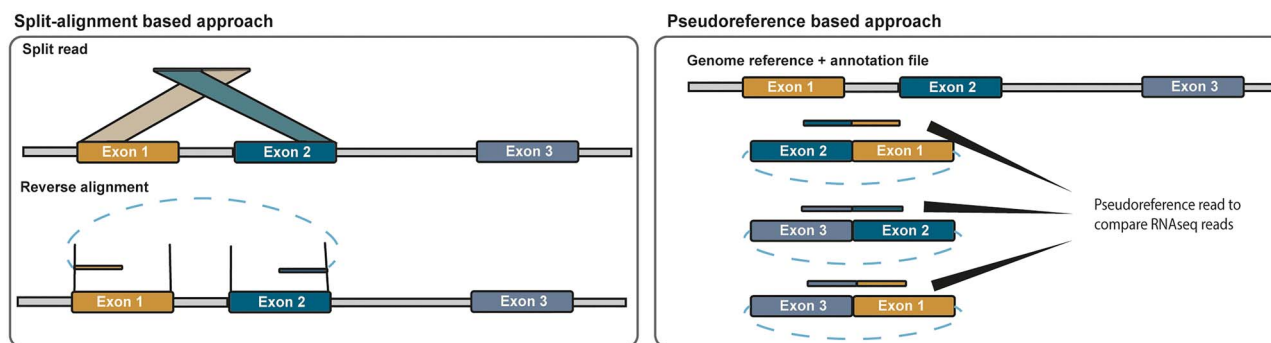


Figure 2. General strategies for circRNA detection. In the split-alignment-based approach, circRNA sequencing reads are split and then reverse aligned to reference genome. In the pseudo reference-based approach, pseudo-sequences are constructed from annotated exons and sequence reads are aligned against the pseudo-sequences.

slight improvement to the find_circ pipeline, which increased the default mapping quality threshold, resulting in a decrease in false discovery rate (FDR).

Over the last decade, the interest of the scientific community on circRNAs has grown, and it is reflecting in an increase in more than 90% in the number of publications related to this molecule in literature (Figure 3A). This interest is also reflected in the numerous tools for circRNAs identification, which together with other bioinformatic tools and databases is helping on the further characterization of this new trend in RNA biology research (Figure 3B). Despite the various algorithms and tools developed by bioinformaticians for circRNA analysis, few systematic evaluations of its performance have been carried out [21, 36, 37]. A study by Hansen et al. (2016) [17] showed that five different tools (circRNA_finder [38], MapSplice [39], CIRCexplorer [40], CIRI [41], find_circ [31]) had only a modest overlap of 16.8% in their predictions, highlighting that the circRNA landscape can differ dramatically depending on the algorithm used. It is worth mentioning that 40% of the predicted circRNAs were obtained by only one tool. To evaluate the true positive candidates on this tool-specific group, they treated samples with RNase R and evaluated their performance, concluding that apart from CIRCexplorer [40] and MapSplice [39], more than half of the potential circRNAs were RNase R sensitive candidates. Thus, revealing to be false positives, suggesting that in most of the cases, the RNAs predicted by several algorithms were likely to be artifacts. To avoid missing potential circRNAs, as well as the identification of false positive artifacts, it is recommended to use more than one algorithm.

Sensitivity and precision are important parameters to evaluate bioinformatic tools performance. In these software, sensitivity refers to the number of reads assigned to each circRNA by the algorithm and is reflected in the total number of circRNAs predicted. On the other hand, precision refers to the accuracy of the algorithm and can be increased by filtering for circRNA candidates by implementing thresholds for read counts. In 2017, Chen and colleagues [37] performed a benchmarking of 11 circRNA detection tools, identifying that NCLScan [42], MapSplice [39], CIRCexplorer [40], DCC [43] and PTESFinder [44] tended to have a low false-positive rates, whereas Segemehl [45], find_circ [31] and UROBORUS [46] yielded the worst performances. In general, NCLScan [42] and CIRCexplorer [40] dominated other tools regarding the precision measuring, whereas KNIFE [47], CIRI [41], Segemehl [45], PTESFinder [44] and CIRCexplorer [40] tended to be more sensitive than other tools. Considering the sensitivity analysis for paired-end data, when both reads span the same

BSJ—which can be the case for small circRNAs, different tools undertake distinct results, evidencing an enormous variation of distances between splicing sites engaged in back splicing regions, ranging typically from a few 100 nucleotides to dozens or even hundreds of kilobases.

Later, in 2018, Hansen [36] evaluated the impact of pre-processed reads on algorithm performance, including low-quality read removal and adaptor trimming. In his benchmarking, some algorithms benefited more than others from read trimming, and the sensitivity measured by read count increased modestly, although most algorithms showed unchanged median expression. Most approaches presented a median expression of 14–20 reads, with DCC [43], circRNA_finder [38] and UROBORUS [46] showing the lowest sensitivity with 11, 9 and 5 median reads per circRNA, respectively. It always depends on the initial quality of the RNA sequencing dataset. Recently, Circall [48] has been developed as a tool to specifically identify EcircRNAs from paired-end RNA-Seq data. It boasts improved true positive detection in comparison to CIRI2 [49], CIRCexplorer [40], MapSplice [39] and find_circ [31]. Furthermore, it is faster in analysis, thanks to its implementation of the RapMap aligner [50], which employs a quasi-mapping strategy.

In addition to the two primary methods for identifying circRNAs, Chen et al. (2021) [51] published a review describing machine learning approaches for circRNA prediction. PredcircRNA [52] and WebCircRNA [53] extract features from transcript sequences to predict the formation of circRNAs, whereas DeepCircCode employs a deep learning model to predict back-splicing for the formation of human circRNAs. This tool also incorporates a visualization method to represent circRNA formation features as sequence motifs. CirRNAPL [54] predicts circRNAs based on structural features and composition of sequences, and the recently released StackCirRNAPred [55] classifies circRNAs from other lncRNAs using a stacking strategy. These tools can complement established split-bases and pseudo reference-based identification methods available in the literature. A complete list of circRNA predictive tools was made available in Table 1.

Before start using and testing these tools, it is also important to consider those that are easy and less time-consuming to install and configure. To facilitate the configuration of the proper environment to run a circRNA predictive analysis we recommend to use tools available in Conda environment (<https://conda.io/>), which is an open-source package management system to facilitate the installation and configuration of software packages written in any language that runs in multi operation systems such as Windows, macOS and Linux.

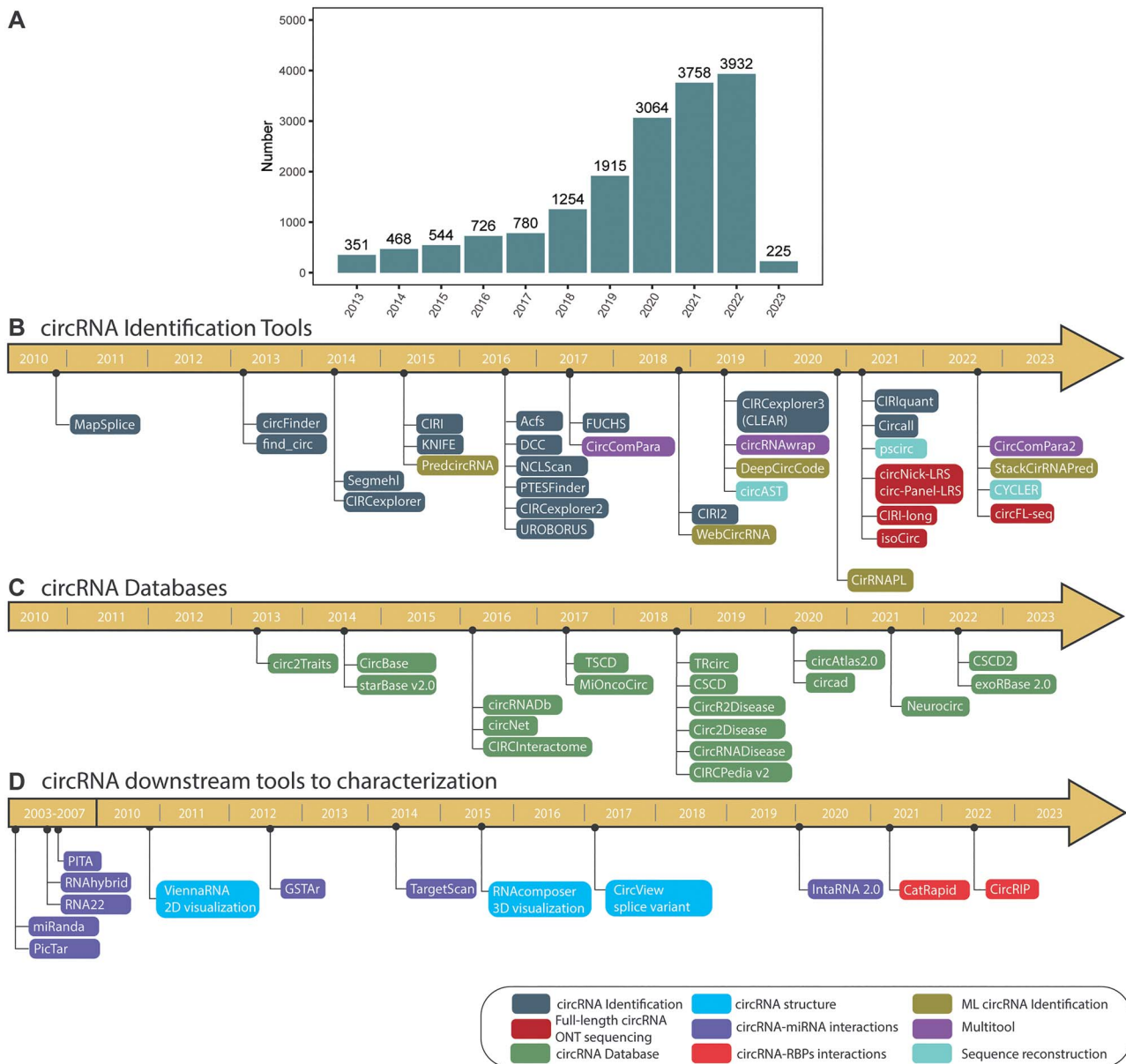


Figure 3. Main circRNA tools in the circRNA research along the last decade. (A) Representation of the number of publications on circRNA investigation containing the term ‘circRNA’ in the title or abstract searched in PubMed by year. (B) Summary of circRNA identification tools, multi-tools software and sequence reconstruction tools; (C) circRNA publicly available databases; and (D) other useful tools for characterization, visualization and functional exploration of circRNAs developed in the last decade.

Multi-tool approaches for circRNA prediction

Multi-tool approaches for circRNA prediction can address the issue of differences in predicted circRNA numbers and high false positives caused by the difference in algorithm methodologies used by different tools. CircComPara [64] integrates CIRCexplorer [40], CIRI [41] and find_circ [31] to improve the accuracy of circRNA prediction, and its recent version, CircComPara2 [57], integrates CIRI [41], find_circ [31], CIRCexplorer2 [56] (using multiple aligners such as STAR [33], BWA [34], Segemehl [45] or TopHat [65]), further enhancing the identification of true circRNAs. CircRNAwrap [58] is another multi-tool approach that integrates eight circRNA detection tools (KNIFE [47], find_circ [31], CIRCexplorer2 [56], CIRCexplorer [40], MapSplice [39], Acfs [66], circRNA_finder [38] and DCC [43]), three sequence reconstruction approaches (RAISE [67], CIRI-AS [68] and CIRCexplorer2

[56]), and several downstream tools (e.g. Sailfish-cir [69]) for expression abundance estimation and other types of computational characterization.

Exploring alternative splicing events in the reconstruction of circRNAs

Several computational tools are used to predict circRNAs by looking for BSJ events. However, the internal composition of circRNAs splicing can be diverse and different from the host linear transcripts. The variety of alternative splicing variants requires the comprehensive analysis of circular transcripts. Therefore, to identify circRNA isoforms, the whole coverage of RNA sequencing reads throughout the circular transcripts is necessary. Two major strategies for identifying inner spliced isoforms of circRNAs are CIRCexplorer3 [29] and CIRI-AS [68], part of CIRI package [41, 49].

CIRCexplorer3 [29] uses information on splice-site coverage and depth based on polyA+ and polyA-/RNase R+ reads, whereas CIRI-AS utilizes paired mate reads of back-spliced reads (BSR) to predict forward splice junctions. This last tool categorizes alternative splice sites into: alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), exon skipping (ES) or intron retention (IR). However, to run CIRI-AS [68], the paired-end reads must have the same length. It is worth mentioning that CIRI2 [49] can process sequencing reads of different lengths, but CIRI-AS [68] cannot. It is recommended to use raw reads directly or trim all reads to the same length before running BWA [34], CIRI2 [49] and CIRI-AS [68].

Other useful computational methods for the full-length reconstruction and alternative spliced circRNA abundance determination from RNA-seq data include circAST [60], psirc [61] and CYCLER [62]. circAST [60] is a downstream analysis tool that employs a multiple splice graph method on RNA-seq data and uses upstream circRNA identification tools such as UROBORUS [46], CIRI2 [49] or CIRCexplorer2 [56] to detect back-spliced events. On the other hand, psirc [61] can identify full-length circRNA isoforms and quantify their expression levels from RNA sequencing with results comparable to CIRCexplorer2 [56] and CIRI2 [49]. Recently, the software CYCLER [62] has been developed to identify and reconstruct the predominant isoform of a circRNA with better performance than CIRCexplorer2 [56] and CIRI-full [70].

Long-read sequencing technologies for the estimation of full-length circRNAs and splicing variants

Ruan et al. (2022) [71] presented a comprehensive overview of different experimental strategies for full-length sequencing splicing variants detection of circRNAs. The authors summarize various library preparation techniques that enrich the sample for circRNAs and pipeline analysis using Oxford Nanopore Technology (ONT) sequencing. These techniques include circNick-LRS [72], circ-Panel-LRS [72], CIRI-long [73], isoCirc [74] and circFL-seq [75]; and their strategies differ in terms of sample preparation and amplification approaches, length of circRNAs obtained and detection of specific features.

The circNick-LRS and circ-Panel-LRS [72] approaches sequence full-length circRNAs through sample enrichment and specific amplification of circRNA transcripts, respectively, followed by ONT sequencing [72]. The circFL-seq [75] and CIRI-long [73] methods use the rolling circle reverse transcription (RCRT) strategy to produce cDNA and detect circRNAs. The circFL-seq assay uses an anchor primer for second-strand synthesis, whereas CIRI-long employs template switching. The isoCirc method [74] utilizes rolling circle amplification (RCA) to enrich BSJ reads, resulting in the detection of circRNAs with sizes ranging from 300–460 nucleotides. circRNA detection using these methods differs in terms of the length of circRNAs obtained. For example, the circNick-LRS [72] approach detects full-length circRNAs with an average size of 795 nt, whereas CIRI-long detects circRNAs less than 500 nt. The circFL-seq approach enables the identification and reconstruction of longer circRNAs up to 2334 nt, and the detection of fusion circRNAs, allowing for isoform-level studies. These approaches also allow for the detection of other specific circRNA sequence features such as microexons inclusion, which are important regulators of the transcriptome [72].

An alternative to explore circRNA splicing variants is to use their identifiers in public databases and look for the mature sequence of circRNA isoforms in repositories like the circAtlas [76], which were reconstructed using the CIRI-full/CIRI-vis pipeline. Unlike linear RNAs, circRNAs have complex and

diverse splicing patterns, with many isoforms from a single circRNA. CircView is a useful tool for visualizing circRNA isoforms obtained from various identification tools [77]. High-throughput techniques such as Illumina-based RNA-seq, microarray [40] and NanoString [78] have been used to profile circRNA expression by detecting and counting unique BSJ sequences, but none can detect the full structure of circRNAs >300 nucleotides or determine their exon composition. Alternatively, ONT platform provides comprehensive characterization of alternative splicing events in circRNAs on a genome-wide scale and has the capacity to study sequence modifications in the RNAs, such as the N6-Methyladenosine (m6A), the most abundant internal modification associated with eukaryotic mRNAs [14]. These epitranscriptomic modifications are detected through the direct sequencing of RNA molecules without the necessity of generating a cDNA.

The structural confirmation is also relevant when studying the impact of full-length circRNA isoforms. Secondary structure and thermodynamic parameters can be obtained using the ViennaRNA Package 2.0 [79], considering a particular parameter devoted specifically for circRNAs ('RNAfold -p -d2 -circ'). Other key tools have been proposed to analyze the superior structures of circRNAs. These include RNAcomposer [80], which can predict the three-dimensional structure of RNA and can be applied to circRNA sequences (<http://rnacomposer.cs.put.poznan.pl/>), as well as RNAfold from the Vienna package [79] and 3dRNA [81]. These tools can provide new insights into circRNA structures and interactions, offering a deeper understanding of this complex molecule. The use of these tools may reveal new information about circRNA interactions at previously uncharacterized levels.

In summary, transitioning from next-generation sequencing (NGS) to third-generation sequencing (TGS) is a viable and effective option to move forward circRNA characterization, as TGS can sequence full-length reads that provide more information in downstream analyses on the complete circRNA sequence, including the splicing variants, characterization of interaction sites, modifications and superior structures. However, TGS is currently more expensive and has a higher error rate than NGS, but novel improved tools are starting to be released to overcome these limitations.

Quantification, normalization and differential expression of circRNAs

Once the repertoire of circRNAs available in a transcriptome assay is defined, the next step is to obtain their expression patterns. Quantifying, normalizing and differentiating the expression of circRNAs remains a challenge in transcriptome analysis. CircRNAs, being rare in nature and often displaying low read counts, pose difficulties in obtaining their expression patterns. This is because a small fraction of the total reads from RNA-seq samples support the detection of BSJ, leading to difficulties in differentiating reads from the inner circRNA sequence and the linear form of the transcript. As a result, the number of total sequenced reads can impact the number of identified circRNAs. To mitigate this challenge, researchers have used the previously mentioned RNase R treatment during library preparation, which enriches the sample with circRNAs and focuses the sequencing procedure on circRNA sequences.

Before conducting differential expression analysis, many authors apply a filter to reduce false positives and increase precision of circRNAs identification. For instance, when generating an expression matrix containing each circRNA expression values to be used in a downstream differential expression comparison, studies normally consider only those circRNAs containing read

counts greater than 2; or to normalize this expression matrix prior its filtering considering mapped back-splice junction reads per million (RPM) or counts per million (CPM) normalized values greater than 0.1 [49].

Different challenges, such as the number of total read sequences, low read counts for each circRNA, high variability between samples and a lack of coverage modeling, can affect differential expression analysis. Some commonly used software for mRNA differential expression, such as DESeq2 [82] and edgeR [83] can be applied to perform differentially expression analysis for circRNAs. CircTest [43] is another approach that can also be employed. It models the data using the beta binomial distribution and performs an ANOVA to identify circRNAs that differ in their relative expression between groups. Other authors use log2(CPM) representation and perform a Mann-Whitney U test to differentiate the expression values of circRNAs within different groups of samples.

Functional prediction of the biological context of circRNAs

CircRNAs have been described to exert their effects through several mechanisms [84]. One of the main mechanisms is their ability to modulate gene expression by acting as a miRNA and protein regulator [2, 85]. They act as a sponge for miRNAs, 'capturing' them through microRNA responsive element (MRE) sites, thereby reducing miRNA availability and influencing mRNA expression levels. Additionally, circRNAs have the potential to modulate alternative splicing or transcription by controlling protein expression and avoiding mRNA transcription. Another mechanism involves circRNAs regulating the parent genes by interacting with RNA pol II, altering its affinity for adjuvants and influencing gene transcription [1]. Finally, some circRNAs have been found to contain open reading frames (ORF) or internal ribosome entry sites (IRES), which have the potential to encode proteins [86, 87]. This is a relatively unexplored area that could have a significant impact at the cellular level.

Computational approaches can be used to predict the interaction between circRNAs and miRNAs or RNA-binding proteins. These predictions can be used to generate interaction networks, by visualizing the results using tools such as Cytoscape [86] or R packages like VizNetwork (<https://datastorm-open.github.io/visNetwork/>) or networkD3 (<http://christophergandrud.github.io/networkD3/>), and integrating data from public protein-protein interaction databases such as STRING [88], IntAct [89] or BIND [90]. Enrichment analysis can also be performed through tools such as enrichR [91], PANTHER [92] or DAVID [93], to identify overrepresented biological pathways and processes according to the coding genes present in the generated network.

CircRNA-miRNA interaction tools

Multiple tools are used to analyze circRNA-miRNA interactions, contributing in the prediction of gene expression patterns on which this interaction could have some kind of effect, favoring its degradation, or changing expression patterns of miRNA and mRNAs through competing endogenous mechanisms. Among them, we can highlight RNAhybrid [94], GStar [95], miRTarget [96], TargetScan [97], miRanda [98], PicTar [99], PITA [100], IntaRNA [101], RNA22 [102] and others. These tools have the potential to identify possible MRE sites within circRNAs. In this way, Riffó-Campos et al. 2016 [103] evaluated the possibility of interaction through numerous parameters depending on the tool used. RNAhybrid [94], GStar [95], miRTarget [96] perform this analysis by calculating the minimum free energy hybridization between

sequences, looking for the most favorable interaction. TargetScan [97] is the tool with better performance, but with a high false negative rate, whereas miRanda [98] has a better sensitivity but with a higher false positive rate. It runs an algorithm based on complementarity, binding energy and taxonomic conservation of the predicted interaction; PicTar aligns seed regions given their high conservation, allowing predicted interactions, and classifies it taxonomically. Finally, PITA [100] calculates the differences between free energies in pairing and unpairing the target binding site to predict the structural accessibility of the seed matching.

CircRNA-protein interaction tools

Main circRNAs and RBP (RNA-binding protein) interactions are involved in the regulation of gene transcription and translation of peptides. Although RBPs can bind both linear RNAs and circRNAs, evidence suggests a preference for RBP-circRNA binding due to the tertiary structures of circRNAs [104]. RBPs play a role in the modulation and control of circRNA biogenesis [105], whereas, at the same time, circRNAs can serve as 'sponges' for RBPs, altering the translation of other proteins [106, 107].

Protein-circRNA interactions are primarily studied through experimental techniques, and specific circRNA tools are limited. Most bioinformatics tools also do not consider the tertiary structure of circRNAs when determining protein-circRNA interactions [27]. A notable online tool is catRAPIDv2.0 (http://service.tartagialab.com/page/catrapid_omics2_group) [108], which evaluates the propensity for circRNA-RBP interactions by analyzing the circRNA sequence with a fragmentation approach against precompiled RBP libraries from several organisms. Advances in this field mainly focus on either predicting the circRNA-binding sites on the protein, or conversely, the protein-binding sites along the circRNA sequence. In the first case, CircInteractome [109] provides some tools, but it does not consider the full-length sequence of circRNAs. Conversely, there are approaches based on neural networking [110, 111], in which the interaction matrices of known protein-circRNA pairs are used to train a neural network to generate improved prediction models. CRIP [110, 111] is a recent deep learning method that uses known sequences and is based on codon splitting to predict interactions. Another novel method, iCircRBP-DHN [112], uses deep hierarchical networks for circRNA-RBP binding site discrimination. circRIP [113], another recently developed algorithm, is an efficient tool for identifying genome-wide circRNA-RBP interactions from RNA immunoprecipitation sequencing (RIP-Seq) and enhanced cross-linking immunoprecipitation (eCLIP) data, optimizing the quantification process.

Experimental procedures for validation

The validation of circRNA presence and functionality is crucial in the study of these non-coding RNAs. RT-qPCR is a sensitive method for detecting the expression levels of circRNAs and comparing them to those of linear RNAs. It can be used to quantify the expression levels in different samples, such as normal and diseased tissues, and to monitor changes in response to various stimuli. RT-qPCR of RNase R-treated samples can also be used to validate the detection of circRNAs. The technique requires divergent primers that target the BSJ, a unique sequence within the human genome [114]. RNase R is a 3'-to-5' exonuclease that degrades linear RNAs, but most circRNAs are resistant to degradation due to their closed circular configuration. To confirm the circular configuration, Northern blot analysis, an RNase R-independent technique, is considered a gold standard. Probes are designed to either span the BSJ or hybridize to the total transcript,

Table 1. Different tools for identifying circRNAs and associated methodology.

Identification Tools	Method	Aligner	Aligner strategy	Link	Reference
CIRCexplorer3 (CLEAR)	S.A.B	HISAT2;STAR	Splice-aware	https://github.com/YangLab/CLEAR	[29]
CIRCexplorer2	S.A.B	STAR	Splice-aware	https://github.com/YangLab/CIRCexplorer2	[56]
CIRIquant	S.A.B	HISAT2;BWA	Versatile	https://github.com/bioinfo-biols/CIRIquant	[30]
CIRI2	S.A.B	BWA	Versatile	https://sourceforge.net/projects/ciri/files/CIRI2/ https://ciri-cookbook.readthedocs.io/en/latest/CIRI-full.html#installation	[49]
Find_circ	S.A.B	Bowtie2	Versatile	https://github.com/marvin-jens/find_circ	[31]
KNIFE	P.R.B	Bowtie2	Versatile	https://github.com/lindaszabo/KNIFE	[47]
DCC	S.A.B	STAR	Splice-aware	https://github.com/dieterich-lab/DCC	[43]
UROBOROS	S.A.B	Bowtie	Versatile	https://github.com/WGLab/UROBOROS http://uroborus.openbioinformatics.org	[46]
NCLScan	P.R.B	BWA	Versatile	https://github.com/TreesLab/NCLscan	[42]
MapSplice	S.A.B	Bowtie	Versatile	https://github.com/ahcarpenter/mapsplice http://www.netlab.uky.edu/p/bioinfo/MapSplice2	[39]
PTESFinder	P.R.B	Bowtie	Versatile	https://github.com/osagiei/pfv2	[44]
Segemehl	S.A.B	Internal	Covariance Models	https://www.bioinf.uni-leipzig.de/Software/segemehl/	[45]
CircCall	P.R.B	RapMap	Quasi-mapping	https://github.com/datngu/Circall	[48]
CircComPara2	S.A.B and R.R.B	Multi-tool	Multi-tool	https://github.com/egaffo/CirComPara2	[57]
circRNAwrap	S.A.B and R.R.B	Multi-tool	Multi-tool	https://github.com/liaoscience/circRNAwrap	[58]
PredcircRNA	M.L	NA	NA	https://github.com/xypan1232/PredcircRNA	[52]
WebCircRNA	ML	NA	NA	https://rth.dk/resources/webcircrna/download	[53]
DeepCirCode	M.L	NA	NA	https://github.com/BioDataLearning/DeepCirCode	[59]
CirRNAPL	M.L	NA	NA	http://server.malab.cn/CirRNAPL/	[54]
StackCirRNAPred	M.L	NA	NA	https://github.com/xwang1427/StackCirRNAPred	[55]
circAST	S.R	-	-	https://github.com/xiaofengsong/CircAST	[60]
psirc	S.R	-	-	https://github.com/Christina-hshi/psirc	[61]
CYCLEr	S.R	-	-	https://github.com/stiv1n/CYCLEr	[62]
CIRI-AS	S.R	-	-	https://sourceforge.net/projects/ciri/files/CIRI-AS/	[63]

Method column abbreviations: S.A.B: split-alignment-based; P.R.B: pseudo reference-based; M.L: machine Learning. S.R.: sequence reconstruction.

and by choosing a suitable gel electrophoresis system (agarose and/or polyacrylamide) the two possible configurations (circular or linear) can be clearly distinguished. Northern blotting can also be combined with RNase R or RNase H ribonuclease digestion. The RNase H cleavage assay has emerged as an elegant method for circRNA validation. Antisense oligonucleotides are designed to target a specific circRNA, and RNase H recognizes DNA–RNA hybrids and cuts the RNA within the duplex. The characteristic shift in mobility during denaturing polyacrylamide gel electrophoresis from the aberrantly slow migration of the circRNA to the expected linear behavior, as well as the cleavage patterns of RNase H digestion, can then be analyzed by Northern blotting [115].

Publicly available repositories for circRNA research

circRNAs are attracting increasing attention from the scientific community and, as a result, a plethora of databases have been established to store extensive information on these RNA molecules (Figure 3C). These repositories provide information such as circRNA sequences, functional predictions, the circRNA–miRNA–mRNA axis and links to diseases [2, 116]. Vromman and colleagues (2020) [18] reviewed these databases and categorized them as curated and non-curated.

In our review, we will address some considerations one should consider when using a repository as a reference. For instance, not all databases use the same genomic coordinates (chr:start-end), which can result from differences in the input file formats used by the databases. BED format-based databases

are 0-based, whereas SAM format-based databases are 1-based, thus requiring adjustments when comparing data between databases. Additionally, some repositories may not provide complete information on their data, such as RNA-seq library type (e.g. single-end, paired-end, stranded, RNase R enriched), primary nucleotide sequence of the mature circRNA, strand of origin (e.g. positive, negative), validation methods (e.g. qPCR, long-read sequencing, Northern blot, etc.). This information is crucial because the presence of an exon junction does not necessarily indicate RNA circularization. For example, RNA-seq poly (A) + decreases sensitivity to detect circRNAs as in general these molecules are not polyadenylated, increasing the quantity of linear RNAs. Nevertheless, studies have shown that even RNA-seq polyadenylated libraries can detect quantifiable amounts of exon junctions, suggesting that their presence alone is not enough to ensure circularization [12, 117]. Another example is related to the presence of Alu sequences in the 3'- and 5'-regions of intronic segments [9, 118], or sequence motifs in the flanking regions of the splicing candidates [117]. All of them are also important features to be considered in the determination of the RNA circularization, and the lack of information makes it difficult to validate and curate the candidates available in these repositories. It is worth mentioning that this kind of mechanism remains poorly understood and it is fundamental to have as many details as possible in the metadata indexed in databases.

Circ2Traits [119] was the first database to provide information on circRNAs, specifically manually curated circRNA–miRNA–disease associations in human datasets. CircBase [120] followed as the second repository published, providing information on

Table 2. Publicly available databases devoted to circRNAs

Database	Type of Data	Organism	Link	Reference
circAtlas	Tissue-specific circRNAs Evolutionary conservation Regulatory network Full-length sequence of circRNAs	<i>Gallus gallus</i> , <i>Homo sapiens</i> , <i>Macaca mulatta</i> , <i>Mus musculus</i> , <i>Rattus norvegicus</i> , <i>Sus scrofa</i> ,	http://159.226.67.237:8080/new/index.php	[76]
circBase	circRNA expression level Sequence	<i>C. elegans</i> , <i>Drosophila melanogaster</i> , <i>H. sapiens</i> , <i>Latimeria chalumnae</i> , <i>M. musculus</i>	http://www.circbase.org	[120]
CIRCpedia v2	circRNAs expression in cell types and tissues from different species	<i>Caenorhabditis elegans</i> , <i>Danio rerio</i> , <i>D. melanogaster</i> , <i>H. sapiens</i> , <i>M. musculus</i> , <i>R. norvegicus</i> ,	http://yang-laboratory.com/circpedia/	[127]
MiOncoCirc	circRNA associated to cancer	<i>H. sapiens</i>	https://mioncocirc.github.io	[121]
CIRCinteractome	Identification of MRE sites Divergent primers and siRNA design	<i>H. sapiens</i>	https://circinteractome.nia.nih.gov	[109]
StarBase v2.0/ENCORI	RNA–RNA interactions	23 species	https://starbase.sysu.edu.cn	[128]
CircRic	circRNAs associated to cancer	<i>H. sapiens</i>	https://hanlab.uth.edu/cRic/	[122]
Circ2Disease	circRNA associated to diseases	<i>H. sapiens</i>	http://bioinformatics.zju.edu.cn/Circ2Disease/index.html	[125]
CircR2Disease	circRNA associated to diseases	<i>M. musculus</i> , <i>R. norvegicus</i> , <i>H. sapiens</i>	https://bio.tools/circR2Disease	[129]
circBank	circRNA interactions	<i>H. sapiens</i>	http://www.circbank.cn	[130]
Circ2Traits	circRNA–miRNA–disease associations	<i>H. sapiens</i>	http://gyanxet-beta.com/circdb/	[119]
CircNet	Tissue-specific circRNA expression patterns circRNA–miRNA–mRNA axis	<i>H. sapiens</i>	https://awi.cuhk.edu.cn/&#x007E;CircNet/php/index.php	[131, 132]
circRNADb	Exonic circRNAs Protein coding–circRNAs annotation	<i>H. sapiens</i>	http://reprod.njmu.edu.cn/cgi-bin/circrnadb/circRNADb.php	[133]
CSCD2	circRNAs in cancer miRNA–circRNA interaction RBP–circRNA interaction ORF in circRNAs	<i>H. sapiens</i>	http://geneyun.net/CSCD2/	[134]
TSCD	Tissue-specific circRNA expression circRNAs in organogenesis	<i>M. musculus</i> , <i>H. sapiens</i>	http://gb.whu.edu.cn/TSCD	[135]
circad	circRNAs associated with diseases	<i>M. musculus</i> , <i>H. sapiens</i> , <i>R. norvegicus</i> , <i>S. scrofa</i>	Rophina et al., 2020 https://clingen.igib.res.in/circad/	[123]
TRcirc	Transcriptional regulation of circRNAs	<i>H. sapiens</i>	http://www.licpathway.net/TRCirc/view/index	[136]
circRNADisease	circRNA associated to diseases	<i>H. sapiens</i>	http://cgga.org.cn:9091/circRNADisease/	[124]
exoRBase 2.0	Extravesicular long RNAs	<i>H. sapiens</i>	http://www.exorbase.org	[137]
NeuroCirc	circRNAs in human brain	<i>H. sapiens</i>	https://voineagulab.github.io/NeuroCirc/	[138]

expression levels and sequence information of circRNAs from human, fly, mouse, coelacanth and *Caenorhabditis elegans*. It has been widely used as a reference in annotation pipelines. CircAtlas [76], another widely used database, relies on the integration of four tools (CIRI2 [49], CIRCexplorer2 [56], DCC [43] and find_circ [31]) to annotate data from human, macaque, mouse, rat, pig and chicken and provides information on tissue-specificity, evolutionary conservation, regulatory networks, disease associations and full-length circRNA sequences. CircAtlas [76] MiOncoCirc [121] and CircRic [122] are repositories focused on cancer-related circRNAs, analyzing biogenesis, transcriptional landscape, integrative analysis and drug response (exclusive to CircRic). Other disease-related databases contain data on various pathologies, such as cardiovascular infirmities, diabetes, immune disorders, etc. Main resources in this category are those with manual curation, such as circad [123], circRNADisease [124], Circ2Traits [119] and Circ2Disease [125].

Each of the publicly available databases (Table 2) has unique features that support circRNA research [2]. However, one of the key challenges within these resources is the lack of a universal naming convention. For instance, some databases use ‘circ_+ host gene + variant’ or ‘circ_+ ID’ when naming the RNAs, which makes it difficult to compare data between databases for a specific circRNA. To address this issue, Chen et al. (2023) [126] proposed a nomenclature for reporting circRNAs to repositories. This involves naming every circRNA with a ‘circ’ prefix, followed by the parental gene name and the number of the exons, e.g. ‘circGENEY(1,3,5)’ for a host gene of name GENEY in which the exons 1, 3 and 5 conformed the circRNA. If an intronic region is retained, the suggested nomenclature includes an ‘RI’ indicating it, such as ‘ciercGENEY(1,RI,3,5)’. This will help to standardize identification and annotation, as well as enable the sharing and cross-reference of information between databases. It is important for researchers to adopt this nomenclature

and publish their data in a format that is consistent with this standard.

In summary, the various circRNA databases have unique characteristics that aid in the study of these RNA molecules. However, there are also challenges that need to be addressed, such as the lack of a universal naming convention, the incompleteness of data in some databases, and differences in genomic coordinates used by different databases. Efforts should be directed toward standardizing these databases to enable more comprehensive and accurate analysis of circRNAs.

CONCLUDING REMARKS

CircRNAs are a growing class of non-coding RNAs that are increasingly being recognized for their role in cellular processes. Despite the availability of various bioinformatic tools to identify and quantify circRNA expression, the lack of a uniform nomenclature system and the diverse expression patterns across different species, genes, tissues and developmental stages still present challenges in circRNA research. The mechanisms of circRNA action, including sponge effect, interaction and peptide coding, are the focus of ongoing studies. However, the limitations of current bioinformatic approaches hinder a comprehensive analysis of circRNAs, such as the inability to accurately quantify expression levels based on the full circRNA sequence and to consider the mature isoform sequences derived from the same BSJ. Direct RNA sequencing or TGS techniques can address these issues and improve the prediction of miRNA and RBP interactions. Most algorithms for predicting circRNA-protein interactions are based on general properties of non-coding RNAs and overlook the tertiary structures of circRNAs, which can greatly influence these interactions. Although some neural network-based algorithms consider these structures, they require a large amount of training data that is currently limited. To advance circRNA research, the development of specific algorithms and better integration of data is necessary. Additionally, a common nomenclature system is crucial for comparing circRNA expression across different conditions, diseases or tissues in standardized ways within different publicly available databases.

It is worth noting that other publications in the literature have reviewed different aspects of the computational tools for circRNA identification and its in silico functional characterization (see [51, 71, 139]). However, unlike other reviews, our work offers a more accessible description of the most popular tools for circRNA identification, including technical aspects, as well as a compilation of the most popular tools for its in silico characterization in order to obtain functional insights of their biological roles. We aimed to cover various aspects of the entire pipeline for circRNA characterization, starting from the reconstruction of the nucleotide sequence to the prediction of interactions with miRNAs and RBPs, and the obtention of potential ORFs generated by that transcript. Additionally, we covered other aspects related to the study of secondary and tertiary structures associated with the predicted circRNAs, and also emphasized the importance of using new sequencing technologies and the preparation of different types of sequencing libraries to obtain the complete sequence and other characteristics of circRNAs, followed by their validation through molecular biology and biochemical experiments, which allow us to characterize their subcellular location, the integrity of their sequence and their interaction with predicted cellular elements. Our aim is to provide an integrative and multidisciplinary vision to new researchers interested in delving into the biology and bioinformatics of circRNAs.

Key Points

- Bioinformatic tools for circRNA prediction and databases should be implemented in an integrative way, to facilitate experimental circRNA-related data interpretation.
- circRNA identification experiments from RNAseq expression data can be evaluated by several detection methods and is case dependent on the implementation of more than one.
- circRNAs have a great potential to regulate other RNA species, and circRNA functions can be explored by interaction prediction pipelines with both miRNAs and RBPs, and should be corroborated by experimental procedures.

AUTHOR CONTRIBUTIONS

C.R., J.S. and V.M.C. conceived the manuscript; C.R., J.S., N.S. and V.M.C. wrote and revised the manuscript; J.S. generated the Figures; V.M.C. obtained the funding for this initiative.

FUNDING

This work was funded by PhD fellowships (grant number 21191016 to CR and 21220964 to JPS) and grants from Agencia Nacional de Investigación y Desarrollo (ANID), Chile. ANID funded this work through the following grants: FONDECYT-ANID (1211731), FONDAP-ANID (15120011), STIC/AmSud-ANID (STIC2020008) and Anillo-ANID (ACT210004 and ATE220016) to VM-C.

DATA AVAILABILITY

No datasets were generated or analyzed during the elaboration of this work.

REFERENCES

1. Xiao J. *Circular RNAs: Biogenesis and Functions*. Singapore: Springer, Singapore 2018.
2. Pandey PR, Munk R, Kundu G, et al. Methods for analysis of circular RNAs. *Wiley Interdiscip Rev RNA* 2020;**11**:e1566.
3. Salgado-Somoza A, Zhang L, Vausort M, et al. The circular RNA MICRA for risk stratification after myocardial infarction. *Int J Cardiol Heart Vasc* 2017;**17**:33–6.
4. Ni H, Li W, Zhuge Y, et al. Inhibition of circHIPK3 prevents angiotensin II-induced cardiac fibrosis by sponging miR-29b-3p. *Int J Cardiol* 2019;**292**:188–96.
5. Tang C-M, Zhang M, Huang L, et al. CircRNA_000203 enhances the expression of fibrosis-associated genes by derepressing targets of miR-26b-5p, Col1a2 and CTGF, in cardiac fibroblasts. *Sci Rep* 2017;**7**:40342.
6. Lei K, Bai H, Wei Z, et al. The mechanism and function of circular RNAs in human diseases. *Exp Cell Res* 2018;**368**:147–58.
7. Wang K, Long B, Liu F, et al. A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. *Eur Heart J* 2016;**37**:2602–11.
8. Mehta SL, Pandi G, Vemuganti R. Circular RNA expression profiles alter significantly in mouse brain after transient focal ischemia. *Stroke* 2017;**48**:2541–8.

9. Jeck WR, Sorrentino JA, Wang K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;**19**:141–57.
10. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol* 2014;**32**:453–61.
11. Xi Y, Fowdur M, Liu Y, et al. Differential expression and bioinformatics analysis of circRNA in osteosarcoma. *Biosci Rep* 2019;**39**:BSR20181514.
12. Salzman J, Gawad C, Wang PL, et al. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 2012;**7**:e30733.
13. Zhang Y, Zhang X-O, Chen T, et al. Circular intronic long noncoding RNAs. *Mol Cell* 2013;**51**:792–806.
14. Timoteo GD, Di Timoteo G, Dattilo D, et al. Modulation of circRNA metabolism by m6A modification. *Cell Rep* 2020;**31**:107641.
15. Gruhl F, Janich P, Kaessmann H, et al. Circular RNA repertoires are associated with evolutionarily young transposable elements. *Elife* 2021;**10**:e67991.
16. Santos-Rodríguez G, Voineagu I, Weatheritt RJ. Evolutionary dynamics of circular RNAs in primates. *Elife* 2021;**10**:10.
17. Hansen TB, Venø MT, Damgaard CK, et al. Comparison of circular RNA prediction tools. *Nucleic Acids Res* 2016;**44**:e58.
18. Vromman M, Vandesompele J, Volders P-J. Closing the circle: current state and perspectives of circular RNA databases. *Brief Bioinf* 2021;**22**:288–97.
19. Panda AC. Circular RNAs act as miRNA sponges. *Adv Exp Med Biol* 2018;**1087**:67–79.
20. Qu S, Yang X, Li X, et al. Circular RNA: a new star of noncoding RNAs. *Cancer Lett* 2015;**365**:141–8.
21. Holdt LM, Stahringer A, Sass K, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun* 2016;**7**:12429.
22. Vausort M, Salgado-Somoza A, Zhang L, et al. Myocardial infarction-associated circular RNA predicting left ventricular dysfunction. *J Am Coll Cardiol* 2016;**68**:1247–8.
23. Li Z, Zhou Y, Yang G, et al. Using circular RNA SMARCA5 as a potential novel biomarker for hepatocellular carcinoma. *Clin Chim Acta* 2019;**492**:37–44.
24. Liang G, Ling Y, Mehrpour M, et al. Autophagy-associated circRNA circCDYL augments autophagy and promotes breast cancer progression. *Mol Cancer* 2020;**19**:65.
25. Aufiero S, Reckman YJ, Pinto YM, et al. Circular RNAs open a new chapter in cardiovascular biology. *Nat Rev Cardiol* 2019;**16**:503–14.
26. Liu C, Liu Y-C, Huang H-D, et al. Biogenesis mechanisms of circular RNA can be categorized through feature extraction of a machine learning model. *Bioinformatics* 2019;**35**:4867–70.
27. Huang A, Zheng H, Wu Z, et al. Circular RNA-protein interactions: functions, mechanisms, and identification. *Theranostics* 2020;**10**:3503–17.
28. López-Jiménez E, Rojas AM, Andrés-León E. RNA sequencing and prediction tools for circular RNAs analysis. *Adv Exp Med Biol* 2018;**1087**:17–33.
29. Ma X-K, Wang M-R, Liu C-X, et al. CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics Proteomics Bioinf* 2019;**17**:511–21.
30. Zhang J, Chen S, Yang J, et al. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun* 2020;**11**:90.
31. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;**495**:333–8.
32. Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;**37**:907–15.
33. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–60.
35. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* 2012;**9**:357–9.
36. Hansen TB. Improved circRNA identification by combining prediction algorithms. *Front Cell Dev Biol* 2018;**6**:20.
37. Zeng X, Lin W, Guo M, et al. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput Biol* 2017;**13**:e1005420.
38. Westholm JO, Miura P, Olson S, 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.
39. Wang K, Singh D, Zeng Z, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res* 2010;**38**:e178.
40. Zhang X-O, Wang H-B, Zhang Y, et al. Complementary sequence-mediated exon circularization. *Cell* 2014;**159**:134–47.
41. Gao Y, Wang J, Zhao F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol* 2015;**16**:4.
42. Chuang T-J, Wu C-S, Chen C-Y, et al. NCLscan: accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. *Nucleic Acids Res* 2016;**44**:e29.
43. Cheng J, Metge F, Dieterich C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* 2016;**32**:1094–6.
44. Izuogu OG, Alhasan AA, Alafghani HM, et al. PTESFinder: a computational method to identify post-transcriptional exon shuffling (PTES) events. *BMC Bioinf* 2016;**17**:31.
45. Hoffmann S, Otto C, Kurtz S, et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol* 2009;**5**:e1000502.
46. Song X, Zhang N, Han P, et al. Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Res* 2016;**44**:e87.
47. Szabo L, Morey R, Palpant NJ, et al. Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol* 2015;**16**:126.
48. Nguyen DT, Trac QT, Nguyen T-H, et al. Circall: fast and accurate methodology for discovery of circular RNAs from paired-end RNA-sequencing data. *BMC Bioinf* 2021;**22**:495.
49. Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. *Brief Bioinf* 2018;**19**:803–10.
50. Srivastava A, Sarkar H, Gupta N, et al. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcripts. *Bioinformatics* 2016;**32**:i192–200.
51. Chen L, Wang C, Sun H, et al. The bioinformatics toolbox for circRNA discovery and analysis. *Brief Bioinf* 2021;**22**:1706–28.
52. Pan X, Xiong K. PredcircRNA: computational classification of circular RNA from other long non-coding RNA using hybrid features. *Mol Biosyst* 2015;**11**:2219–26.
53. Pan X, Xiong K, Anthon C, et al. WebCircRNA: classifying the circular RNA potential of coding and noncoding RNA. *Genes* 2018;**9**:536.

54. Niu M, Zhang J, Li Y, et al. CirRNAPL: a web server for the identification of circRNA based on extreme learning machine. *Comput Struct Biotechnol J* 2020;**18**:834–42.
55. Wang X, Liu Y, Li J, et al. StackCirRNAPred: computational classification of long circRNA from other lncRNA based on stacking strategy. *BMC Bioinf* 2022;**23**:563.
56. Zhang X-O, Dong R, Zhang Y, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* 2016;**26**:1277–87.
57. Gaffo E, Buratin A, Dal Molin A, Bortoluzzi S. Sensitive, reliable and robust circRNA detection from RNA-seq with CirComPara2. *Brief Bioinf* 2022;**23**:bbab418.
58. Li L, Bu D, Zhao Y. CircRNAwrap - a flexible pipeline for circRNA identification, transcript prediction, and abundance estimation. *FEBS Lett* 2019;**593**:1179–89.
59. Wang J, Wang L. Deep learning of the back-splicing code for circular RNA formation. *Bioinformatics* 2019;**35**:5235–42.
60. Wu J, Li Y, Wang C, et al. CircAST: full-length assembly and quantification of alternatively spliced isoforms in circular RNAs. *Genomics Proteomics Bioinf* 2019;**17**:522–34.
61. Yu KH-O, Shi CH, Wang B, et al. Quantifying full-length circular RNAs in cancer. *Genome Res* 2021;**31**:2340–53.
62. Stefanov SR, Meyer IM. CYCLer-a novel tool for the full isoform assembly and quantification of circRNAs. *Nucleic Acids Res* 2023;**51**:e10.
63. Zheng Y, Zhao F. Detection and reconstruction of circular RNAs from transcriptomic data. *Methods Mol Biol* 2018;**1724**:1–8.
64. Gaffo E, Bonizzato A, Kronnie GT, et al. CirComPara: a multi-method comparative bioinformatics pipeline to detect and study circRNAs from RNA-seq data. *Noncoding RNA* 2017;**3**:8.
65. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;**25**:1105–11.
66. You X, Conrad TO. Acfs: accurate circRNA identification and quantification from RNA-Seq data. *Sci Rep* 2016;**6**:38820.
67. Li L, Zheng Y-C, Kayani MUR, et al. Comprehensive analysis of circRNA expression profiles in humans by RAISE. *Int J Oncol* 2017;**51**:1625–38.
68. Gao Y, Wang J, Zheng Y, et al. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat Commun* 2016;**7**:12060.
69. Li M, Xie X, Zhou J, et al. Quantifying circular RNA expression from RNA-seq data using model-based framework. *Bioinformatics* 2017;**33**:2131–9.
70. Zheng Y, Ji P, Chen S, et al. Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Med* 2019;**11**:2.
71. Ruan H, Wang P-C, Han L. Characterization of circular RNAs with advanced sequencing technologies in human complex diseases. *Wiley Interdiscip Rev RNA* 2023;**14**:e1759.
72. Rahimi K, Venø MT, Dupont DM, et al. Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nat Commun* 2021;**12**:4825.
73. Zhang J, Hou L, Zuo Z, et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat Biotechnol* 2021;**39**:836–45.
74. Xin R, Gao Y, Gao Y, et al. isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat Commun* 2021;**12**:266.
75. Liu Z, Yang E. circFL-seq, a full-length circRNA sequencing method. *Bio-Protocol* 2022;**12**:e4384.
76. Wu W, Ji P, Zhao F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* 2020;**21**:1–14.
77. Feng J, Xiang Y, Xia S, et al. CircView: a visualization and exploration tool for circular RNAs. *Brief Bioinf* 2018;**19**:1310–6.
78. Dahl M, Daugaard I, Andersen MS, et al. Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab Invest* 2018;**98**:1657–69.
79. Lorenz R, Bernhart SH, Siederdisen CH, et al. ViennaRNA package 2.0. *Algorithms Mol Biol* 2011;**6**:26.
80. Biesiada M, Pachulska-Wieczorek K, Adamiak RW, et al. RNA-Composer and RNA 3D structure prediction for nanotechnology. *Methods* 2016;**103**:120–7.
81. Zhang Y, Wang J, Xiao Y. 3dRNA: 3D structure prediction from linear to circular RNAs. *J Mol Biol* 2022;**434**:167452.
82. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
83. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.
84. Fan X, Weng X, Zhao Y, et al. Circular RNAs in cardiovascular disease: an overview. *Biomed Res Int* 2017;**2017**:1–9.
85. Zhou M-Y, Yang J-M, Xiong X-D. The emerging landscape of circular RNA in cardiovascular diseases. *J Mol Cell Cardiol* 2018;**122**:134–9.
86. Legnini I, Di Timoteo G, Rossi F, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in Myogenesis. *Mol Cell* 2017;**66**:22–37.e9.
87. Pamudurti NR, Bartok O, Jens M, et al. Translation of CircRNAs. *Mol Cell* 2017;**66**:9–21.e7.
88. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13.
89. Hermjakob H, Montecchi-Palazzi L, Lewington C, et al. IntAct: an open source molecular interaction database. *Nucleic Acids Res* 2004;**32**:D452–5.
90. Bader GD, Betel D, Hogue CWV. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* 2003;**31**:248–50.
91. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;**44**:W90–7.
92. Mi H, Muruganujan A, Casagrande JT, et al. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 2013;**8**:1551–66.
93. Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 2007;**8**:R183.
94. Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 2006;**34**:W451–4.
95. Brousse C, Liu Q, Beaclair L, et al. A non-canonical plant microRNA target site. *Nucleic Acids Res* 2014;**42**(8):5270–79.
96. Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* 2019;**20**:18.
97. Agarwal V, Bell GW, Nam J-W, et al. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015;**4**:e05005.

98. Enright AJ, John B, Gaul U, et al. MicroRNA targets in drosophila. *Genome Biol* 2003;**5**:R1.
99. Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet* 2005;**37**:495–500.
100. Kertesz M, Iovino N, Unnerstall U, et al. The role of site accessibility in microRNA target recognition. *Nat Genet* 2007;**39**:1278–84.
101. Wright PR, Georg J, Mann M, et al. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 2014;**42**:W119–23.
102. Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 2006;**126**:1203–17.
103. Riffo-Campos AL, Riquelme I, Brebi-Mieville P. Tools for sequence-based miRNA target prediction: what to choose? *Int J Mol Sci* 2016;**17**:1987.
104. Zang J, Lu D, Xu A. The interaction of circRNAs and RNA binding proteins: an important part of circRNA maintenance and function. *J Neurosci Res* 2020;**98**:87–97.
105. Janas T, Janas MM, Sapoń K, et al. Mechanisms of RNA loading into exosomes. *FEBS Lett* 2015;**589**:1391–8.
106. Abdelmohsen K, Panda AC, Munk R, et al. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol* 2017;**14**:361–9.
107. Du WW, Yang W, Liu E, et al. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res* 2016;**44**:2846–58.
108. Armaos A, Colantoni A, Proietti G, et al. catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions. *Nucleic Acids Res* 2021;**49**:W72–9.
109. Dudekula DB, Panda AC, Grammatikakis I, et al. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* 2016;**13**:34–42.
110. Zhang K, Pan X, Yang Y, et al. CRIP: predicting circRNA-RBP-binding sites using a codon-based encoding and hybrid deep neural networks. *RNA* 2019;**25**:1604–15.
111. Wang Z, Lei X. Matrix factorization with neural network for predicting circRNA-RBP interactions. *BMC Bioinf* 2020;**21**:229.
112. Yang Y, Hou Z, Ma Z, et al. iCircRBP-DHN: identification of circRNA-RBP interaction sites using deep hierarchical network. *Brief Bioinf* 2021;**22**:bbaa274.
113. Dong X, Chen K, Chen W, et al. circRIP: an accurate tool for identifying circRNA-RBP interactions. *Brief Bioinf* 2022;**23**:bbac186.
114. Panda A, Gorospe M. Detection and analysis of circular RNAs by RT-PCR. *Bio-Protocol* 2018;**8**:e2775–e2775.
115. Pfafenrot C, Preußner C. Establishing essential quality criteria for the validation of circular RNAs as biomarkers. *Biomol Detect Quantif* 2019;**17**:100085.
116. Dori M, Biciato S. Integration of bioinformatic predictions and experimental data to identify circRNA-miRNA associations. *Genes* 2019;**10**:642.
117. Guo JU, Agarwal V, Guo H, et al. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* 2014;**15**:409.
118. Pervouchine DD. Circular exonic RNAs: when RNA structure meets topology. *Biochim Biophys Acta Gene Regul Mech* 2019;**1862**:194384.
119. Ghosal S, Das S, Sen R, et al. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet* 2013;**4**:283.
120. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;**20**:1666–70.
121. Zhao S, Li S, Liu W, et al. Circular RNA signature in lung adenocarcinoma: a MiOncoCirc database-based study and literature review. *Front Oncol* 2020;**10**:523342.
122. Ruan H, Xiang Y, Ko J, et al. Comprehensive characterization of circular RNAs in ~ 1000 human cancer cell lines. *Genome Med* 2019;**11**:55.
123. Rophina M, Sharma D, Poojary M, et al. Circad: a comprehensive manually curated resource of circular RNA associated with diseases. *Database* 2020;**2020**:baaa019.
124. Zhao Z, Wang K, Wu F, et al. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. *Cell Death Dis* 2018;**9**:475.
125. Yao D, Zhang L, Zheng M, et al. Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. *Sci Rep* 2018;**8**:11018.
126. Chen L-L, Bindereif A, Bozzoni I, et al. A guide to naming eukaryotic circular RNAs. *Nat Cell Biol* 2023;**25**:1–5.
127. Dong R, Ma X-K, Li G-W, et al. CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinf* 2018;**16**:226–33.
128. Li J-H, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014;**42**:D92–7.
129. Fan C, Lei X, Tie J, et al. CircR2Disease v2.0: an updated web server for experimentally validated circRNA-disease associations and its application. *Genomics Proteomics Bioinf* 2022;**20**:435–45.
130. Liu M, Wang Q, Shen J, et al. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol* 2019;**16**:899–905.
131. Liu Y-C, Li J-R, Sun C-H, et al. CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res* 2016;**44**:D209–15.
132. Chen Y, Yao L, Tang Y, et al. CircNet 2.0: an updated database for exploring circular RNA regulatory networks in cancers. *Nucleic Acids Res* 2022;**50**:D93–101.
133. Chen X, Han P, Zhou T, et al. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* 2016;**6**:34985.
134. Feng J, Chen W, Dong X, et al. CSCD2: an integrated interactional database of cancer-specific circular RNAs. *Nucleic Acids Res* 2022;**50**:D1179–83.
135. Xia S, Feng J, Lei L, et al. Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinf* 2017;**18**:984–92.
136. Tang Z, Li X, Zhao J, et al. TRCirc: a resource for transcriptional regulation information of circRNAs. *Brief Bioinf* 2019;**20**:2327–33.
137. Lai H, Li Y, Zhang H, et al. exoRBase 2.0: an atlas of mRNA, lncRNA and circRNA in extracellular vesicles from human biofluids. *Nucleic Acids Res* 2022;**50**:D118–28.
138. Voineagu I, Walsh K, Gokool A, et al. NeuroCirc: an integrative resource of circular RNA expression in the human brain. *Bioinformatics* 2021;**37**(20):3664–66.
139. Jakobi T, Dieterich C. Computational approaches for circular RNA analysis. *Wiley Interdiscip Rev RNA* 2019;**10**(3):e1528.