Chapter 10

Genome-Wide Annotation of circRNAs and Their Alternative Back-Splicing/Splicing with CIRCexplorer Pipeline

Rui Dong, Xu-Kai Ma, Ling-Ling Chen, and Li Yang

Abstract

Circular RNAs (circRNAs) derived from back-spliced exons were sporadically identified about 25 years ago, and have been recently re-discovered genome-wide across different species. Interestingly, one gene locus can generate multiple circRNAs through alternative back-splicing and/or alternative splicing, thus expanding our understanding on the diversity and complexity of transcriptomes. Precise annotation of circRNAs with their alternative back-splicing and alternative splicing events is the basis for the functional characterization of different categories of circRNAs. Here we describe a step-by-step computational scheme to annotate circRNAs from publicly available RNA sequencing datasets with the CIRCexplorer2 pipeline.

Key words CircRNA, CIRCexplorer, Computational biology, Alternative back-splicing, Alternative splicing

1 Introduction

Only a handful of back-spliced circular RNAs (circRNAs) were individually identified in the past decades [1–3], and were thought to be as by-products of spliceosome-mediated splicing errors (one type of mis-splicing), thus unlikely to play important roles in biological processes. Largely due to their covalently closed structure feature without a polyadenylated tail, circRNAs were missed in massive transcriptomic analyses of polyadenylated (m)RNAs (poly (A)+ RNA-seq) [4]. Recently, by taking advantage of biochemical enrichment of non-polyadenylated RNAs for deep sequencing (poly(A)– or Ribo– RNA-seq) and bioinformatic approaches to identify fragments mapped to back-splice sites, numerous circRNAs have been systematically uncovered as being co-expressed with their linear counterparts from various tissues/cell lines and across different species [4–11]. Importantly, increasing lines of evidence have begun to show that at least some circRNAs play a role in gene expression with distinct mechanisms of action [7, 8, 12–18]. In addition, although expressed at a low level in general, some

Narendra Wajapeyee and Romi Gupta (eds.), *Epitranscriptomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1870, https://doi.org/10.1007/978-1-4939-8808-2_10, © Springer Science+Business Media, LLC, part of Springer Nature 2019

circRNAs are more abundant than their linear (m)RNA cognates expressed from the same gene loci [5].

Back-splicing is catalyzed by the canonical spliceosomal machinery, albeit in a low efficiency [19]. Similar to canonical splicing, back-splicing for circRNA formation is also regulated by both *cis*-elements and *trans*-factors [12, 20]. Recent studies have demonstrated that the biogenesis of circRNAs is mediated by orientation-opposite complementary sequences in introns that flank back-spliced exons [9, 11, 21, 22], and is also affected by many RNA binding proteins (RBPs) [11, 15, 23, 24]. Such a multiple-level regulation by *cis*-elements and *trans*-factors on back-splicing thus leads to the phenomenon that a single gene locus can produce multiple circRNAs, referred to as alternative circularization [9]. Both alternative back-splicing and alternative splicing within circRNAs contribute to alternative circularization [4]. Specifically, alternative back-splicing (including both alternative 5′ and 3′ back-splicing found only in circRNAs) generates multiple circRNAs with different back-spliced junction sites, while alternative splicing (including all four basic types found in linear RNAs) within circRNAs results in different internal sequences between two circRNAs from the same gene locus [4]. It is worthwhile noting that both alternative back-splicing and alternative splicing are diversely regulated among different tissues/cell lines [4].

Given that a large number of circRNAs are expressed in a tissue- and cell line- specific manner, reliable circRNA annotation with computational algorithms is of great importance for the subsequent depiction of their functions in a given sample. So far, multiple computational methods have been developed to annotate circRNAs [8–10, 25, 26]. Although it has been suggested that several algorithms should ideally be combined to achieve reliable predictions, the pipeline, CIRCexplorer, developed in our lab has been reported as one of the best circRNA prediction toolkits with the most reliable outputs of circRNA prediction [26]. In addition, the upgraded CIRCexplorer2 pipeline has been further developed to annotate the complex alternative back-splicing and alternative splicing events in circRNAs [4]. Importantly, with the de novo transcript assembly embedded in CIRCexplorer2 pipeline, novel back-spliced/spliced exons in circRNAs were also revealed [4].

In this chapter, we describe a step-by-step computational scheme to illustrate how to identify circRNAs from non-polyadenylated RNA-seq datasets by CIRCexplorer2. The example we use here is specific for human (reference genome: GRCh37/hg19), but the method can be expanded to other species with related reference genomes. Of note, TopHat2 and TopHat-Fusion are used in this analysis; but many other aligners, such as STAR, can also be used for a similar analysis. The CIRCexplorer2 can be downloaded from https://github.com/YangLab/CIRCexplorer2.

2 Materials

2.1 Hardware Requirements

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

2.2 Software Requirements

1. CIRCexplorer2

   CIRCexplorer2 is available at https://github.com/YangLab/CIRCexplorer2, and the detailed document is at http://circexplorer2.readthedocs.io/en/latest/. This chapter is based on the version 2.3.

2. Python 2.7

   Information and manuals of python can be downloaded at https://www.python.org.

3. Perl 5

   Information and manuals of perl can be downloaded at https://www.perl.org/.

4. TopHat2 and TopHat-Fusion (version 2.0.9 or later).

   The latest TopHat2 version can be downloaded at https://ccb.jhu.edu/software/tophat/. And the manual of TopHat2 is at https://ccb.jhu.edu/software/tophat/manual.shtml.

5. Cufflinks (version 2.1.1 or later)

   Cufflinks can be downloaded at http://cole-trapnell-lab.github.io/cufflinks/.

6. BEDTools

   The homepage of BEDTools is at http://bedtools.readthedocs.io/en/latest/.

7. UCSC utilities

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

8. Other python-related packages

   Other python-related packages, such as pysam (version 0.8.4 or later), pybedtools (require pandas if pybedtools version $\geq$ 0.7.6), docopt, and scipy, are available at https://pypi.python.org/.

2.3 Reference Genome and RNA-Seq Datasets (Used in this Chapter)

1. hg19.fa

   "hg19.fa" contains human reference genome sequence (version GRCh37/hg19), which can be downloaded at http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz.

2. hg19_kg.gtf

"hg19_kg.gtf" is a General Transfer Format (GTF) file, which can be transferred by genePredToGtf from knownGene.txt. The human genome knownGene.txt can be downloaded at http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/knownGene.txt.gz.

3. refFlat.txt

"refFlat.txt" lists RefSeq Genes with Gene Names and can be downloaded at http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refFlat.txt.gz.

4. bowtie1_index and bowtie2_index

"bowtie1_index" is the index file for bowtie1 (TopHat-Fusion) and can be downloaded at ftp://ftp.ccb.jhu.edu/pub/data/bowtie_indexes/hg19.ebwt.zip. "bowtie2_index" is the index file for bowtie2 (TopHat2) and can be downloaded at ftp://ftp.ccb.jhu.edu/pub/data/bowtie2_indexes/hg19.zip.

5. RNA-seq files

RNA-seq files can be generated in house or downloaded from publicly available databases, such as Gene Expression Omnibus (GEO). Before subsequent analyses, RNA-seq datasets are required to be tested for quality control and preprocessing [27].

3 Methods

3.1 Genome-Wide Annotation of circRNAs by CIRCexplorer2 (Fig. 1)

1. **RNA-seq read alignment for back-spliced junctions**

With one combined command, RNA-seq dataset is mapped to reference genome with TopHat2 (default) to obtain RNA-seq reads that are mapped to the genome and the colinear exon–exon junctions, and the TopHat2-unmapped reads were then mapped by TopHat-Fusion (default) to retrieve reads that are mapped to back-spliced exon–exon junctions. In addition to TopHat2/TopHat2-Fusion, other aligners can also be used for this mapping step (*see* **Note 1**). Reads that are mapped to back-spliced exon–exon junctions are labeled with XF tag after TopHat-Fusion mapping. Only the back-spliced exon–exon junction reads that are mapped to the same chromosomes are extracted into the output "back_spliced_junction.bed" file for further study.

In this step, the input files are gene annotation GTF file ("hg19_kg.gtf"), bowtie1 index of reference genome ("bowtie1_index") for TopHat2-Fusion, bowtie2 index of reference genome ("bowtie2_index") for TopHat2 and the examined RNA-seq dataset ("pA_minus.fastq"). A "back_spliced_junction.bed" output file with reads that are mapped to back-spliced exon–exon junctions and an "alignment" output folder

containing alignment results for TopHat2 (in a "tophat" folder), for TopHat-Fusion (in a "tophat_fusion" folder) are generated. The "back_spliced_junction.bed" is a BED6 file, listing back-spliced junction chrom, back-spliced junction start, back-spliced junction end, back-spliced junction name, BED6 score (default as "0"), and strand information (*see* **Notes 2** and **3**).

Command line: *CIRCexplorer2 align -G* ***hg19_kg.gtf*** *-i* ***bowtie1_index*** *-j* ***bowtie2_index*** *-f* ***pA_minus.fastq*** *-o* ***alignment*** *-b* ***back_spliced_junction.bed***

2. **Annotation of circRNAs**

This step is to annotate circRNAs with known RefSeq genes. The input files include RefSeq Gene file ("refFlat.txt"), the reference genome file ("hg19.fa"), and the "back_spliced_junction.bed" file with aforementioned back-spliced exon–exon junction information. The output file is "circularRNA_known.txt". The "circularRNA_known.txt" is an extended BED12 format file, listing circRNA chrom, circRNA start, circRNA end, circRNA name, BED12 score, strand, thickStart (same as "circRNA start"), thickEnd (same as "circRNA start"), itemRgb (default as "0,0,0"), exonCount, exonSizes, exonStarts information with additional six fields as framgent number, circRNA type (circRNA or circular intronic RNA (ciRNA) [28]), gene name, isoform name, exonStart-exonEnd, and circRNA-flanking intron information (*see* **Note 4**).

Command line: *CIRCexplorer2 annotate -r* ***refFlat.txt*** *-g* ***hg19.fa*** *-b* ***back_spliced_junction.bed*** *-o* ***circularRNA_known.txt***

3. **Assemble circRNA transcripts**

This step is to de novo assemble circRNA transcripts with applicable non-polyadenylated RNA-seq datasets, such as poly(A)–, Ribo–, poly(A)–/RNase R, and/or Ribo–/RNase R samples.

The input files are RefSeq Gene file ("refFlat.txt") and the "alignment/tophat" folder containing TopHat2 mapping result. The output is an "assemble" folder including all de novo assembly results by Cufflinks (*see* **Note 5**).

Command line: *CIRCexplorer2 assemble -r* ***refFlat.txt*** *-m* ***alignment/tophat*** *-o* ***assemble***

4. **Annotation of circRNA with novel and known exons**

This step is to annotate circRNAs with the de novo assembled circRNA transcripts (*see* **Note 6**).

The input files are RefSeq Gene file ("refFlat.txt"), the reference genome file ("hg19.fa"), the "back_spliced_junction.bed" file, and the "assemble" folder. A new

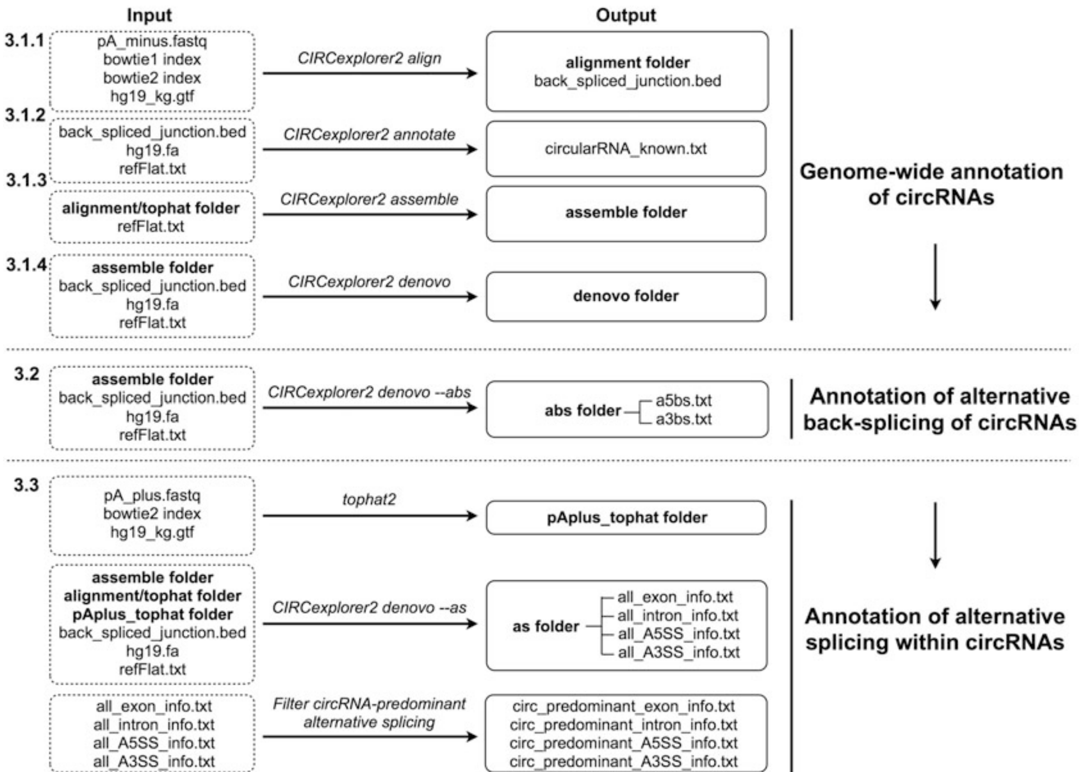

Fig. 1 Flow of circRNA annotation and alternative back-splicing/splicing analyses by CIRCexplorer pipeline

"circularRNA_full.txt" in extended BED12 format to annotate circRNAs will be generated within a new "denovo" folder. The format of "circularRNA_full.txt" is same as "circularRNA_known.txt" generated in Subheading 3.1, Step 2.

Command line: *CIRCexplorer2 denovo -r* ***refFlat.txt*** *-g* ***hg19.fa*** *-b* ***back_spliced_junction.bed*** *-d* ***assemble*** *-o* ***denovo***

3.2 Annotation of Alternative Back-Splicing of circRNAs (Figs. 1 and 2a)

Multiple circRNAs can be processed in single gene loci through alternative back-splice and/or alternative splice site selection [4]. There are two types of alternative back-splicing, alternative 5′ back-splicing and alternative 3′ back-splicing.

An "--abs" parameter is incorporated into the CIRCexplorer2 denovo step to annotate alternative back-splicing events. Two output files "a5bs.txt" and "a3bs.txt" are generated in the "abs" folder to individually list alternative 5′ or 3′ back-splicing information, including circRNA chrom, circRNA start, circRNA end, strand, alternative back-splice site, back-spliced framgment counts, and Percent Circularized-site Usage (PCU) (*see* **Note 7**).

Command line: *CIRCexplorer2 denovo --abs* ***abs*** *-r* ***refFlat.txt*** *-g* ***hg19.fa*** *-b* ***back_spliced_junction.bed*** *-d* ***assemble*** *-o* ***denovo***

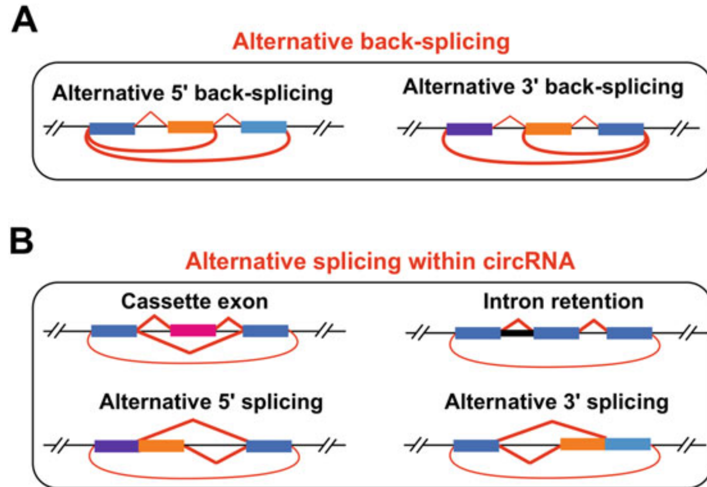

Fig. 2 Different types of alternative back-splicing/splicing of circRNAs. (**a**) Two types of alternative back-splicing of circRNAs. (**b**) Four basic types of alternative splicing within circRNAs

3.3 Annotation of Alternative Splicing Predominated in circRNAs (Figs. 1 and 2b)

All four basic types of alternative splicing (including cassette exon, retained intron, alternative 5′ splicing, alternative 3′ splicing) that are commonly identified in linear RNAs can also be identified in circRNAs. This step is to characterize canonical alternative splicing events that are predominated in circRNAs.

To annotate circRNA-predominated alternative splicing events, poly(A)+ RNA-seq dataset from the same sample is also mapped for alternative splicing analysis. Basically, poly(A)+ RNA-seq is mapped to reference genome by TopHat2. The input files are poly(A)+ RNA-seq dataset ("pA_plus.fastq"), bowtie2 index file ("bowtie2_index") for TopHat2 and the GTF annotation file (hg19_kg.gtf). The output is a new folder ("pAplus_tophat") with poly(A)+ RNA-seq dataset mapping result.

Command line: *tophat2 -a 6 --microexon-search -m 2 -p 10 -g 1 -G* ***hg19_kg.gtf*** *-o* ***pAplus_tophat bowtie2_index pA_plus.fastq***

An "--as" parameter is incorporated into the CIRCexplorer2 denovo step to annotate alternative splicing events in circRNAs. With additional input folders "alignment/tophat" containing poly (A)−/poly(A)− RNaseR RNA-seq dataset mapping result and "pAplus_tophat" containing poly(A)+ RNA-seq dataset mapping result, four output files "all_exon_info.txt", "all_intron_info.txt", "all_A5SS_info.txt", and "all_A3SS_info.txt" files with all four types of alternative splicing events are generated in the "as" folder (*see* **Note 8**).

Command line: *CIRCexplorer2 denovo --as* ***as*** *-r* ***refFlat.txt*** *-g* ***hg19.fa*** *-b* ***back_spliced_junction.bed*** *-d* ***assemble*** *-m* ***alignment/tophat*** *-n* ***pAplus_tophat*** *-o* ***denovo***

Then, circRNA-predominant alternative splicing is determined based on the following criteria:

1. Cassette exons

   $P_{(\text{circular percent spliced in (PSI)} > \text{linear PSI, fisher exact test})} < 0.01$

   Inclusion $\text{reads}_{\text{circular}} \geq 10$

   Exclusion $\text{reads}_{\text{linear}} \geq 5$

   The input file is "all_exon_info.txt" and the output file is "circ_predominant_exon_info.txt" with circRNA-predominant cassette exon.

   Command line: *perl -alne 'print if ($F[11] < 0.01 and $F[13] >= 10 and $F[16] >= 5)'* ***all_exon_info.txt*** > ***circ_predominant_exon_info.txt***

2. Intron retention

   Percent Intron Retention $(\text{PIR})_{\text{circular}} > \text{PIR}_{\text{linear}}$

   $P_{(\text{exon-intron reads} \mathrel{!=} \text{intron reads, binomial test})} < 0.05$

   $\text{Exon1-Intron}_{\text{circular}} + \text{Intron-Exon2}_{\text{circular}} \geq 1$

   $\text{Exon1-Exon2}_{\text{linear}} \geq 5$

   The input file is "all_intron_info.txt" and the output file is "circ_predominant_intron_info.txt" with circRNA-predominant retained intron.

   Command line: *perl -alne 'print if ($F[9] > $F[10] and $F[11] < 0.05 and $F[13] >= 1 and $F[17] >= 5)'* ***all_intron_info.txt*** > ***circ_predominant_intron_info.txt***

3. Alternative 5′ splicing

   Percent Splice site Usage $(\text{PSU})_{\text{circular}} > \text{PSU}_{\text{linear}}$

   $0 < \text{PSU}_{\text{circular}} < 100\%$

   Total junction reads in splice site ≥ 5

   The input file is "all_A5SS_info.txt" and the output file is "circ_predominant_A5SS_info.txt" with circRNA-predominant alternative 5′ splicing.

   Command line: *perl -alne 'print if ($F[6] > $F[9] and $F[6] >0 and $F[6] < 100 and $F[5] >= 5)'* ***all_A5SS_info.txt*** > ***circ_predominant_A5SS_info.txt***

4. Alternative 3′ splicing

   $\text{PSU}_{\text{circular}} > \text{PSU}_{\text{linear}}$

   $0 < \text{PSU}_{\text{circular}} < 100\%$

   Total junction reads in splice site ≥ 5

   The input file is "all_A3SS_info.txt" and the output file is "circ_predominant_A3SS_info.txt" with circRNA-predominant alternative 3′ splicing.

   Command line: *perl -alne 'print if ($F[6] > $F[9] and $F[6] >0 and $F[6] < 100 and $F[5] >= 5)'* ***all_A3SS_info.txt*** > ***circ_predominant_A3SS_info.txt***

3.4 CircRNA Expression Calculation

The expression of circRNAs is usually represented by the fragments that are mapped to the back-spliced exon–exon junction sites. In addition to the raw fragment numbers, normalized RNA-seq fragments that are mapped to a specific back-spliced exon–exon junction by total mapped fragments is used to quantitate circRNA expression. With FPM (Fragments mapped to back-spliced junction Per Million mapped fragments), circRNAs from different samples with distinct sequencing depths can be directly compared (*see* **Note 9**). The formula for calculating FPM is:

$$\mathrm{FPM} = \frac{\text{Fragments mapped to a specific back-spliced junction}}{\text{Total mapped fragments}} \times 10^6$$

Fragments mapped to back-spliced junctions are annotated in "circularRNA_known.txt" and "denovo/circularRNA_full.txt" files. And the total mapped fragments are the sum of TopHat2 and TopHat-Fusion mapped fragments.

3.5 Visualization of circRNAs by Using UCSC Genome Browser (Fig. 3)

1. **Visualization of sequencing data**

   In this step, a "--bw" parameter is specifically included for sequencing read visualization. Correspondingly, a bigWig file "accepted_hits.bw" is generated in "alignment/tophat" folder. The "accepted_hits.bw" can then be uploaded to UCSC genome browser for visualization.

   Command line: *CIRCexplorer2 align --bw -G* ***hg19_kg.gtf*** *-i* ***bowtie1_index*** *-j* ***bowtie2_index*** *-f* ***pA_minus.fastq*** *-o* ***alignment*** *-b* ***back_spliced_junction.bed***

2. **Visualization of circRNA**

   The "circularRNA_known.txt" and "denovo/circularRNA_-full.txt" can be converted into a BED12 format file and then

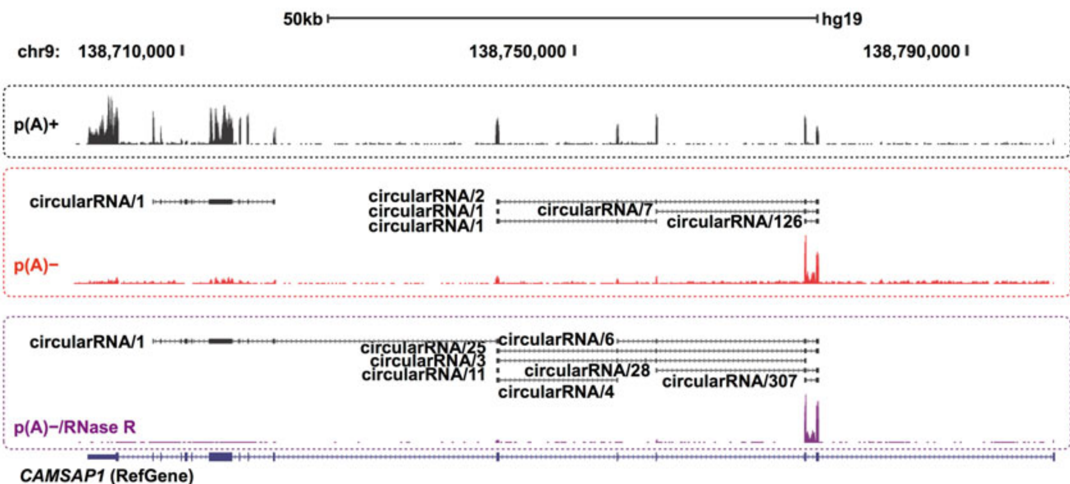

Fig. 3 Identification and visualization of circRNAs in *CAMSAP1* locus from PA1 cell line. Different types of RNA-seq datasets from p(A)+, p(A)– and p(A)–/RNase R are shown

uploaded to UCSC Genome browser for circRNA visualization.

The input file is "circularRNA_known.txt" or "denovo/circularRNA_full.txt" and the output file is "upload_circularRNA_known.bed" or "upload_circularRNA_full.bed".

Command line: *cut -f 1-12* ***circularRNA_known.txt*** > ***upload_circularRNA_known.bed;*** *cut -f 1-12* ***denovo/circularRNA_full.txt*** > ***upload_circularRNA_full.bed***

3.6 CIRCpedia: An Integrative Database of circRNAs with Detected Alternative Back-Splicing and Alternative Splicing (http:/www.picb.ac.cn/rnomics/circpedia/)

All identified alternative back-splicing and alternative splicing events in circRNAs, including newly identified exons, are available in the CIRCpedia database (http://www.picb.ac.cn/rnomics/circpedia). In this online database, multiple circRNAs produced from each individual gene locus in different cell lines can be searched, browsed, and downloaded. Currently, the database contains circRNA back-splicing and alternative splicing from six species (human, mouse, rat, zebrafish, fruitfly, and elegans) and information on a wider spectrum of cell-line, tissue, and species samples will be constructed when additional high-quality RNA-seq datasets are available.

A simple search is available from the search page of CIRCpedia. Users can query circRNA information in different cell lines and different types of back-splicing. CIRCpedia provides query support by gene symbols and genomic locations. A specific gene symbol (/genomic location) retrieves all circRNAs that have been identified in a given gene locus (/genomic location), together with relevant alternative back-splicing. In addition, users can also restrict their query to a specific cell line by different setting options. After the query, an informative table with CIRCpedia ID, species, host gene, isoform, location, strand, FPM, ExonStart-ExonEnd, seq type, cell line, conservation between human and mouse, annotation information by MapSplice alignment and enrichment fold change after RNase R treatment information will be available to check online or download for further analysis. Useful links are also available to access more information or gene descriptions in GeneCard websites.

4 Notes

1. The other optional aligners such as HISAT2, STAR, MapSplice, segmehl, and BWA can also be used for RNA-seq read alignment. CIRCexplorer2 provides additional commands to extract back-spliced exon–exon junction information for subsequent analyses. All commands are available at http://circexplorer2.readthedocs.io/en/latest/. For instance, the input file "Chimeric.out.junction" is parsed to extract

alignment results with STAR aligner, and the output file is "back_spliced_junction.bed".

Command line: *CIRCexplorer2 parse -t STAR* ***Chimeric.out.junction*** *-b* ***back_spliced_junction.bed***

2. In this step, an output file "back_spliced_junction.bed" is generated. The "score" (the color parameter, set as off) field is set as "0" to match the BED6 format.
3. The detailed information of CIRCexplorer2 output files is in http://circexplorer2.readthedocs.io/en/latest/.
4. The "name" field in "circularRNA_known.txt" is set as "Circular RNA/Junction fragments". In order to match the BED12 format, the "score" (the color parameter, set as off), "thickStart" (the coding sequence starting position at which the feature is drawn thickly), "thickEnd" (the coding sequence ending position at which the feature is drawn thickly), and "itemRgb" (the color parameter, set as off) fields are set as "0", circRNA starting position, circRNA starting position, and "0,0,0", respectively. Since circular RNAs are determined as noncoding RNAs, the "thickStart" and "thickEnd" parameters are basically useless here and both are set as circRNA starting position. Two main subfamilies of spliceosome-dependent circular RNAs, circRNAs from back-spliced exons and ciRNAs from spliced introns [29], are labeled to describe the circRNA types identified in this step.
5. In this de novo assembly step, CIRCexplorer2 ignores rRNA assembly by setting the parameter "--remove-rRNA". Currently, this parameter is only applicable with hg19 reference.
6. The last two steps are specific for de novo assembly of circRNAs. So, they are not required for circRNA annotation with known RefSeq genes.
7. The Percent Circularized-site Usage (PCU) formula is used to evaluate the percentage of examined alternative back-splicing event, defined as:

$$\mathrm{PCU} = \frac{\text{Fragments mapped to a specific back−spliced junction}}{\text{Total fragments mapped to correlated back−splice junctions}} \times 100$$

8. Both the poly(A)+ and poly(A)– (with/without RNase R) RNA-seq datasets from the same sample are required for alternative splicing analyses within circRNAs. In general, poly(A)– samples with RNase R treatment, rather than without RNase R treatment, are highly recommended in this step.
9. FPM $\geq$ 0.1 is generally used to select highly expressed circRNAs.

References

1. Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B (1991) Scrambled exons. Cell 64 (3):607–613. https://doi.org/10.1016/0092-8674(91)90244-S
2. Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell 73(5):1019–1030. https://doi.org/10.1016/0092-8674(93)90279-Y
3. Cocquerelle C, Mascrez B, Hetuin D, Bailleul B (1993) Mis-splicing yields circular RNA molecules. FASEB J 7(1):155–160. https://doi.org/10.1096/fasebj.7.1.7678559
4. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Chen LL, Yang L (2016) Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. Genome Res 26(9):1277–1287. https://doi.org/10.1101/gr.202895.115
5. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO (2013) Cell-type specific features of circular RNA expression. PLoS Genet 9(9): e1003777. https://doi.org/10.1371/journal.pgen.1003777
6. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 7 (2):e30733. https://doi.org/10.1371/journal.pone.0030733
7. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495 (7441):384–388. https://doi.org/10.1038/nature11993
8. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495(7441):333–338. https://doi.org/10.1038/nature11928
9. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L (2014) Complementary sequence-mediated exon circularization. Cell 159 (1):134–147. https://doi.org/10.1016/j.cell.2014.09.001
10. Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, Lai EC (2014) Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. Cell Rep 9(5):1966–1980. https://doi.org/10.1016/j.celrep.2014.10.062
11. Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon EY, Landthaler M, Dieterich C, Rajewsky N (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep 10(2):170–177. https://doi.org/10.1016/j.celrep.2014.12.019
12. Chen LL (2016) The biogenesis and emerging roles of circular RNAs. Nat Rev Mol Cell Biol 17(4):205–211. https://doi.org/10.1038/nrm.2015.32
13. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L, Zhu P, Chang Z, Wu Q, Zhao Y, Jia Y, Xu P, Liu H, Shan G (2015) Exon-intron circular RNAs regulate transcription in the nucleus. Nat Struct Mol Biol 22(3):256–264. https://doi.org/10.1038/nsmb.2959
14. Guarnerio J, Bezzi M, Jeong JC, Paffenholz SV, Berry K, Naldini MM, Lo-Coco F, Tay Y, Beck AH, Pandolfi PP (2016) Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. Cell 165(2):289–302. https://doi.org/10.1016/j.cell.2016.03.020
15. Li X, Liu CX, Xue W, Zhang Y, Jiang S, Yin QF, Wei J, Yao RW, Yang L, Chen LL (2017) Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. Mol Cell 67 (2):214–227 e217. https://doi.org/10.1016/j.molcel.2017.05.023
16. Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-Hernandez D, Ramberger E, Shenzis S, Samson M, Dittmar G, Landthaler M, Chekulaeva M, Rajewsky N, Kadener S (2017) Translation of CircRNAs. Mol Cell 66(1):9–21 e27. https://doi.org/10.1016/j.molcel.2017.02.021
17. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, Fatica A, Santini T, Andronache A, Wade M, Laneve P, Rajewsky N, Bozzoni I (2017) Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. Mol Cell 66 (1):22–37 e29. https://doi.org/10.1016/j.molcel.2017.02.017
18. Chen YG, Kim MV, Chen X, Batista PJ, Aoyama S, Wilusz JE, Iwasaki A, Chang HY (2017) Sensing self and foreign circular RNAs by intron identity. Mol Cell 67(2):228–238

e225. https://doi.org/10.1016/j.molcel.2017.05.022

19. Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, Bindereif A (2015) Exon circularization requires canonical splice signals. Cell Rep 10(1):103–111. https://doi.org/10.1016/j.celrep.2014.12.002
20. Chen LL, Yang L (2015) Regulation of circRNA biogenesis. RNA Biol 12(4):381–388. https://doi.org/10.1080/15476286.2015.1020271
21. Liang D, Wilusz JE (2014) Short intronic repeat sequences facilitate circular RNA production. Genes Dev 28(20):2233–2247. https://doi.org/10.1101/gad.251926.114
22. Dong R, Ma XK, Chen LL, Yang L (2017) Increased complexity of circRNA expression during species evolution. RNA Biol 14(8):1064–1074. https://doi.org/10.1080/15476286.2016.1269999
23. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S (2014) circRNA biogenesis competes with pre-mRNA splicing. Mol Cell 56(1):55–66. https://doi.org/10.1016/j.molcel.2014.08.019
24. Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ (2015) The RNA binding protein quaking regulates formation of circRNAs. Cell 160(6):1125–1134. https://doi.org/10.1016/j.cell.2015.02.014
25. Gao Y, Wang J, Zhao F (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol 16:4. https://doi.org/10.1186/s13059-014-0571-3
26. Hansen TB, Veno MT, Damgaard CK, Kjems J (2016) Comparison of circular RNA prediction tools. Nucleic Acids Res 44(6):e58. https://doi.org/10.1093/nar/gkv1458
27. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864. https://doi.org/10.1093/bioinformatics/btr026
28. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL (2013) Circular intronic long noncoding RNAs. Mol Cell 51(6):792–806. https://doi.org/10.1016/j.molcel.2013.08.017
29. Yang L (2015) Splicing noncoding RNAs from the inside out. Wiley Interdiscip Rev RNA 6(6):651–660. https://doi.org/10.1002/wrna.1307