

Review

Approaches and challenges in genome-wide circular RNA identification and quantification

Xu-Kai Ma,^{1,*} Si-Nan Zhai,^{1,2} and Li Yang^{1,*}

Numerous circular RNAs (circRNAs) produced from back-splicing of exon(s) have been recently revealed on a genome-wide scale across species. Although generally expressed at a low level, some relatively abundant circRNAs can play regulatory roles in various biological processes, prompting continuous profiling of circRNA in broader conditions. Over the past decade, distinct strategies have been applied in both transcriptome enrichment and bioinformatic tools for detecting and quantifying circRNAs. Understanding the scope and limitations of these strategies is crucial for the subsequent annotation and characterization of circRNAs, especially those with functional potential. Here, we provide an overview of different transcriptome enrichment, deep sequencing and computational approaches for genome-wide circRNA identification, and discuss strategies for accurate quantification and characterization of circRNA.

Highlights

Multiple datasets and computational approaches have been developed for genome-wide circular RNA (circRNA) profiling.

The unified key principle in genome-wide circRNA profiling is to identify reads mapped to circRNA-featured back-splicing junctions (BSJs).

It is essential to consider cognate linear RNA expression and effects for functional circRNA quantification and characterization.

Transcriptome-wide identification of circRNAs

Covalently closed RNA molecules in nature were first observed in plants as pathogenic viroid RNA genomes by electron micrographs in the 1970s [1]. Since then, a variety of circles of RNA had been sporadically reported to be produced in other viruses, such as the RNA genomes in hepatitis δ virus [2]; from (precursors of) noncoding RNAs, including (but not limited to) mitochondrial RNAs [3,4], rRNAs [5,6], and tRNAs [7,8]; and from mRNA precursors (pre-mRNAs) in eukaryotic cells [9,10] (see also reviews in [11–13]). Interestingly, the observation of eukaryotic RNA circles could be traced back to the original perception by electron micrographs in the late 1970s, mainly in extracted cytoplasmic fractions of HeLa cells, but the origin and function were unknown [14].

More than a decade later, with distinct biochemical assays, such as PCR cloning, Northern blotting, and RNase H digestion, the exons or introns of some eukaryotic protein-coding genes were shown to possess the capability to generate stable circular transcripts through different mechanisms [9,10]. On the one hand, a few of the eukaryotic exons can be joined at their consensus splice sites, but in a scrambled manner, which is now referred to as back-splicing, a process that links a downstream 5' back-splicing site reversely to an upstream 3' back-splicing site [12,15] to form a circular transcript (Figure 1, Key figure) [9,16–20]. Strikingly, the circular transcript generated from the mouse sex-determining *Sry* gene was shown to be the predominant output of its host gene locus, suggesting functional potential of circular *Sry* transcripts [17]. On the other hand, intronic sequences could be also accumulated after splicing as lariats, and their location near transcription sites in the nucleus of eukaryotic cells also indicates their functionality [10,21]. However, only a handful of scrambled exon-producing RNA circles or intron lariats were sporadically reported in the 1990s [9,10,16–21].

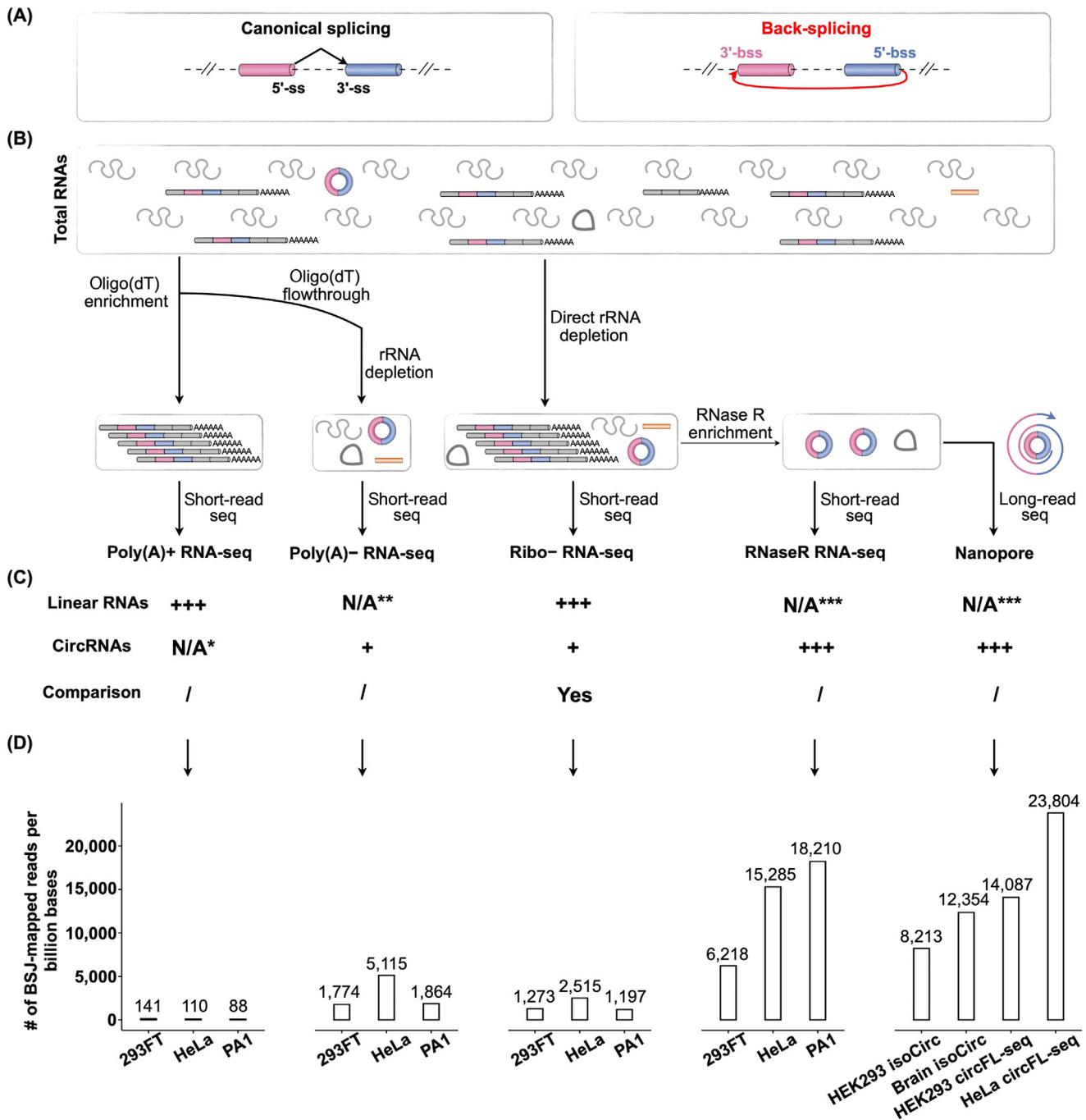
Whole-transcriptome analyses of circular transcripts were successfully achieved in the early 2010s when nonpolyadenylated RNAs were specifically fractionated for further investigation [22–24]. Importantly, independent laboratories have convincingly demonstrated that some

¹Center for Molecular Medicine, Children's Hospital, Fudan University and Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China
²Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

*Correspondence: maxukai@fudan.edu.cn (X.-K. Ma) and liyang_fudan@fudan.edu.cn (L. Yang).

Key figure

Different biochemical approaches for transcriptome-wide circRNA profiling



Trends in Genetics
(See figure legend at the bottom of the next page.)

circRNAs generated from back-splicing of exons, or circular intronic RNAs (ciRNAs) stabilized with a lariat structure could play previously underappreciated roles in the regulation of gene expression [25–28], strongly promoting studies on circular RNAs. Nowadays, we have learned that a large number of circRNAs and ciRNAs can be generated from eukaryotic pre-(m)RNAs in a spliceosome-dependent manner, and the vast majority of them are circRNAs featuring back-splicing junctions (BSJs) [13,29]. In addition, the production of circRNAs is regulated in multiple layers and competes with that of linear RNAs [30–35]. Furthermore, some highly expressed circRNAs have been suggested to be involved in a variety of functions, such as immune responses and tumorigenesis [36–38] (see also reviews in [13,15,39]) with distinct modes of action [39–41], shedding new light on the biological significance of circRNAs. Moreover, due to their stability inside cells and in extracellular fluids, circRNAs are thought to be potential therapeutic targets and diagnostic biomarkers for diseases such as cancer [42,43]. Finally, both engineered and endogenous circRNAs were shown to be translatable [44–48], highlighting the use of RNA circles as novel translational platforms for RNA medicine.

Compared with canonical splicing with a colinear order in linear RNAs, noncolinear BSJs feature in circRNAs (Figure 1A). In this scenario, identifying sequencing reads mapped to circRNA-specific BSJs has led to the computational detection of a substantial number of circRNAs from deep sequencing datasets that are still growing quickly. Since they are largely coexpressed with their cognate linear RNAs with almost fully overlapping sequences, precisely annotating and quantifying circRNAs on a genome-wide scale has been challenging [12,15,49]. In this review, we discuss different strategies for the purification and enrichment of circRNA, distinct sequencing approaches, and the corresponding computational setups for genome-wide circRNA profiling, and emphasize the importance of accurate circRNA quantification for cross-sample comparisons when using distinct strategies.

Different enrichment approaches and sequencing platforms for circRNA profiling

As they are covalently closed without the canonical 3'-polyadenylation tail, large-scale identification of circular transcripts fell under the radar of early whole-transcriptomic profiling on polyadenylated [poly(A)⁺] RNAs enriched by oligo(dT) beads [50,51], named poly(A)⁺ RNA-seq (Figure 1B), but it was achieved by other (circular) RNA enrichment strategies. First, by collecting fractionations that are not associated with oligo(dT) beads and further depleting redundant rRNAs, nonpolyadenylated [poly(A)⁻] RNAs were extracted for deep sequencing [22,52], named poly(A)⁻ RNA-seq (Figure 1B), leading to genome-wide identification of diverse types of nonpolyadenylated RNAs, including circRNAs and ciRNAs [23,24,27,32]. Alternatively, rRNA-depleted (ribo⁻) RNAs, which contain both poly(A)⁺ and poly(A)⁻ transcripts, are now widely applied to deep sequencing, named ribo⁻ RNA-seq (Figure 1B), for profiling circRNA and ciRNA [26,53,54]. Of note, other types of nonpolyadenylated RNAs other than circRNAs could be also detected in poly(A)⁻ and ribo⁻ RNA-seq datasets [55]. Moreover, when further enriched by additional RNase R treatment, which digests linear RNAs, circRNAs and ciRNAs could be

Figure 1. (A) Schematic of canonical splicing (left) and back-splicing (right). (B) Schematic of different RNA-seq approaches with distinct enrichment strategies and/or sequencing platforms. (C) Contents of circular RNAs (circRNAs) and linear RNAs that can be theoretically detected by different RNA-seq approaches. *, some circRNAs might be nonspecifically detected in polyadenylated [poly(A)⁺] RNA-seq. **, some nonpolyadenylated linear RNAs could be detected in poly(A)⁻ RNA-seq. *** some linear RNAs might be nonspecifically detected in RNaseR RNA-seq due to incomplete degradation by RNase R. (D) Statistics on back-splicing junction (BSJ)-mapped reads that can be detected by different sequencing methods, normalized by sequencing depth (per billion mapped bases). Deep sequencing datasets are from 293FT [103] (GEO: GSE172193 and GSE149691), HeLa [22,109,110] (GEO: GSE24399, GSE53328, and GSE90247), and PA1 [35,86] (GSE75733, GSE73325) cell lines; isoCirc datasets [84] (GEO: GSE141693); and circFL-seq datasets [85] (BioProject: PRJNA722575) were downloaded for this analysis by the CiRCexplorer3/CLEAR or corresponding pipeline [58]. Of note, different cell lines, sequencing platforms, and other batch effects also could lead to the fluctuations in the numbers of the BSJ-mapped reads. Abbreviations: bss, back-splicing site; N/A, not available; ss, splicing site.

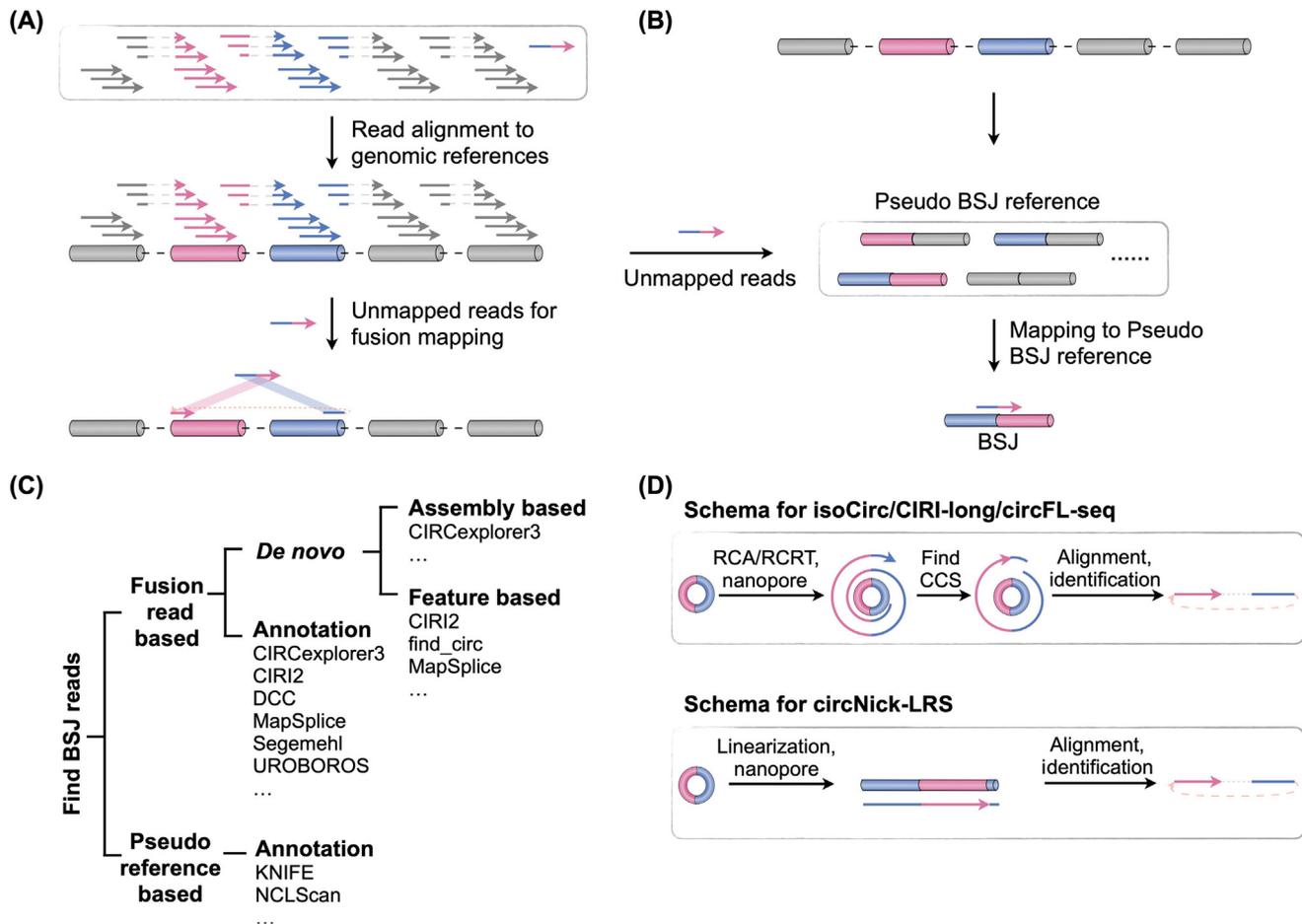
significantly enriched and overwhelmed in subsequent deep sequencing, named RNaseR RNA-seq (Figure 1B). Finally, circRNA profiling has been recently achieved by taking advantage of long-read sequencing platforms, such as Oxford Nanopore, which generally incorporate RNase R-treatment before long-read sequencing for circRNA enrichment (Figure 1B).

All these aforementioned datasets have been applied to the genome-wide detection of circRNAs, and each of them presents unique opportunities and challenges in circRNA profiling due to their distinct (circular) RNA enrichment strategies and sequencing platforms. Briefly, poly(A)⁺ RNA-seq was originally designed to profile mRNAs [and long noncoding (lnc)RNAs] with a 3'-polyadenylated tail [56] (Figure 1C), but was not intended to identify circRNAs that have no open ends. In this scenario, poly(A)⁺ RNA-seq datasets are far from ideal for genome-wide circRNA profiling. However, possibly due to nonspecific binding of circRNAs with oligo(dT)-beads and/or other reasons, some poly(A)⁺ RNA-seq reads could be still mapped to circRNA-featuring BSJs, leading to less efficient circRNA profiling by poly(A)⁺ RNA-seq [57]. Instead, poly(A)⁻ and ribo⁻ RNA-seq datasets are more widely used for genome-wide circRNA profiling than poly(A)⁺ ones. Indeed, about tenfold more BSJ-mapped reads could be detected by poly(A)⁻ and ribo⁻ RNA-seq (Figure 1D). Theoretically, poly(A)⁻ RNA-seq datasets do not contain polyadenylated (linear) RNAs, while ribo⁻ RNA-seq datasets consist of both polyadenylated and nonpolyadenylated (circular) RNAs (Figure 1C), making them suitable for the comparison of circRNAs with their cognate linear ones than other sequencing datasets [58,59]. To further enrich circRNAs, pretreatment with RNase R, which is a 3'-to-5' exoribonuclease that efficiently degrades linear RNAs [24,32], before deep sequencing is generally applied in circRNA identification. About three- to tenfold more BSJ-mapped reads could be detected in RNase R-treated short- and long-read sequencing datasets than non-RNase R-treated ones (Figure 1D). Since RNase R treatment removes the false positives resulting from *trans*-splicing and/or reverse transcriptase template switching, it has been widely adopted for validating circRNA [49,60–62]. However, prolonged RNase R incubation can lead to degradation of some circRNAs [63], especially those with longer sequences [64,65], requiring attention during studies.

Unified principle of circRNA annotation by identifying the reads mapped to BSJ sites

Genome-wide profiling of circRNAs from high-throughput sequencing data mainly relies on computational pipelines to detect BSJs featuring circRNA [24,32] (Figure 1A). Regardless of the distinct (circ)RNA purification/enrichment strategies and different sequencing platforms used for genome-wide identification of circRNA (Figure 1B,C), a unified principle is applied in nearly all computational pipelines for reliable annotation of circRNA: identifying sequencing reads mapped to BSJs. So far, more than a dozen computational methods have been developed for profiling circRNA from short-read RNA-seq datasets [66]. Depending on how they identify the reads mapped to BSJs, these computational methods can be simply divided into two categories: fusion-read-based (Figure 2A) or pseudo-reference-based (Figure 2B) methods [66].

Fusion-read-based methods directly align the sequencing reads to genomic references, followed by detecting the fusion (or chimeric) reads mapped in a noncolinear manner (Figure 2A). The accuracy and precision of detecting fusion reads are essential for the characterization of circRNA. Common pipelines capable of extracting fusion reads include TopHat-Fusion [67], STAR [68], and BWA [69]. After identifying fusion reads, some pipelines, such as CIRCexplorer3/CLEAR (using TopHat-Fusion, STAR, or BWA) [58], CIRI2 (using BWA) [70], DCC (using STAR) [71], and MapSplice [72], make use of annotated gene exon information to improve the accuracy of identifying BSJ-based circRNA. Additionally, some pipelines have also developed unique strategies for *de novo* discoveries of circRNAs containing unannotated exons. For example,



Trends In Genetics

Figure 2. Calls of back-splicing junction (BSJ)-mapped reads for circular RNA (circRNA) annotation. (A) Schematic drawing for fusion-read-based methods to call BSJ-mapped reads. (B) Schematic of pseudo-reference-based methods to call BSJ-mapped reads. (C) Summary of available bioinformatic tools to call BSJ-mapped reads from RNA-seq datasets. (D) Schematic of the available bioinformatic tools used to call BSJ-mapped reads from long-read sequencing datasets. Abbreviations: CCS, cyclic consensus sequences; RCA, rolling circle amplification; RCRT, rolling circle reverse transcription.

CIRCexplorer3 uses tools such as StringTie [73] or Cufflinks [74] to assemble new transcripts, while CIRI2 and find_circ use splicing signals, splice site distances, and some other factors to achieve a similar goal (Figure 2C).

Pseudo-reference-based methods require initial construction of pseudo-BSJ references according to existing gene annotation data before mapping (Figure 2B). The pseudo-reference-based method has been implemented in the KNIFE [75] and NCLScan [76] pipelines to identify reads mapped to the pseudo-BSJ references (Figure 2C). To remove potential false positives, these pipelines typically align reads to the genome and transcriptome references first, and then to the pseudo-BSJ references (Figure 2B). As they rely on the known gene annotations, pseudo-reference-based methods may not be suitable for *de novo* discovery of circRNA [75,76].

Although they have been developed by different laboratories with distinct setups and strategies, most of these computational methods can be used for analyzing all types of short-read RNA-seq datasets, including the aforementioned poly(A)⁺, poly(A)⁻, ribo⁻ and RNaseR methods. By taking

advantage of this commonality, these publicly available computational methods have been extensively compared. It has been shown that certain tools, such as the CIRCexplorer series, MapSplice, and CIRI series, exhibit relatively better performance with fewer false positives [77,78]. It was also suggested that using multiple tools for detecting circRNA can provide more accurate results [77,78], leading to new pipelines that amalgamate a variety of tools. For example, CircComPara2 [79] integrates CIRCexplorer2, find_circ, and CIRI, whereas CircRNAwrap [80] unites up to eight circRNA detection tools and multiple circRNA analysis tools. Nevertheless, for examining RNA-seq datasets with super-high depths (up to 300 million of 2×150 paired-end reads), a very recent study also highlighted that different levels of sensitivity but not precision were observed in most computational tools for circRNA profiling [81].

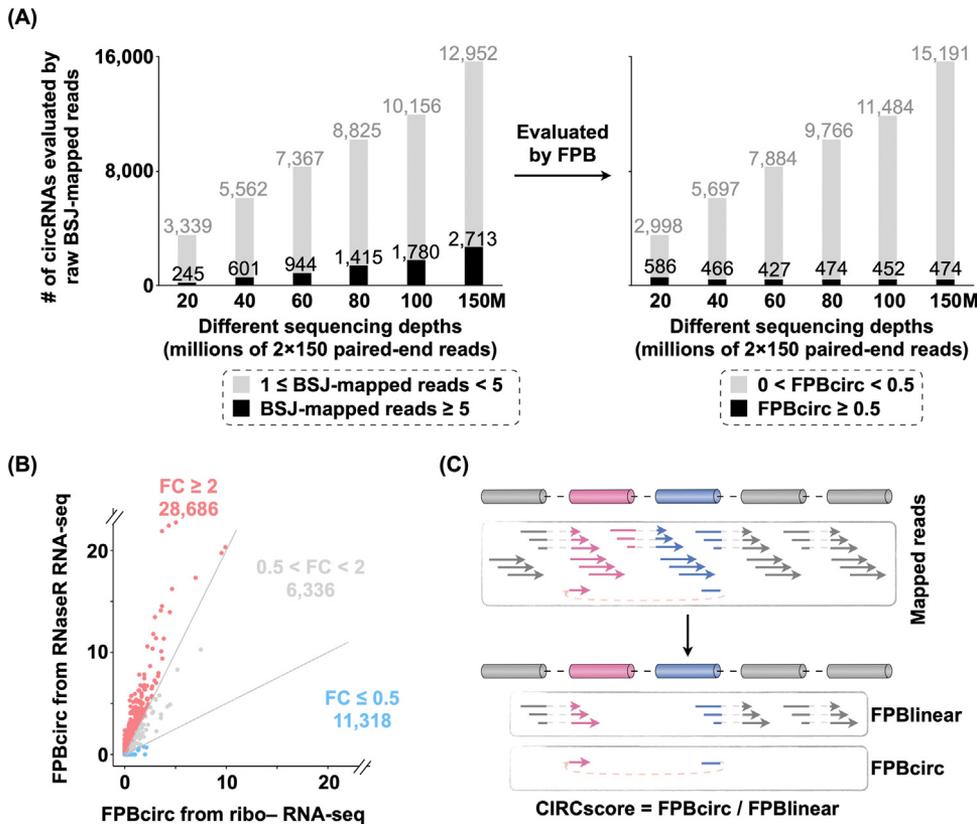
Other than short-read sequencing, circRNA profiling has also successfully been performed by long-read sequencing. Currently, four types of long-read sequencing datasets, namely circNick-LRS [82], CIRI-long [83], isoCirc [84], and circFL-seq [85], have been reported for genome-wide examinations of circRNA. Although they have all been adapted to the same Oxford Nanopore platform, these four datasets were individually generated with different sample preparation and processed with corresponding specific computational algorithms for identifying circRNA. More specifically, CIRI-long, circFL-seq, isoCirc require rolling circle reverse transcription (RCRT) or rolling circle amplification (RCA) processes and need a step for identifying cyclic consensus sequences (CCS) or consensus sequences (Figure 2D). By contrast, circNick-LRS skips the RCRT or RCA step for direct long-read sequencing of circRNAs (Figure 2D). Compared with short-read datasets, long-read sequencing datasets have obvious advantages for examining full-length circRNA and identifying isoforms. For instance, more alternative circularization and internal alternative splicing events could be detected by long-read sequencing [82–85] than by short-read RNA-seq methods [29,86]. Thus, despite the limitations of high cost and high sequencing error rates [83–85], long-read sequencing is at least a beneficial supplement for full-length annotation of circRNAs [29].

Interestingly, by incorporating cutting-edge approaches such as machine learning, additional software tools, such as PredcircRNA [87] and WebCircRNA [88], have been recently constructed to predict the existence of circRNAs with extracted genomic sequence features only. However, these approaches based on artificial intelligence models cannot be used to explain the differential expression of circRNAs across cell lines and tissues that have the same genomic DNA sequences. Of note, several studies have reported systematic comparisons of different strategies, such as short-read and long-read sequencing datasets, for circRNA quantification [81,89], and additional tips to ensure efficient circRNA profiling with distinct bioinformatic pipelines have been summarized [77,78,81].

Challenges in circRNA quantification and subsequent cross-sample comparison

By taking advantage of different computational pipelines to identify reads mapped to BSJs from various types of sequencing datasets, a significant number of circRNAs have been detected in various cell lines and tissues under diverse contexts and across distinct species [32,75,90–92], annotated in different databases, including (but not limited to) CSCD [93], CSCD2 [94], CircRic [95], TSCD [96], CIRCpedia [97], and CircAtlas [98]. Accurate quantification of these circRNAs is essential for subsequent analyses, such as cross-sample comparison to find circRNAs with biological potential.

In line with the unified principle to identify sequencing reads mapped to BSJs by all computational pipelines for circRNA annotation, the raw number of BSJ-mapped reads has been naturally used to quantify circRNAs. Obviously, circRNAs with more BSJ-mapped reads are expressed at



Trends in Genetics

Figure 3. Genome-wide quantification of circular RNAs (circRNAs). (A) Comparison of circRNA quantification by raw numbers of back-splicing junction (BSJ)-mapped reads (left) or normalized fragments per billion mapped bases (FPB) values of BSJ-mapped reads (right). Black bars indicate circRNAs with raw numbers of BSJ-mapped reads ≥ 5 (left) or FPBcirc ≥ 0.5 (right). Gray bars indicate circRNAs with raw numbers of BSJ-mapped reads < 5 (left) or FPBcirc < 0.5 (right). Six virtual RNA-seq datasets were randomly extracted from the original HLF cell rRNA-depleted (ribo⁻) RNA-seq dataset [81] (BioProject: PRJNA789110) containing about 300 million of 2x150 paired-end reads to mimic different sequencing depths, including 20, 40, 60, 80, 100 and 150 million of 2x150 paired-end reads. (B) Comparison of detected expression levels of circRNAs from ribo⁻ or RNaseR RNA-seq samples in HLF cells [81]. Light red dots, 28 686 RNase R enriched circRNAs with FC ≥ 2 in HLF cells. Light blue dots, 11 318 RNase R-depleted circRNAs with FC ≤ 0.5 in HLF cells; gray dots, 6336 circRNAs with $0.5 < FC < 2$ in HLF cells. (C) Schematic drawing for direct circRNA and linear RNA expression comparison from ribo⁻ RNA-seq by FPB values. The expression of circRNA and its cognate linear RNA is individually evaluated by normalized BSJ- and SJ- mapped reads, as FPBcirc and FPBlinear, respectively. The CIRCscore is further determined by dividing FPBcirc by FPBlinear to demonstrate the relative circRNA expression. Abbreviations: FC, fold change of FPBcirc from RNaseR RNA-seq versus FPBcirc from ribo⁻ RNA-seq, evaluated by the CIRCexplorer3/CLEAR pipeline [58]; FPBcirc, BSJ-mapped fragments per billion mapped bases; FPBlinear, SJ-mapped fragments per billion mapped bases; SJ, splice junction.

higher levels than those with fewer BSJ-mapped reads in the same datasets. However, since variable sequencing depths are generally applied in different datasets and studies, using the absolute numbers of BSJ-mapped reads for cross-sample comparisons can be biased. For example, along with the increase in sequencing depths from 20, 40, 60, 80, and 100 to 150 million reads randomly extracted from the same HLF ribo⁻ RNA-seq dataset [81] (BioProject: PRJNA789110), the number of total circRNAs with ≥ 1 BSJ-mapped reads and of highly expressed circRNAs with ≥ 5 BSJ-mapped reads both scaled up accordingly (Figure 3A, left). However, when the read values were normalized by sequencing depth, such as BSJ-mapped fragments per billion mapped bases (FPB), as used by the CIRCexplorer3/CLEAR pipeline [58], the number of highly-expressed circRNAs with ≥ 0.5 FPB, which is similar to ≥ 6 BSJ-mapped reads from 40 million of

2×150 paired-end reads, remains stable with an increase in the sequencing depths from 20 to 150 million, whereas more circRNAs with lower expression levels ($0 < \text{FPB} < 0.5$) could be identified by the increase in sequencing depths (Figure 3A, right). These results suggested that although circRNAs with lower expression levels could be further detected by the increase in sequencing depths, highly expressed circRNAs (with ≥ 0.5 FPB) could be detected with sequencing depths of even 20 million of 2×150 paired-end reads. It is worthwhile noting that FPB also considers the read lengths (such as 1×50 versus 1×101) and sequencing strategies (such as single-end versus paired-end) to normalize BSJ-mapped reads and thus is more tolerant to changes between samples with different sequencing read lengths and/or strategies [58].

In addition, different purification and enrichment strategies also affect the numbers of BSJ-mapped reads called from different datasets, such as those with or without RNase R treatment. As shown in Figure 1D, three- to tenfold more BSJ-mapped reads could be detected in samples treated with RNase R than in those without RNase R treatment. Since a nonnegotiable portion of the circRNAs were shown to be depleted by RNase R treatment (Figure 3B, light blue dots indicate RNase R-depleted circRNAs) and different RNase R treatment conditions could also affect circRNA enrichment [63], quantifying the circRNAs with RNase R datasets and their cross-sample comparisons might be biased. It is also worth noting that some types of enzymatic artifacts, such as template switching, RCA, and ligation artifacts, can occur during the library preparation step, leading to false positives and/or biased quantification results [49,99,100]. Meanwhile, nearby homologous genes can also give rise to canonical linear splice junctions that mimic BSJs, resulting in false signals being detected by bioinformatic algorithms [101,102].

Last but not least, linear RNA expression should be considered when evaluating the expression of circRNA because of the coexistence of circRNAs and their cognate linear RNAs originating from the same genomic loci. Functional studies of circRNAs might be overshadowed by the highly expressed linear RNA isoforms due to their almost totally overlapping sequences [58,59,103,104]. One possible solution is to find highly expressed circRNAs with a background low-expression of linear forms. However, direct expression comparison of circRNAs with cognate linear RNAs is difficult, not only because of their sequence similarity, but also because of the distinct strategies used for quantifying circular or linear RNA. Generally, linear RNA expression is calculated by mapped reads normalized by gene length and sequencing depth, such as fragments per kilobase of transcript per million mapped reads (FPKM) or transcripts per million (TPM) [105]. However, circRNAs are calculated by BSJ-mapped reads only, such as FPB. Since FPKM and TPM are not scaled to FPB, the expression of circRNAs measured by FPB is not similar to that of linear RNAs measured by FPKM or TPM (Figure 3C). To solve this problem, Sailfish-cir [106] reports TPMs for both circular and linear isoforms for their comparison. CIRCexplorer3/CLEAR, CIRI2, CIRIquant, and DCC report BSJ-mapped reads and reads aligned to colinear exon–exon junctions to evaluate the expression levels of circular or linear RNA, respectively. The latter approach has been extended for direct comparison of the expression of circular and linear RNA (Figure 3C), such as obtaining the CIRCscore with CIRCexplorer3/CLEAR [58], or the circular to linear ratios with CIRIquant [55]. By using the expression of cognate linear RNA as the background, highly expressed circRNAs with low-expression linear RNAs can be selected for subsequent functional studies [58]. Of special note, given their specific transcript enrichment strategies, ribo⁻ RNA-seq datasets are more suitable for direct circular and linear RNA expression comparison than others, such as poly(A)⁻ or RNase R datasets. Nevertheless, except for a small portion [90], the vast majority of circRNAs are much less abundant than their cognate linear RNAs, and useful approaches to ensure the characterization of functional circRNAs with highly expressed linear RNAs have been demonstrated [64,65].

Concluding remarks and future perspectives

In the past decade, multiple (circular) RNA enrichment strategies and sequencing technologies have been adopted to achieve genome-wide characterization of circRNA (Figure 1). By taking advantage of special computational frameworks to identify reads mapped to BSJs featuring circRNAs, a large number of circRNAs have been uncovered from various types of sequencing datasets (Figure 2). Methodologies have also been developed and applied to precisely quantify the expression of circRNA, normalized by sequencing depths and the expression levels of cognate linear RNA, facilitating cross-sample comparisons for highly expressed circRNAs with biological significance (Figure 3). Despite these achievements, ambiguous read mapping events, which occur when short sequencing reads align to multiple genomic loci with similar sequence content, as well as the batch effects arising from sample preparation, library construction, and different sequencing runs, challenge precise identification of circRNA on a genome-wide scale [58,59]. Carefully designing experiments and including biological replicates should be considered for accurate assessment of circRNA detection. It is worth noting that other types of circRNAs, such as ciRNAs, have been co-detected with circRNAs via transcriptomic profiling, although their mechanisms of biogenesis and modes of action are distinct. Several open questions remain (see Outstanding questions). For example, how can we efficiently and precisely profile circRNAs at the single-cell level [107,108]? Most single-cell RNA-seq datasets basically contain poly(A)⁺ RNAs, which are still far from ideal for the characterization of circRNA at single-cell resolution. More specific approaches are therefore needed to enrich circRNAs from single cells for subsequent sequencing and bioinformatic analysis. In addition, seamless analytical pipelines from profiling to quantification and then to functional prediction of circRNAs are also desired.

To conclude, the intricacy of genome-wide circRNA profiling and their accurate quantification stems from the coexpression of circular and linear RNA transcripts with overlapping sequences. Diverse transcriptome enrichment strategies, sequencing technologies, and computational approaches have facilitated the identification of circRNA and subsequent cross-sample comparison to inspect their functional roles in various biological settings. Additional endeavors to develop innovative techniques specifically designed for circRNAs will unquestionably result in a deeper comprehension of these fascinating noncoding RNA molecules with a unique circular formation and their prospective application in diagnosis and therapeutics.

Acknowledgments

We apologize to colleagues whose work could not be discussed owing to space limitations. This work was supported by the National Natural Science Foundation of China (31925011), the Ministry of Science and Technology of China (2021YFA1300503, 2019YFA0802804), and the Chinese Academy of Sciences, China (XDB38040300) to L.Y., and by Science and Technology Commission of Shanghai Municipality, China (STCSM) (23YF1407400) to X.K.M.

Declaration of interests

The authors declare no competing interests.

References

1. Sanger, H.L. *et al.* (1976) Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U. S. A.* 73, 3852–3856
2. Kos, A. *et al.* (1986) The hepatitis delta (delta) virus possesses a circular RNA. *Nature* 323, 558–560
3. Arnberg, A.C. *et al.* (1980) Some yeast mitochondrial RNAs are circular. *Cell* 19, 313–319
4. Hensgens, L.A. *et al.* (1983) Variation, transcription and circular RNAs of the mitochondrial gene for subunit I of cytochrome c oxidase. *J. Mol. Biol.* 164, 35–58
5. Grabowski, P.J. *et al.* (1981) The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of *Tetrahymena*. *Cell* 23, 467–476
6. Kjems, J. and Garrett, R.A. (1988) Novel splicing mechanism for the ribosomal RNA intron in the archaeobacterium *Desulfurococcus mobilis*. *Cell* 54, 693–703
7. Clouet d'Orval, B. *et al.* (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res.* 29, 4518–4529
8. Singh, S.K. *et al.* (2004) Sequential 2'-O-methylation of archaeal pre-tRNATrp nucleotides is guided by the intron-encoded but trans-acting box C/D ribonucleoprotein of pre-tRNA. *J. Biol. Chem.* 279, 47661–47671

Outstanding questions

Differential expression of circRNAs has been observed among cell lines and tissues. Can experimental (and possibly computational) approaches be developed to address dynamic circRNA expression at the single-cell level?

Can a comprehensive circRNA analysis process be created to facilitate identification, quantification, and downstream functional annotation of circRNAs?

Both *cis*-regulatory elements and *trans*-acting factors regulate circRNA expression. Is it feasible to model these factors in order to predict circRNA expression under different conditions?

9. Cocquerelle, C. *et al.* (1992) Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* 11, 1095–1098
10. Qian, L. *et al.* (1992) A spliced intron accumulates as a lariat in the nucleus of T cells. *Nucleic Acids Res.* 20, 5345–5350
11. Lasda, E. and Parker, R. (2014) Circular RNAs: diversity of form and function. *RNA* 20, 1829–1842
12. Chen, L.L. (2016) The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 17, 205–211
13. Yang, L. *et al.* (2022) Biogenesis and regulatory roles of circular RNAs. *Annu. Rev. Cell Dev. Biol.* 38, 263–289
14. Hsu, M.T. and Coca-Prados, M. (1979) Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* 280, 339–340
15. Chen, L.L. (2020) The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 21, 475–490
16. Nigro, J.M. *et al.* (1991) Scrambled exons. *Cell* 64, 607–613
17. Capel, B. *et al.* (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* 73, 1019–1030
18. Cocquerelle, C. *et al.* (1993) Mis-splicing yields circular RNA molecules. *FASEB J.* 7, 155–160
19. Dubin, R.A. *et al.* (1995) Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* 167, 245–248
20. Pasmán, Z. *et al.* (1996) Exon circularization in mammalian nuclear extracts. *RNA* 2, 603–610
21. Kopczyński, C.C. and Muskavitch, M.A. (1992) Introns excised from the Delta primary transcript are localized near sites of Delta transcription. *J. Cell Biol.* 119, 503–512
22. Yang, L. *et al.* (2011) rRNA-depleted (ribo⁻) RNAs characterization of non-polyadenylated RNAs. *Genome Biol.* 12, R16
23. Salzman, J. *et al.* (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 7, e30733
24. Jeck, W.R. *et al.* (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157
25. Hansen, T.B. *et al.* (2013) Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388
26. Memczak, S. *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338
27. Zhang, Y. *et al.* (2013) Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806
28. Li, X. *et al.* (2021) Linking circular intronic RNA degradation and function in transcription by RNase H1. *Sci. China Life Sci.* 64, 1795–1809
29. Chen, L.L. *et al.* (2023) A guide to naming eukaryotic circular RNAs. *Nat. Cell Biol.* 25, 1–5
30. Ashwal-Fluss, R. *et al.* (2014) circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66
31. Ivanov, A. *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 10, 170–177
32. Zhang, X.O. *et al.* (2014) Complementary sequence-mediated exon circularization. *Cell* 159, 134–147
33. Chen, L.L. and Yang, L. (2015) Regulation of circRNA biogenesis. *RNA Biol.* 12, 381–388
34. Starke, S. *et al.* (2015) Exon circularization requires canonical splice signals. *Cell Rep.* 10, 103–111
35. Zhang, Y. *et al.* (2016) The biogenesis of nascent circular RNAs. *Cell Rep.* 15, 611–624
36. Liu, C.X. *et al.* (2019) Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell* 177, 865–880
37. Guarnerio, J. *et al.* (2016) Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. *Cell* 165, 289–302
38. Wu, K. *et al.* (2019) Circular RNA F-circSR derived from SLC34A2-ROS1 fusion gene promotes cell migration in non-small cell lung cancer. *Mol. Cancer* 18, 98
39. Liu, C.X. and Chen, L.L. (2022) Circular RNAs: characterization, cellular roles, and applications. *Cell* 185, 2016–2034
40. Wilusz, J.E. (2018) A 360° view of circular RNAs: From biogenesis to functions. *Wiley Interdiscip. Rev. RNA* 9, e1478
41. Kristensen, L.S. *et al.* (2019) The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691
42. Li, Y. *et al.* (2015) Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* 25, 981–984
43. Memczak, S. *et al.* (2015) Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* 10, e0141214
44. Chen, C.K. *et al.* (2021) Structured elements drive extensive circular RNA translation. *Mol. Cell* 81, 4300–4318 e4313
45. Li, X. *et al.* (2017) Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol. Cell* 67, 214–227
46. Liu, C.X. *et al.* (2022) RNA circles with minimized immunogenicity as potent PKR inhibitors. *Mol. Cell* 82, 420–434
47. Wesselhoeft, R.A. *et al.* (2019) RNA circularization diminishes immunogenicity and can extend translation duration *in vivo*. *Mol. Cell* 74, 508–520
48. Yang, Y. *et al.* (2017) Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res.* 27, 626–641
49. Szabo, L. and Salzman, J. (2016) Detecting circular RNAs: bioinformatic and experimental challenges. *Nat. Rev. Genet.* 17, 679–692
50. Graveley, B.R. (2008) Molecular biology: power sequencing. *Nature* 453, 1197–1198
51. Wilhelm, B.T. *et al.* (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239–1243
52. Yin, Q.F. *et al.* (2015) Fractionation of non-polyadenylated and ribosomal-free RNAs from mammalian cells. *Methods Mol. Biol.* 1206, 69–80
53. Cui, P. *et al.* (2010) A comparison between ribo-minus RNA-seq and polyA-selected RNA-seq. *Genomics* 96, 259–265
54. Gardner, E.J. *et al.* (2012) Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev.* 26, 2550–2559
55. Yin, Q.F. *et al.* (2012) Long noncoding RNAs with snoRNA ends. *Mol. Cell* 48, 219–230
56. Marioni, J.C. *et al.* (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18, 1509–1517
57. Lu, T. *et al.* (2015) Transcriptome-wide investigation of circular RNAs in rice. *RNA* 21, 2076–2087
58. Ma, X.K. *et al.* (2019) CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics Proteomics Bioinformatics* 17, 511–521
59. Zhang, J. *et al.* (2020) Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat. Commun.* 11, 90
60. Chuang, T.J. *et al.* (2018) Integrative transcriptome sequencing reveals extensive alternative *trans*-splicing and *cis*-backsplicing in human cells. *Nucleic Acids Res.* 46, 3671–3691
61. Jeck, W.R. and Sharpless, N.E. (2014) Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32, 453–461
62. Yu, C.Y. *et al.* (2014) Is an observed non-co-linear RNA product spliced *in trans*, *in cis* or just *in vitro*? *Nucleic Acids Res.* 42, 9410–9423
63. Zhang, Y. *et al.* (2016) Characterization of circular RNAs. *Methods Mol. Biol.* 1402, 215–227
64. Nielsen, A.F. *et al.* (2022) Best practice standards for circular RNA research. *Nat. Methods* 19, 1208–1220
65. Dodbele, S. *et al.* (2021) Best practices to ensure robust investigation of circular RNAs: pitfalls and tips. *EMBO Rep.* 22, e52072
66. Rebollo, C. *et al.* (2023) Computational approaches for circRNAs prediction and *in silico* characterization. *Brief. Bioinform.* 24, bbad154
67. Kim, D. and Salzberg, S.L. (2011) TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* 12, R72
68. Dobin, A. *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21
69. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760
70. Gao, Y. *et al.* (2018) Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* 19, 803–810
71. Cheng, J. *et al.* (2016) Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* 32, 1094–1096

72. Wang, K. *et al.* (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* 38, e178
73. Pertea, M. *et al.* (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295
74. Trapnell, C. *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578
75. Szabo, L. *et al.* (2015) Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol.* 16, 126
76. Chuang, T.J. *et al.* (2016) 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.* 44, e29
77. Hansen, T.B. *et al.* (2016) Comparison of circular RNA prediction tools. *Nucleic Acids Res.* 44, e58
78. Hansen, T.B. (2018) Improved circRNA identification by combining prediction algorithms. *Front. Cell Dev. Biol.* 6, 20
79. Gaffo, E. *et al.* (2022) Sensitive, reliable and robust circRNA detection from RNA-seq with CirComPara2. *Brief Bioinform.* 23, bbab418
80. Li, L. *et al.* (2019) CircRNAwrap – a flexible pipeline for circRNA identification, transcript prediction, and abundance estimation. *FEBS Lett.* 593, 1179–1189
81. Vromman, M. *et al.* (2023) Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nat. Methods* 20, 1159–1169
82. Rahimi, K. *et al.* (2021) Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nat. Commun.* 12, 4825
83. Zhang, J. *et al.* (2021) Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat. Biotechnol.* 39, 836–845
84. Xin, R. *et al.* (2021) isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat. Commun.* 12, 266
85. Liu, Z. *et al.* (2021) circFL-seq reveals full-length circular RNAs with rolling circular reverse transcription and nanopore sequencing. *Elife* 10, e69457
86. Zhang, X.O. *et al.* (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287
87. Pan, X. and Xiong, K. (2015) PredcircRNA: computational classification of circular RNA from other long non-coding RNA using hybrid features. *Mol. BioSyst.* 11, 2219–2226
88. Pan, X. *et al.* (2018) WebCircRNA: classifying the circular RNA potential of coding and noncoding RNA. *Genes (Base)* 9, 536
89. Rahimi, K. *et al.* (2021) Nanopore long-read sequencing of circRNAs. *Methods* 196, 23–29
90. Salzman, J. *et al.* (2013) Cell-type specific features of circular RNA expression. *PLoS Genet.* 9, e1003777
91. Westholm, J.O. *et al.* (2014) Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.* 9, 1966–1980
92. Rybak-Wolf, A. *et al.* (2015) Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885
93. Xia, S. *et al.* (2018) CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.* 46, D925–D929
94. Feng, J. *et al.* (2022) CSCD2: an integrated interactional database of cancer-specific circular RNAs. *Nucleic Acids Res.* 50, D1179–D1183
95. Ruan, H. *et al.* (2019) Comprehensive characterization of circular RNAs in ~1000 human cancer cell lines. *Genome Med.* 11, 55
96. Xia, S. *et al.* (2017) Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief. Bioinform.* 18, 984–992
97. Dong, R. *et al.* (2018) CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinformatics* 16, 226–233
98. Wu, W. *et al.* (2020) CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.* 21, 101
99. Conn, V. and Conn, S.J. (2019) SplintQuant: a method for accurately quantifying circular RNA transcript abundance without reverse transcription bias. *RNA* 25, 1202–1210
100. Dahl, M. *et al.* (2018) Enzyme-free digital counting of endogenous circular RNA molecules in B-cell malignancies. *Lab. Investig.* 98, 1657–1669
101. Kristensen, L.S. *et al.* (2018) Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. *RNA Biol.* 15, 280–291
102. Garcia-Rodriguez, J.L. *et al.* (2023) Spatial profiling of circular RNAs in cancer reveals high expression in muscle and stromal cells. *Cancer Res.* Published online July 21, 2023. <https://doi.org/10.1158/0008-5472.CAN-23-0748>
103. Gao, X. *et al.* (2022) Knockout of circRNAs by base editing back-splice sites of circularized exons. *Genome Biol.* 23, 16
104. Li, S. *et al.* (2021) Screening for functional circular RNAs using the CRISPR-Cas13 system. *Nat. Methods* 18, 51–59
105. Conesa, A. *et al.* (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13
106. Li, M. *et al.* (2017) Quantifying circular RNA expression from RNA-seq data using model-based framework. *Bioinformatics* 33, 2131–2139
107. Fan, X. *et al.* (2015) Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol.* 16, 148
108. Wu, W. *et al.* (2022) Exploring the cellular landscape of circular RNAs using full-length single-cell RNA sequencing. *Nat. Commun.* 13, 3242
109. Mercer, T.R. *et al.* (2015) Genome-wide discovery of human splicing branchpoints. *Genome Res.* 25, 290–303
110. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74