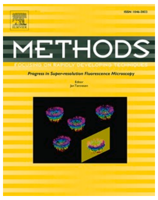
Methods 196 (2021) 3–10

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth



CIRCexplorer pipelines for circRNA annotation and quantification from non-polyadenylated RNA-seq datasets

Xu-Kai Ma [a], Wei Xue [a], Ling-Ling Chen [b,c,d], Li Yang [a,c,*]

[a] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

[b] State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

[c] School of Life Science and Technology, ShanghaiTech University, 393 Middle Huaxia Road, Shanghai 201210, China

[d] School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

ARTICLE INFO

Keywords:
Circular RNA
CircRNA
CIRCexplorer
Alternative back-splicing
Non-polyadenylated RNA-seq

ABSTRACT

Covalently closed circular RNAs (circRNAs) produced by back-splicing of exon(s) are co-expressed with their cognate linear RNAs from the same gene loci. Most circRNAs are fully overlapped with their cognate linear RNAs in sequences except the back-spliced junction (BSJ) site, thus challenging the computational detection, experimental validation and hence functional evaluation of circRNAs. Nevertheless, specific bioinformatic pipelines were developed to identify fragments mapped to circRNA-featured BSJ sites, and circRNAs were pervasively identified from non-polyadenylated RNA-seq datasets in different cell lines/tissues and across species. Precise identification and quantification of circRNAs provide a basis to further understand their functions. Here, we describe detailed computational steps to annotate and quantify circRNAs using a series of CIRCexplorer pipelines.

1. Introduction

Various types of covalently closed circular RNA transcripts are produced by different mechanisms [1–3]. Among them, two major subgroups of circular RNAs are characteristically processed from eukaryotic RNA precursors in a spliceosome-dependent manner: circular RNAs (circRNAs) from back-splicing of exons and circular intronic RNAs (ciRNAs) from excised intron lariats [1,4,5]. Although discovered as early as in 1990s [6], only a handful of these spliceosome-dependent circular RNAs were reported in following decades [7–10]. Mainly owing to their circular formation without 3' polyadenylated tails, most circular RNAs were excluded in the early days transcriptomic profiling by deep sequencing of polyadenylated [poly(A)+] RNAs, referred to as mRNA-seq [11,12]. They have been recently re-discovered by examining transcriptomic RNAs without polyadenylated tails [13,14]. Understanding the biogenesis and potential biological significance of circRNAs from back-spliced exons, which are widely expressed, has become the focus in the field [1,15–18].

Different types of RNA-seq datasets have been utilized to profile non-polyadenylated RNAs [poly(A)−], non-polyadenylated RNAs together with polyadenylated ones [ribo−], or non-polyadenylated circular RNAs after RNase R treatment [RNaseR], respectively (Fig. 1A) [19–21]. Although all three types of datasets are applicable to identify circRNAs [22], distinct groups of RNAs could be detected in these RNA-seq datasets owing to different strategies of RNA enrichments. On one hand, most linear polyadenylated RNAs are found in the poly(A)+ RNA-seq datasets; while circRNAs are found to be with other subgroups of non-polyadenylated RNAs, such as sno-lncRNAs [23,24] and ciRNAs [25] in the poly(A)− RNA-seq datasets (Fig. 1A)[14,21]. On the other hand, non-polyadenylated (circ)RNAs are co-purified with polyadenylated RNAs in the ribo− RNA-seq analysis, or are largely reserved by RNase R digestion in RNaseR RNA-seq datasets (Fig. 1A) [22]. Of note, despite the fact that RNase R is generally used to enrich circular RNAs [22,26–28], some circRNAs were found to be sensitive to RNase R treatment [22,29].

With specific computational approaches to identify fragments mapped to back-spliced junction (BSJ) sites [26–28,30,31], hundreds of thousands of circRNAs have been computationally detected in different cell lines/tissues and across species [15,32–35]. Subsequent studies have shown unique features of circRNAs in biogenesis [36–39],

* Corresponding author at: CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China.

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

https://doi.org/10.1016/j.ymeth.2021.02.008
Received 30 November 2020; Received in revised form 2 February 2021; Accepted 5 February 2021
Available online 12 February 2021

structure and degradation [40]. Importantly, lines of evidence suggest that circRNAs play important roles in cell proliferation [41,42], brain function [43], and innate immune response [40,44]. It has been shown that most circRNAs are produced from middle exons of annotated genes [39], resulting in the overlapped sequences between circRNAs and their cognate linear RNAs, except circRNA-characteristic BSJ sites (Fig. 1B). Such sequence similarities bring challenges to characterize circRNA biogenesis and function because most circRNAs are co-expressed with their cognate linear RNAs [39]. Nonetheless, specific and reliable computational pipelines are required for circRNA identification and annotation in order to provide the basis for the study of circRNA biogenesis and function.

Almost all these computational pipelines, such as find_circ [45], CIRI [46] and CIRCexplorer (version one) [39], were set to identify deep sequencing fragments specifically aligned to BSJ sites for circRNA annotation in early days (Table 1). To identify the complex alternative back-splicing regulation and unique internal splicing regulation within circRNAs, other pipelines, such as CIRI-AS [47], CIRCexplorer2 [35], and CircSplice [48], were updated or developed to inspect alternative circularization, a phenomenon that multiple circRNAs are produced from one gene locus including inner alternative splicing events (Table 1). Most recently, to tackle the difficulty of comparing circular and cognate linear RNA expression at the same time, additional pipelines, mainly CIRCexplorer3-CLEAR [20], CIRIquant [49] and DCC [50], were constructed to quantitate the relative expression of circRNAs by normalizing to that of individual cognate linear RNAs (Table 1). In this chapter, we provide the step-by-step protocol to computationally annotate and quantify circRNAs with well-developed CIRCexplorer toolkits on a series of RNA-seq datasets in PA1 cells. Of note, in addition to HISAT2 [51] and TopHat-Fusion [52] that are shown in this analysis, other aligners, such as STAR [53] and BWA [54], can be also used in the CIRCexplorer pipeline to identify circular RNAs.

2. Materials

2.1. Hardware and software

1. 64-bit computer running Linux
2. 8 GB of RAM (16 GB preferred)
3. CIRCexplorer2

CIRCexplorer2 is available at https://github.com/YangLab/CIRCexplorer2, and the detailed document is at https://circexplorer2.readthedocs.io/en/latest. This chapter is based on the version 2.4.0.

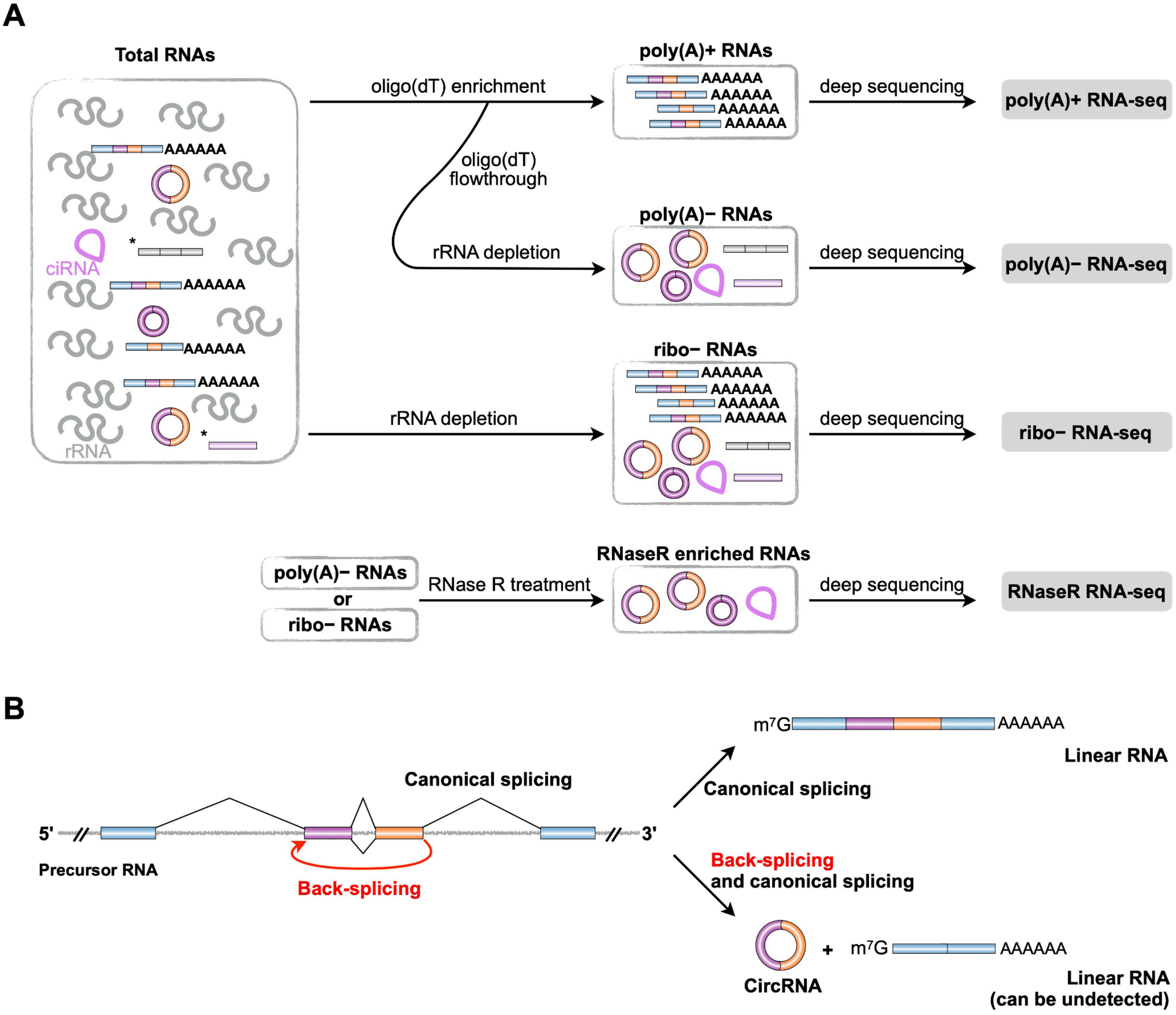


Fig. 1. Illustration of different types of RNA-seq datasets for circRNA annotation.

Table 1
Summary of computational pipelines for circRNA annotation and quantification. Names, aligners, links and references of these pipelines are provided.

Pipeline	Mapper	URL	Reference
I. General annotation of circRNAs with BSJ sites			
ACFS	BWA	https://github.com/arthuryxt/acfs	[64]
CIRCexplorer	TopHat/STAR	https://github.com/YangLab/CIRCexplorer	[39]
circRNA_finder	STAR	https://github.com/orzechoj/circRNA_finder	[68]
CIRI2	BWA	https://sourceforge.net/projects/ciri/files/CIRI2	[46]
find_circ	Bowtie2	https://github.com/rajewsky-lab/find_circ	[45]
KNIFE	Bowtie2	https://github.com/lindaszabo/KNIFE	[69]
MapSplice	Bowtie	http://www.netlab.uky.edU/p/bioinfo/MapSplice2	[70]
Uroborus	TopHat	https://github.com/WGLab/UROBORUS	[71]
II. Alternative (back-) splicing landscape of circRNAs			
CIRCexplorer2	TopHat/ STAR/ MapSplice/ BWA/ segemehl	https://github.com/YangLab/CIRCexplorer2	[35]
CIRI-AS	/	https://sourceforge.net/projects/ciri/files/CIRI-AS	[47]
CircSplice	STAR	https://github.com/GeneFeng/CircSplice	[48]
III. Direct comparison of circRNAs with linear RNAs			
CLEAR	TopHat/ STAR/ MapSplice/ BWA/ segemehl	https://github.com/YangLab/CLEAR	[20]
CIRIquant	BWA	https://github.com/Kevinzjy/CIRIquant	[49]
DCC	STAR	https://github.com/dieterich-lab/DCC	[50]
Sailfish-cir	/	https://github.com/zerodel/sailfish-cir	[72]

2.2. *CIRCexplorer3-CLEAR*

CIRCexplorer3-CLEAR is available at https://github.com/YangLab/CLEAR. This chapter is based on the version 1.0.1.

2.3. *Python 2.7*

The python can be fetched from https://www.python.org.

2.4. *Perl 5*

The perl can be fetched from https://www.perl.org.

2.5. *Bowtie*

Bowtie [55] can be downloaded from http://bowtie-bio.sourceforge.net/index.shtml, and the manual of Bowtie is at http://bowtie-bio.sourceforge.net/manual.shtml. This chapter is based on the version 0.12.9.

2.6. *Bowtie2*

Bowtie2 [56] can be downloaded from http://bowtie-bio.sourceforge.net/bowtie2/index.shtml, and the manual of Bowtie2 is at http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml. This chapter is based on the version 2.2.9.

2.7. *Hisat2*

HISAT2 [51] can be downloaded from http://daehwankimlab.github.io/hisat2/download/, and the manual of HISAT2 is at http://daehwankimlab.github.io/hisat2/manual/. This chapter is based on the version 2.0.5.

2.8. *TopHat2 and TopHat-Fusion*

The TopHat2 [52] can be downloaded from https://ccb.jhu.edu/software/tophat, and the information of TopHat2 is at https://ccb.jhu.edu/software/tophat/manual.shtml. This chapter is based on the version 2.0.12.

2.9. *StringTie*

The StringTie [57] can be downloaded from https://ccb.jhu.edu/software/stringtie, and the information of StringTie is at https://ccb.jhu.edu/software/stringtie/index.shtml?t=manual. This chapter is based on the version 1.3.6.

2.10. *Samtools*

The samtools [58] can be fetched from https://sourceforge.net/projects/samtools/files/samtools. This chapter is based on the version 0.1.19.

2.11. *BEDtools*

The BEDTools [59] can be fetched from http://bedtools.readthedocs.io/en/latest. This chapter is based on the version 2.26.0.

2.12. *RegTools*

The RegTools [60] can be downloaded from https://github.com/griffithlab/regtools. This chapter is based on the version 0.5.2.

2.13. *UCSC utilities*

UCSC utilities, including genePredToGtf, gtfToGenePred, bedGraphToBigWig, and bedToBigBed, are available at https://hgdownload.soe.ucsc.edu/admin/exe.

2.14. *Other python-related packages*

Python-related packages, including pysam, pybedtools, docopt, and scipy, are available at https://pypi.python.org.

2.15. *Reference genome and gene annotation*

1. hg38.fa

"hg38.fa" contains human primary reference genome sequence (version GRCh38/hg38), which can be downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.chromFa.tar.gz.

2. hg38_gencode.gtf

"hg38_gencode.gtf" is a General Transfer Format (GTF) file for gene annotation, which can be downloaded from ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_31/gencode.v31.annotation.gtf.gz.

This chapter is based on the version 31.

2.16. *RNA-Seq datasets*

Previously-published poly(A)+, poly(A)− and ribo− RNA-seq datasets of PA1 cells are used in this chapter [35,37], which can be downloaded from NCBI GEO (GSE75733 and GSE73325).

3. Analysis of circRNAs

3.1. Selection of different types of RNA-seq datasets

Given the fact that different repertories of RNA molecules are isolated for poly(A)+, poly(A)−, ribo−, or RNaseR RNA-seq (Fig. 1), it is important to choose appropriate types of RNA-seq together with suitable pipelines for designed circRNA profiling. Generally speaking, poly(A)+ RNA-seq datasets are not expected for circRNA detection, despite that some circRNAs might be unintentionally examined in poly(A)+ RNA-seq datasets, possibly due to non-specific binding of circRNAs with oligo(dT) beads. In contrast, all types of poly(A)−, ribo−, or RNaseR RNA-seq datasets can be used for circRNA analyses with the series of CIRCexplorer pipelines (Fig. 2A). To precisely detect uniquely back-spliced exons and the internal landscape of circRNAs, paired poly(A)+ and poly (A)− datasets are suggested to be compared by taking advantage of alternative (back-)splicing analysis module in the CIRCexplorer2 pipeline (Fig. 2B). Importantly, ribo− RNA-seq datasets are more applicable than other types of non-polyadenylated RNA-seq for direct expression comparison of circular and linear RNAs, such as by using CIRCexporer3-CLEAR pipeline (Fig. 2C).

3.2. Quality control of RNA-seq datasets for circRNA identification

Prior to genome-wide analysis, the quality of the RNA-seq datasets should be generally evaluated by applying specific quality control tools, such as FastQC. If needed, RNA-seq datasets can be further cleaned according to the report of quality control tools. Some other factors are also suggested to be considered before subsequent circRNA analyses. For example, the length of RNA-seq fragments affects the efficiency of fragments aligned to BSJ sites, hence having impact on the quantification of circRNAs [20]. In addition, although normally removed, some rRNAs might be still included in poly(A)−, ribo−, or RNaseR RNA-seq datasets owing to their extremely high expression. Similarly, some linear RNAs might be also detectable in RNaseR RNA-seq datasets owing to their unexpected resistance to RNase R treatment [22]. Taken together, cross-sample comparison of circRNAs should be carefully evaluated when using different sets of RNA-seq datasets for analysis.

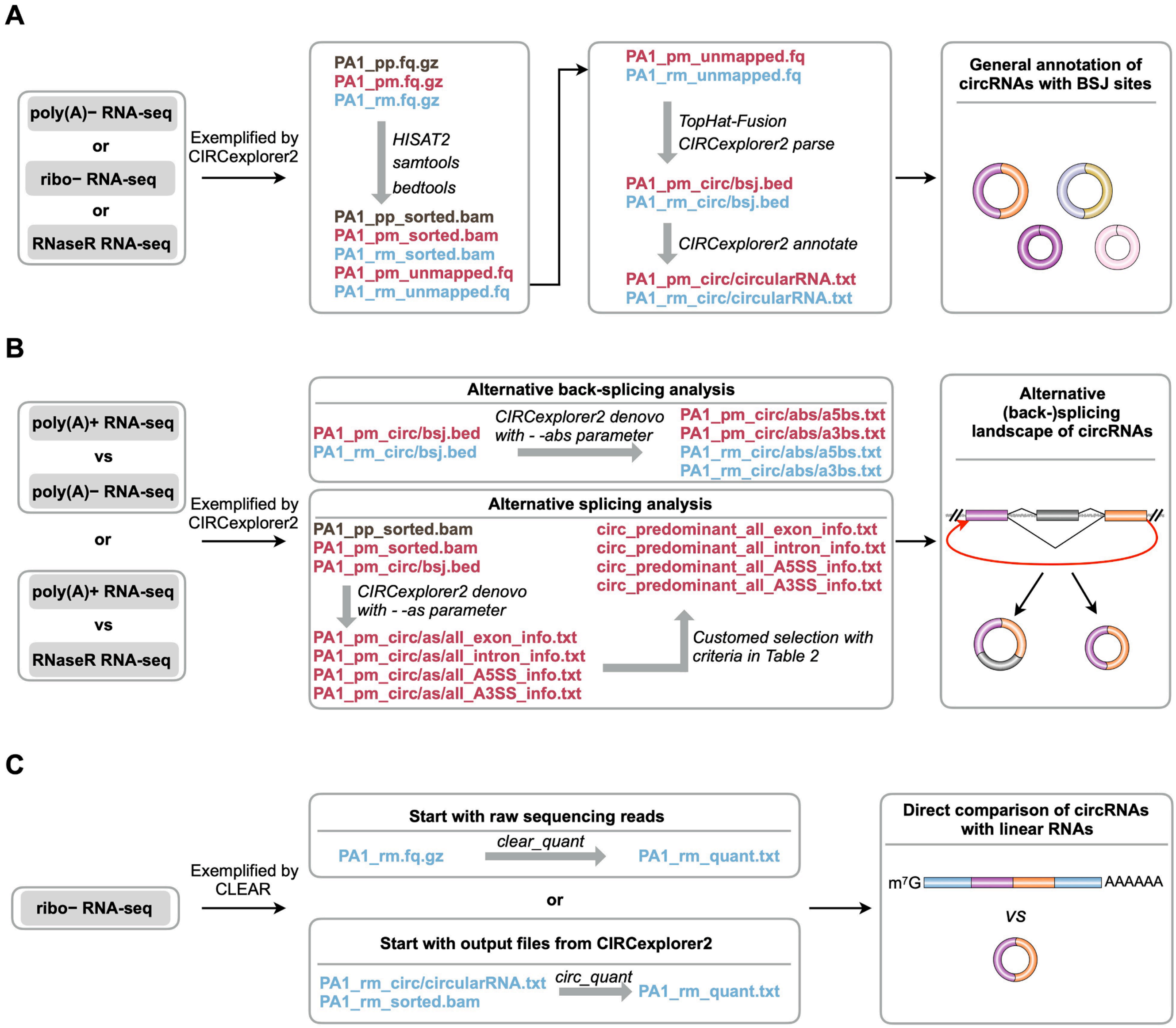

A
poly(A)− RNA-seq
or
ribo− RNA-seq
or
RNaseR RNA-seq
Exemplified by CIRCexplorer2
PA1_pp.fq.gz
PA1_pm.fq.gz
PA1_rm.fq.gz
HISAT2
samtools
bedtools
PA1_pp_sorted.bam
PA1_pm_sorted.bam
PA1_rm_sorted.bam
PA1_pm_unmapped.fq
PA1_rm_unmapped.fq
PA1_pm_unmapped.fq
PA1_rm_unmapped.fq
TopHat-Fusion
CIRCexplorer2 parse
PA1_pm_circ/bsj.bed
PA1_rm_circ/bsj.bed
CIRCexplorer2 annotate
PA1_pm_circ/circularRNA.txt
PA1_rm_circ/circularRNA.txt
General annotation of circRNAs with BSJ sites
B
poly(A)+ RNA-seq
vs
poly(A)− RNA-seq
or
poly(A)+ RNA-seq
vs
RNaseR RNA-seq
Exemplified by CIRCexplorer2
Alternative back-splicing analysis
PA1_pm_circ/bsj.bed
PA1_rm_circ/bsj.bed
CIRCexplorer2 denovo with - -abs parameter
PA1_pm_circ/abs/a5bs.txt
PA1_pm_circ/abs/a3bs.txt
PA1_rm_circ/abs/a5bs.txt
PA1_rm_circ/abs/a3bs.txt
Alternative splicing analysis
PA1_pp_sorted.bam
PA1_pm_sorted.bam
PA1_pm_circ/bsj.bed
CIRCexplorer2 denovo with - -as parameter
PA1_pm_circ/as/all_exon_info.txt
PA1_pm_circ/as/all_intron_info.txt
PA1_pm_circ/as/all_A5SS_info.txt
PA1_pm_circ/as/all_A3SS_info.txt
circ_predominant_all_exon_info.txt
circ_predominant_all_intron_info.txt
circ_predominant_all_A5SS_info.txt
circ_predominant_all_A3SS_info.txt
Customed selection with criteria in Table 2
Alternative (back-)splicing landscape of circRNAs
C
ribo− RNA-seq
Exemplified by CLEAR
Start with raw sequencing reads
PA1_rm.fq.gz
clear_quant
PA1_rm_quant.txt
or
Start with output files from CIRCexplorer2
PA1_rm_circ/circularRNA.txt
PA1_rm_sorted.bam
circ_quant
PA1_rm_quant.txt
Direct comparison of circRNAs with linear RNAs
m7G
AAAAAA
vs

Fig. 2. Flowcharts of circRNA annotation, circRNA-predominant alternative back-splicing/splicing and circRNA quantification with the series of CIRCexplorer pipelines.

3.3. Genome-wide annotation of circular RNAs

The essential step for circRNA identification is to find fragments mapped to circRNA-featured BSJ sites, adapted in all widely-used computational pipelines (Table 1). Here, we apply CIRCexplorer2 for circRNA profiling from the ribo− RNA-seq dataset in PA1 cells (Fig. 2A). First, indexes of human reference genomes (hg38.fa) are created.

Command line:

```
bowtie-build hg38.fa hg38
bowtie2-build hg38.fa hg38
hisat2-build hg38.fa hg38
```

Then, the ribo− RNA-seq dataset in PA1 cells is aligned to human reference genome (GRCh38/hg38) by aligner HISAT2. The input files contain HISAT2 genome index file (hg38), the GENCODE splice site annotation file (hg38_gencode_sp.txt) and ribo− RNA-seq dataset files (PA1_rm.fq.gz). The output file is a HISAT2 aligned result file in SAM format (PA1_rm.sam).

Command Line:

```
hisat2 –no-softclip –score-min L,-16,0 –mp 7,7 –rfg 0,7 –rdg 0,7 –dta -k 1 –max-seeds 20 -p 20 -x hg38 –known-splicesite-infile hg38_gencode_sp.txt -U PA1_rm.fq.gz -S PA1_rm.sam &>! PA1_rm_hisat2.log
```

As an intermediate step, samtools and bedtools are used to transfer the "PA1_rm.sam" to BAM format file (PA1_rm_sorted.bam) and to obtain HISAT2-unmapped fragments in FASTQ format (PA1_rm_unmapped.fq).

Command Line:

```
samtools view -Sb -f 4 PA1_rm.sam > PA1_rm_unmapped.bam
bamToFastq -i PA1_rm_unmapped.bam -fq PA1_rm_unmapped.fq
samtools view -Sb -F 4 PA1_rm.sam > PA1_rm_mapped.bam
samtools sort PA1_rm_mapped.bam PA1_rm_sorted
samtools index PA1_rm_sorted.bam
```

The HISAT2-unmapped result file (PA1_rm_unmapped.fq) is re-aligned by TopHat-Fusion to retrieve fragments aligned to BSJ sites. Except "PA1_rm_unmapped.fq", input files also include bowtie1 index file for reference genome (hg38). The output is a folder containing TopHat2 mapping results (PA1_rm_fusion).

Command Line:

```
tophat2 -o PA1_rm_fusion -p 20 –fusion-search –keep-fasta-order –bowtie1 –no-coverage-search hg38 PA1_rm_unmapped.fq &>! PA1_rm_fusion.log
```

CIRCexplorer2 is used to extract fragments mapped to BSJ sites with XF tag (*see* Note 1). The input file is TopHat-Fusion mapping result file (PA1_rm_fusion/accepted_hits.bam). The output is a new folder (PA1_rm_circ) containing fragments aligned to BSJ sites (bsj.bed).

Command line:

```
mkdir PA1_rm_circ
CIRCexplorer2 parse -f -t TopHat-Fusion PA1_rm_fusion/accepted_hits.bam -b PA1_rm_circ/bsj.bed &>! PA1_rm_parse.log
```

Finally, circRNAs are annotated with known GENCODE annotation genes. The bias from base offset around exon–intron junctions can be corrected in this step by CIRCexplorer2 embedding fine-tuning function. The input files include GENCODE gene annotation file (hg38_gencode.txt), the reference genome file (hg38.fa), and the "bsj.bed" file with aforementioned BSJ information. The output file is "circularRNA.txt" (*see* Note 2) in extended BED12 format, including circRNA chrom, circRNA start, circRNA end, circRNA name, BED12 score, strand, thickStart (same as "circRNA start"), thickEnd (same as "circRNA start"), itemRgb, exonCount, exonSizes, exonStarts information with additional six fields as fragment number that aligned to back-splicing junctions, circular RNA type (circRNA or circular intronic RNA. *see* Note 3), gene name, isoform name, exonStart, exonEnd, and circRNA-flanking intron information.

Command Line:

```
CIRCexplorer2 annotate -r hg38_gencode.txt -g hg38.fa -b PA1_rm_circ/bsj.bed -o PA1_rm_circ/circularRNA.txt &>! PA1_rm_annotate.log
```

3.4. Detection of alternative back-splicing

Currently, there are two basic strategies to investigate the alternative back-splicing of circRNAs and their internal alternative splicing events. One is to compare poly(A)+ with poly(A)− or RNaseR RNA-seq datasets to fetch circRNA predominant back-splicing/splicing events. The other one is to utilize long paired-end fragments to check the internal structures of circRNAs. Here, we use CIRCexplorer2 as an example to examine the alternative back-splicing landscape of circRNAs based on ribo− or poly(A)− RNA-seq datasets in PA1 cells (Fig. 2B).

There are two types of alternative back-splicing events, including alternative 5' back-splicing and alternative 3' back-splicing [35]. With an "–abs" parameter added into the CIRCexplorer2 annotation step, two output files, referred to as "a5bs.txt" and "a3bs.txt", are generated in a default folder (PA1_rm_circ/abs) to list all alternative 5' a 3' back-splicing related information, including circRNA chrom, circRNA start, circRNA end, strand, alternative back-splice site, back-spliced fragment counts, and Percent Circularized-site Usage (PCU).

Command line:

```
CIRCexplorer2 denovo –abs PA1_rm_circ/abs -r hg38_gencode.txt -g hg38.fa -b PA1_rm_circ/bsj.bed -o PA1_rm_circ/tmp &>! PA1_rm_abs.log
```

A typical 5' alternative back-splicing of circRNA at *FAT1* gene locus is observed in PA1 cells from both poly(A)− and ribo− RNA-seq datasets (Fig. 3A).

3.5. Detection of alternative splicing events within circRNAs

Similar to those in linear RNAs, four basic types of alternative splicing events (cassette exon, retained intron, alternative 5' splicing and alternative 3' splicing) can be observed in the internal region of circRNAs by CIRCexplorer2 [35]. In principle, CIRCexplorer2 is used to predict circRNA-predominated alternative splicing events by comparing paired poly(A)+ and poly(A)− RNA-seq datasets in PA1 cells (Fig. 2B). PA1 poly(A)− RNA-seq dataset is analyzed by commands in "3.3 Identify back-splicing junction reads for circRNA with known gene annotation" and "3.4 Detection of alternative back-splicing". Meanwhile, the PA1 poly(A)+ RNA seq dataset (PA1_pp.fq.gz) is mapped to reference genome to obtain an output file named as "PA1_pp_sorted.bam".

Alternative splicing events in circRNAs are predicted by add "–as" parameter into the CIRCexplorer2 annotate. With additional input file "PA1_pm_sorted.bam" containing poly(A)− RNA-seq dataset HISAT2 mapping result and "PA1_pp_sorted.bam" containing poly(A)+ RNA-seq dataset HISAT2 mapping result, information of all four basic type alternative splicing events will be separately exported to output files "all_exon_info.txt", "all_intron_info.txt", "all_A5SS_info.txt", and "all_A3SS_info.txt" files in the "as" folder (*see* Note 4).

Command line:

```
CIRCexplorer2 denovo –as PA1_pm_circ/as -r hg38_gencode.txt -g hg38.fa -b PA1_pm_circ/bsj.bed -m PA1_pm_sorted.bam -n PA1_pp_sorted.bam -o PA1_pm_circ/tmp &>! PA1_pm_as.log
```

With strict criteria, circRNA-predominant alternative splicing events are then determined (Table 2). Comprehensive alternative back-splicing and splicing landscape of circRNAs are available at CIRCpedia v2 from hundreds of RNA-seq samples across multiple species [61]. A typical cassette exon inclusion or exclusion in the internal region of circRNA at *TRPC1* gene locus is demonstrated by comparing poly(A)+ and poly (A)− datasets from PA1 cells (Fig. 3B).

3.6. Comparison of circular RNA and linear RNA expression

In addition to their complete sequence overlapping, different strategies are also applied to quantify circular or linear RNAs, which impedes the direct comparison of circular and linear RNA expression. A few computational pipelines, such as CIRCexplorer3-CLEAR, CIRIquant, DCC and Sailfish-cir, are built to overcome this obstacle (Table 1). Here, we use CIRCexplorer3-CLEAR as an example to realize this direct

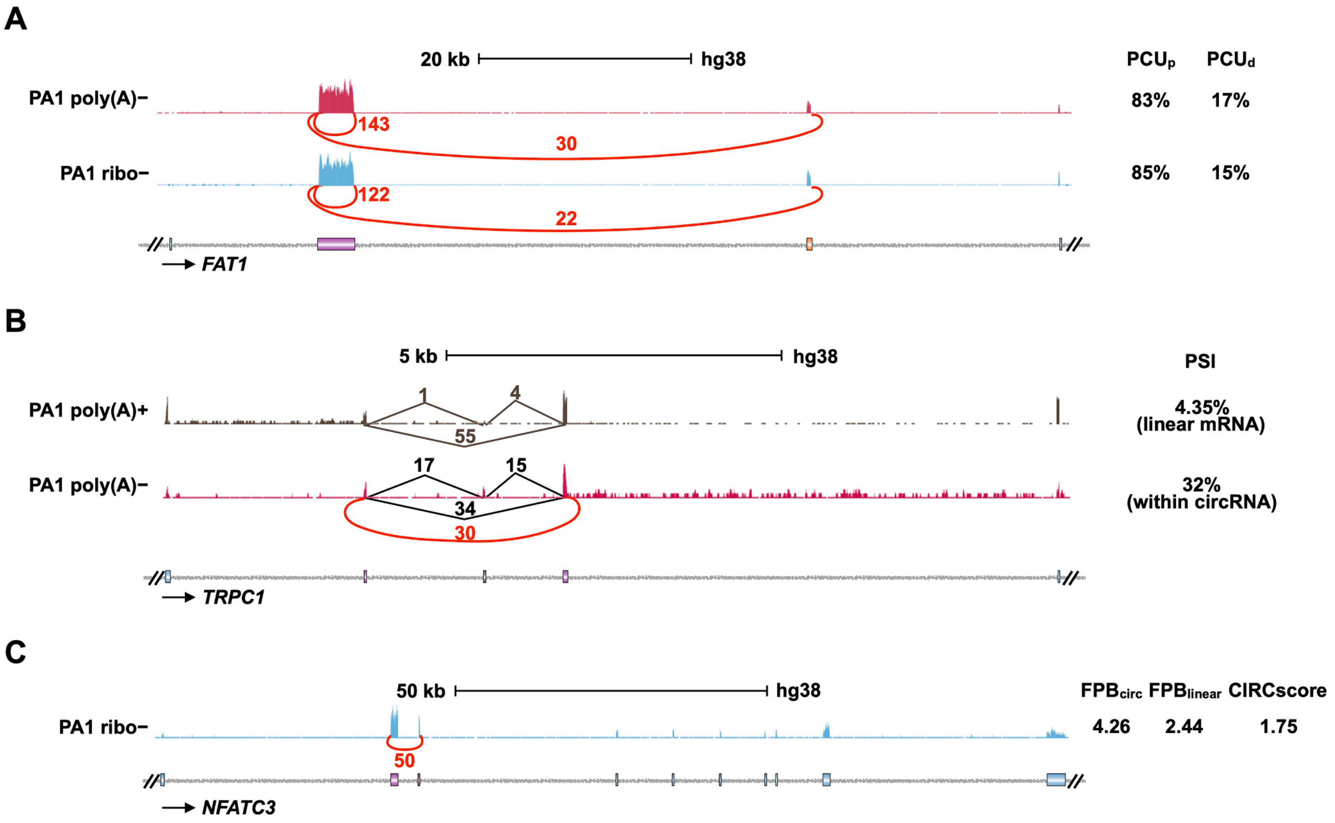


Fig. 3. Visualization of circRNAs with CIRCexplorer pipelines.

Table 2
Criteria to determine alternative splicing events within circRNAs. All four basic types of alternative splicing events can be then determined in the internal region of circRNAs.

Alternative splicing	Criteria
Cassette exons	$P_{\text{(circular percent spliced in (PSI) > linear PSI, fisher exact test)}} < 0.01$ Inclusion $\text{reads}_{\text{circular}} \geq 10$ Exclusion $\text{reads}_{\text{linear}} \geq 5$
Intron retention	Percent Intron Retention (PIR) $_{\text{circular}} > \text{PIR}_{\text{linear}}$ $P_{\text{(exon-intron reads != intron reads, binomial test)}} < 0.05$ $\text{Exon1-Intron}_{\text{circular}} + \text{Intron-Exon2}_{\text{circular}} \geq 1$ $\text{Exon1-Exon2}_{\text{linear}} \geq 5$
Alternative 5' splicing	Percent Splice site Usage (PSU) $_{\text{circular}} > \text{PSU}_{\text{linear}}$ $0 < \text{PSU}_{\text{circular}} < 100\%$ Total junction reads in splice site ≥ 5
Alternative 3' splicing	$\text{PSU}_{\text{circular}} > \text{PSU}_{\text{linear}}$ $0 < \text{PSU}_{\text{circular}} < 100\%$ Total junction reads in splice site ≥ 5

comparison (Fig. 2C) from the prevalent ribo− RNA-seq datasets in PA1 cells (*see* Note 5). CIRCexplorer3-CLEAR pipeline can either directly parse raw RNA-seq datasets in fastq format or output files from CIRCexplorer2 for the comparison. Both modes are introduced below.

To start with raw sequencing reads, the input files include PA1 ribo− RNA-seq dataset (PA1_rm.fq.gz), the reference genome HISAT2/Bowtie1 index files (hg38), and GENCODE gene annotation file (hg38_gencode.gtf).

Command line:

```
clear_quant −1 PA1_rm.fq.gz -g hg38.fa -i hg38 -j hg38 -G hg38_gencode.gtf -o PA1_rm_quant &>! PA1_rm_quant.log
```

To start with output files after CIRCexplorer2, the input files include CIRCexplorer2 annotate result file (PA1_rm_circ/circularRNA.txt), the HISAT2 mapping result file (PA1_rm_sorted.bam), and GENCODE gene annotation file (hg38_gencode.txt).

Command line:

```
circ_quant -c PA1_rm_circ/circularRNA.txt -b PA1_rm_sorted.bam -t -r hg38_gencode.txt -o PA1_rm_circ/quant.txt &>! PA1_rm_quant.log
```

Both modes lead to the same output file, “PA1_rm_quant.txt”. The “PA1_rm_quant.txt” is an extended BED12 format file containing the same columns as PA1_rm_circ/circularRNA.txt, but with three additional columns, including FPB for expression of circRNA (FPB_{circ}), FPB for expression of cognate linear RNA ($\text{FPB}_{\text{linear}}$), and CIRCscore for relative expression of circRNA ($\text{FPB}_{\text{circ}}/\text{FPB}_{\text{linear}}$). A new quantification parameter, fragments per billion mapped fragments (FPB), is invented to quantify both circular and linear RNA expression independent of sequencing strategies (paired-end or single-end RNA-seq) and read lengths. An additional new CIRCscore parameter is also introduced to direct compare circular and linear RNA expression by dividing FPB_{circ} with $\text{FPB}_{\text{linear}}$, which evaluates the relative circRNA expression with linear RNA expression as the background (illustrated at the *NFATC3* gene locus in Fig. 3C).

4. Discussion

By applying recently-developed computational pipelines to retrieve fragments mapped to BSJ sites from various types of RNA-seq datasets, a tremendous amount of circRNAs have been annotated and quantified in a tissue- and species- specific manner (*see* Note 6). In this chapter, we apply a series of CIRCexplorer pipelines as examples to illustrate how to identify and quantify circRNAs from different RNA-seq datasets. However, multiple events could potentially lead to false positives for circRNA annotation, such as the temple switching during RNA-seq library preparation and possible genetic rearrangements of exons [33,62,63]. Thus, beyond computational annotation, experimental validations, such as northern blotting and with RNase R treatment must be applied prior to selecting circRNAs of interest for the subsequent functional analyses [22,33] (*see* Note 7). Nevertheless, other than aforementioned shortread RNA-seq datasets and pipelines for circRNA detection, long-read sequencing technology has been also used to profile circRNAs with decoded full-length sequences [64–67]. In the future, the combination of the short- and long- read sequencing technologies will provide better solutions for circRNA profiling across cells and tissues.

5. Notes

1. In the CIRCexplorer2 parse step, it is suggested to add "-f" parameter to make the analysis consistent for either paired-end or single-end RNA-seq datasets. With "-f" parameter, only paired fragments, but not individual reads, are counted.

2. *De novo* assembly of circRNAs by CIRCexplorer2 is applied to detect unannotated exons for back-splicing.

3. To our knowledge, ciRNAs from intron lariats could be also identified together with circRNAs by using CIRCexplorer pipelines.

4. Detailed information for alternative splicing output files is in https://circexplorer2.readthedocs.io/en/latest/modules/denovo.

5. To achieve more reliable comparison of circRNA and linear RNA expression, ribo– RNA-seq datasets are suggested for analysis with CIRCexplorer3-CLEAR.

6. It is highly suggested to use multiple computational pipelines to achieve reliable results for circRNA analyses [26,30].

7. RNaseR RNA-seq datasets are generally used for validation of circRNA in a genome-wide scale, but not all circRNAs are inert to RNase R incubation [22].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research in Yang Lab is supported by National Natural Science Foundation of China [grant numbers 31925011, 31730111, 91940306, 31801073]; National Key R&D Program of China [grant number 2019YFA0802804]; and the Youth Innovation Promotion Association CAS from Chinese Academy of Sciences.

References

[1] L.L. Chen, The biogenesis and emerging roles of circular RNAs, Nat. Rev. Mol. Cell Biol. 17 (4) (2016) 205–211.
[2] M. Danan, S. Schwartz, S. Edelheit, R. Sorek, Transcriptome-wide discovery of circular RNAs in Archaea, Nucleic Acids Res. 40 (7) (2012) 3131–3142.
[3] E. Lasda, R. Parker, Circular RNAs: diversity of form and function, RNA 20 (12) (2014) 1829–1842.
[4] H. Wu, L. Yang, L.L. Chen, The diversity of long noncoding RNAs and their generation, Trends Genet. 33 (8) (2017) 540–552.
[5] L. Yang, Splicing noncoding RNAs from the inside out, Wiley Interdiscip. Rev. RNA 6 (6) (2015) 651–660.
[6] J.M. Nigro, K.R. Cho, E.R. Fearon, S.E. Kern, J.M. Ruppert, J.D. Oliner, K. W. Kinzler, B. Vogelstein, Scrambled exons, Cell 64 (3) (1991) 607–613.
[7] B. Capel, A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow, R. Lovell-Badge, Circular transcripts of the testis-determining gene Sry in adult mouse testis, Cell 73 (5) (1993) 1019–1030.
[8] C. Cocquerelle, P. Daubersies, M.A. Majerus, J.P. Kerckaert, B. Bailleul, Splicing with inverted order of exons occurs proximal to large introns, EMBO J. 11 (3) (1992) 1095–1098.
[9] C. Cocquerelle, B. Mascrez, D. Hetuin, B. Bailleul, Mis-splicing yields circular RNA molecules, FASEB J. 7 (1) (1993) 155–160.
[10] L. Qian, M.N. Vu, M. Carter, M.F. Wilkinson, A spliced intron accumulates as a lariat in the nucleus of T cells, Nucleic Acids Res. 20 (20) (1992) 5345–5350.
[11] B.R. Graveley, Molecular biology: power sequencing, Nature 453 (7199) (2008) 1197–1198.
[12] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5 (7) (2008) 621–628.
[13] J. Salzman, C. Gawad, P.L. Wang, N. Lacayo, P.O. Brown, Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types, PLoS ONE 7 (2) (2012), e30733.
[14] L. Yang, M.O. Duff, B.R. Graveley, G.G. Carmichael, L.L. Chen, Genomewide characterization of non-polyadenylated RNAs, Genome Biol. 12 (2) (2011) R16.
[15] L.L. Chen, The expanding regulatory mechanisms and cellular functions of circular RNAs, Nat. Rev. Mol. Cell Biol. 21 (8) (2020) 475–490.
[16] L.S. Kristensen, M.S. Andersen, L.V.W. Stagsted, K.K. Ebbesen, T.B. Hansen, J. Kjems, The biogenesis, biology and characterization of circular RNAs, Nat. Rev. Genet. 20 (11) (2019) 675–691.
[17] J.E. Wilusz, A 360 degrees view of circular RNAs: From biogenesis to functions, Wiley Interdiscip Rev RNA 9 (4) (2018), e1478.
[18] M.S. Xiao, Y. Ai, J.E. Wilusz, Biogenesis and functions of circular RNAs come into focus, Trends Cell Biol. 30 (3) (2020) 226–240.
[19] R. Dong, X.K. Ma, L.L. Chen, L. Yang, Genome-wide annotation of circRNAs and their alternative back-splicing/splicing with CIRCexplorer pipeline, Methods Mol. Biol. 1870 (2019) 137–149.
[20] X.K. Ma, M.R. Wang, C.X. Liu, R. Dong, G.G. Carmichael, L.L. Chen, L. Yang, CIRCexplorer3: A CLEAR pipeline for direct comparison of circular and linear RNA expression, Genom. Proteomics Bioinf. 17 (5) (2019) 511–521.
[21] Q.F. Yin, L.L. Chen, L. Yang, Fractionation of non-polyadenylated and ribosomal-free RNAs from mammalian cells, Methods Mol. Biol. 1206 (2015) 69–80.
[22] Y. Zhang, L. Yang, L.L. Chen, Characterization of circular RNAs, Methods Mol. Biol. 1402 (2016) 215–227.
[23] Y.H. Xing, R.W. Yao, Y. Zhang, C.J. Guo, S. Jiang, G. Xu, R. Dong, L. Yang, L. L. Chen, SLERT regulates DDX21 rings associated with pol I transcription, Cell 169 (4) (2017) 664–678 e16.
[24] Q.F. Yin, L. Yang, Y. Zhang, J.F. Xiang, Y.W. Wu, G.G. Carmichael, L.L. Chen, Long noncoding RNAs with snoRNA ends, Mol. Cell 48 (2) (2012) 219–230.
[25] Y. Zhang, X.O. Zhang, T. Chen, J.F. Xiang, Q.F. Yin, Y.H. Xing, S. Zhu, L. Yang, L. L. Chen, Circular intronic long noncoding RNAs, Mol. Cell 51 (6) (2013) 792–806.
[26] T.B. Hansen, M.T. Veno, C.K. Damgaard, J. Kjems, Comparison of circular RNA prediction tools, Nucleic Acids Res. 44 (6) (2016), e58.
[27] W.R. Jeck, N.E. Sharpless, Detecting and characterizing circular RNAs, Nat. Biotechnol. 32 (5) (2014) 453–461.
[28] L. Szabo, J. Salzman, Detecting circular RNAs: bioinformatic and experimental challenges, Nat. Rev. Genet. 17 (11) (2016) 679–692.
[29] M.S. Xiao, J.E. Wilusz, An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends, Nucleic Acids Res. 47 (16) (2019) 8755–8769.
[30] T.B. Hansen, Improved circRNA identification by combining prediction algorithms, Front. Cell Dev. Biol. 6 (2018) 20.
[31] Y. Gao, F. Zhao, Computational strategies for exploring circular RNAs, Trends Genet. 34 (5) (2018) 389–400.
[32] P. Ji, W. Wu, S. Chen, Y. Zheng, L. Zhou, J. Zhang, H. Cheng, J. Yan, S. Zhang, P. Yang, F. Zhao, Expanded expression landscape and prioritization of circular RNAs in mammals, Cell Rep. 26 (12) (2019) 3444–3460 e5.
[33] X. Li, L. Yang, L.L. Chen, The biogenesis, functions, and challenges of circular RNAs, Mol. Cell 71 (3) (2018) 428–442.
[34] A. Rybak-Wolf, C. Stottmeister, P. Glazar, M. Jens, N. Pino, S. Giusti, M. Hanan, M. Behm, O. Bartok, R. Ashwal-Fluss, M. Herzog, L. Schreyer, P. Papavasileiou, A. Ivanov, M. Ohman, D. Refojo, S. Kadener, N. Rajewsky, Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed, Mol. Cell. 58 (5) (2015) 870–885.
[35] X.O. Zhang, R. Dong, Y. Zhang, J.L. Zhang, Z. Luo, J. Zhang, L.L. Chen, L. Yang, Diverse alternative back-splicing and alternative splicing landscape of circular RNAs, Genome Res. 26 (9) (2016) 1277–1287.
[36] W. Xue, X.-K. Ma, L. Yang, Fast and Furious: insights of back splicing regulation during nascent RNA synthesis, SCIENCE CHINA Life Sciences 2021 doi:10.1007/s11427-020-1881-1.
[37] Y. Zhang, W. Xue, X. Li, J. Zhang, S. Chen, J.L. Zhang, L. Yang, L.L. Chen, The biogenesis of nascent circular RNAs, Cell Rep 15 (3) (2016) 611–624.
[38] L.L. Chen, L. Yang, Regulation of circRNA biogenesis, RNA Biol. 12 (4) (2015) 381–388.
[39] X.O. Zhang, H.B. Wang, Y. Zhang, X. Lu, L.L. Chen, L. Yang, Complementary sequence-mediated exon circularization, Cell 159 (1) (2014) 134–147.
[40] C.X. Liu, X. Li, F. Nan, S. Jiang, X. Gao, S.K. Guo, W. Xue, Y. Cui, K. Dong, H. Ding, B. Qu, Z. Zhou, N. Shen, L. Yang, L.L. Chen, Structure and degradation of circular RNAs regulate PKR activation in innate immunity, Cell 177 (4) (2019) 865–880 e21.
[41] S. Chen, V. Huang, X. Xu, J. Livingstone, F. Soares, J. Jeon, Y. Zeng, J.T. Hua, J. Petricca, H. Guo, M. Wang, F. Yousif, Y. Zhang, N. Donmez, M. Ahmed, S. Volik, A. Lapuk, M.L.K. Chua, L.E. Heisler, A. Foucal, N.S. Fox, M. Fraser, V. Bhandari, Y.J. Shiah, J. Guan, J. Li, M. Orain, V. Picard, H. Hovington, A. Bergeron, L. Lacombe, Y. Fradet, B. Tetu, S. Liu, F. Feng, X. Wu, Y.W. Shao, M.A. Komor, C. Sahinalp, C. Collins, Y. Hoogstrate, M. de Jong, R.J.A. Fijneman, T. Fei, G. Jenster, T. van der Kwast, R.G. Bristow, P.C. Boutros, H.H. He, Widespread and Functional RNA Circularization in Localized Prostate Cancer, Cell 176(4) (2019) 831-843 e22.
[42] S.Q. Li, X. Li, W. Xue, L. Zhang, L. Yang, L.Z. Yang, S.M. Cao, Y.N. Lei, C.X. Liu, S. K. Guo, L. Shan, M. Wu, X. Tao, J.L. Zhang, X. Gao, J. Zhang, J. Wei, J.S. Li, L. Yang, L.L. Chen, Screening for functional circular RNAs using the CRISPR-Cas13 system, Nat. Methods 18 (1) (2021) 51–59.
[43] M. Piwecka, P. Glazar, L.R. Hernandez-Miranda, S. Memczak, S.A. Wolf, A. Rybak-Wolf, A. Filipchyk, F. Klironomos, C.A. Cerda Jara, P. Fenske, T. Trimbuch, V. Zywitza, M. Plass, L. Schreyer, S. Ayoub, C. Kocks, R. Kuhn, C. Rosenmund, C. Birchmeier, N. Rajewsky, Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function, Science 357 (6357) (2017).

[44] X. Li, C.X. Liu, W. Xue, Y. Zhang, S. Jiang, Q.F. Yin, J. Wei, R.W. Yao, L. Yang, L. L. Chen, Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection, Mol. Cell 67 (2) (2017) 214–227 e7.

[45] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L.H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, N. Rajewsky, Circular RNAs are a large class of animal RNAs with regulatory potency, Nature 495 (7441) (2013) 333–338.

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

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

[48] J. Feng, K. Chen, X. Dong, X. Xu, Y. Jin, X. Zhang, W. Chen, Y. Han, L. Shao, Y. Gao, C. He, Genome-wide identification of cancer-specific alternative splicing in circRNA, Mol. Cancer 18 (1) (2019) 35.

[49] J. Zhang, S. Chen, J. Yang, F. Zhao, Accurate quantification of circular RNAs identifies extensive circular isoform switching events, Nat. Commun. 11 (1) (2020) 90.

[50] J. Cheng, F. Metge, C. Dieterich, Specific identification and quantification of circular RNAs from sequencing data, Bioinformatics 32 (7) (2016) 1094–1096.

[51] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (4) (2015) 357–360.

[52] D. Kim, S.L. Salzberg, TopHat-Fusion: an algorithm for discovery of novel fusion transcripts, Genome Biol. 12 (8) (2011) R72.

[53] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (1) (2013) 15–21.

[54] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760.

[55] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (3) (2009) R25.

[56] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359.

[57] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (3) (2015) 290–295.

[58] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.S. Durbin, Genome project data processing, the sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.

[59] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (2010) 841–842.

[60] K.C. Cotto, Y.-Y. Feng, A. Ramu, Z.L. Skidmore, J. Kunisaki, M. Richters, S. Freshour, Y. Lin, W.C. Chapman, R. Uppaluri, R. Govindan, O.L. Griffith, M. Griffith, RegTools: Integrated analysis of genomic and transcriptomic data for the discovery of splicing variants in cancer, bioRxiv (2021) 436634.

[61] R. Dong, X.K. Ma, G.W. Li, L. Yang, CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison, Genomics Proteomics Bioinf. 16 (4) (2018) 226–233.

[62] T.J. Chuang, Y.J. Chen, C.Y. Chen, T.L. Mai, Y.D. Wang, C.S. Yeh, M.Y. Yang, Y. T. Hsiao, T.H. Chang, T.C. Kuo, H.H. Cho, C.N. Shen, H.C. Kuo, M.Y. Lu, Y.H. Chen, S.C. Hsieh, T.W. Chiang, Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells, Nucleic Acids Res. 46 (7) (2018) 3671–3691.

[63] C.Y. Yu, H.J. Liu, L.Y. Hung, H.C. Kuo, T.J. Chuang, Is an observed non-co-linear RNA product spliced in trans, in cis or just in vitro? Nucleic Acids Res. 42 (14) (2014) 9410–9423.

[64] X. You, I. Vlatkovic, A. Babic, T. Will, I. Epstein, G. Tushev, G. Akbalik, M. Wang, C. Glock, C. Quedenau, X. Wang, J. Hou, H. Liu, W. Sun, S. Sambandan, T. Chen, E. M. Schuman, W. Chen, Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity, Nat. Neurosci. 18 (4) (2015) 603–610.

[65] K. Rahimi, M.T. Venø, D.M. Dupont, J. Kjems, Nanopore sequencing of full-length circRNAs in human and mouse brains reveals circRNA-specific exon usage and intron retention, BioRxiv (2019), 567164.

[66] Y. Zheng, P. Ji, S. Chen, L. Hou, F. Zhao, Reconstruction of full-length circular RNAs enables isoform-level quantification, Genome Med. 11 (1) (2019) 2.

[67] R. Xin, Y. Gao, Y. Gao, R. Wang, K.E. Kadash-Edmondson, B. Liu, Y. Wang, L. Lin, Y. Xing, isoCirc catalogs full-length circular RNA isoforms in human transcriptomes, Nat. Commun. 12 (1) (2021) 266.

[68] J.O. Westholm, P. Miura, S. Olson, S. Shenker, B. Joseph, P. Sanfilippo, S. E. Celniker, B.R. Graveley, E.C. Lai, Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation, Cell Rep 9 (5) (2014) 1966–1980.

[69] L. Szabo, R. Morey, N.J. Palpant, P.L. Wang, N. Afari, C. Jiang, M.M. Parast, C. E. Murry, L.C. Laurent, J. Salzman, Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development, Genome Biol. 16 (1) (2015) 126.

[70] K. Wang, D. Singh, Z. Zeng, S.J. Coleman, Y. Huang, G.L. Savich, X. He, P. Mieczkowski, S.A. Grimm, C.M. Perou, J.N. MacLeod, D.Y. Chiang, J.F. Prins, J. Liu, MapSplice: accurate mapping of RNA-seq reads for splice junction discovery, Nucleic Acids Res. 38 (18) (2010), e178.

[71] X. Song, N. Zhang, P. Han, B.S. Moon, R.K. Lai, K. Wang, W. Lu, Circular RNA profile in gliomas revealed by identification tool UROBORUS, Nucleic Acids Res 44 (9) (2016) e87.

[72] M. Li, X. Xie, J. Zhou, M. Sheng, X. Yin, E.A. Ko, T. Zhou, W. Gu, Quantifying circular RNA expression from RNA-seq data using model-based framework, Bioinformatics 33 (14) (2017) 2131–2139.