nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-023-00815-w

Check for updates

Full-length circular RNA profiling by nanopore sequencing with CIRI-long

Lingling Hou[1,4], Jinyang Zhang[1,4]✉ and Fangqing Zhao[1,2,3]✉

Circular RNAs (circRNAs) have important roles in regulating developmental processes and disease progression. As most circRNA sequences are highly similar to their cognate linear transcripts, the current short-read sequencing-based methods rely on the back-spliced junction signal for distinguishing circular and linear reads, which does not allow circRNAs' full-length structure to be effectively reconstructed. Here we describe a long-read sequencing-based protocol, CIRI-long, for the detection of full-length circular RNAs. The CIRI-long protocol combines rolling circular reverse transcription and nanopore sequencing to capture full-length circRNA sequences. After poly(A) tailing, RNase R treatment, and size selection of polymerase chain reaction products, CIRI-long achieves an increased percentage (6%) of circular reads in the constructed library, which is 20-fold higher compared with previous Illumina-based strategies. This method can be applied in cell lines or tissue samples, enabling accurate detection of full-length circRNAs in the range of 100–3,000 bp. The entire protocol can be completed in 1 d, and can be scaled up for large-scale analysis using the nanopore barcoding kit and PromethION sequencing device. CIRI-long can serve as an effective and user-friendly protocol for characterizing full-length circRNAs, generating direct and convincing evidence for the existence of detected circRNAs. The analytical pipeline offers convenient functions for identification of full-length circRNA isoforms and integration of multiple datasets. The assembled full-length transcripts and their splicing patterns provide indispensable information to explore the biological function of circRNAs.

Introduction

Circular RNAs (circRNAs) are single-stranded, covalently closed RNA loops with a 3′–5′ phosphodiester bond generated by back-splicing events[1,2]. The emergence of next-generation sequencing technology and circRNA enrichment methods has enabled the comprehensive investigation of circRNAs across various species[3–6]. From then on, studies have profiled the abundant and conserved expression of circRNAs across species and organisms[7,8]. Recent studies have also demonstrated the important roles of circRNAs, including sponging microRNAs[9,10], interacting with RNA-binding proteins[11–13], and being translated into short peptides[14–16]. These essential roles of circRNAs rely on the regulatory sequence elements contained within the circular transcripts[17]. Although most circRNAs are derived from exonic regions[18], recent studies have also revealed that many circRNAs harbor a specific exon structure that is distinct from the structure of their messenger RNA counterparts, suggesting the unique alternative splicing mechanism of these circular transcripts[19–21]. Thus, the determination of full-length circRNA sequences has become a crucial aspect of circRNA analysis.

Here we describe our CIRI-long method for effective detection of full-length circRNA isoforms with a variety of lengths using nanopore sequencing[22]. The CIRI-long protocol implements rolling circle reverse transcription (RT) for the capture of full-length circRNA templates and includes poly(A) tailing, RNase R treatment and library size selection steps for further enrichment of circRNAs. Altogether, the CIRI-long protocol can serve as an easy-to-use protocol for the effective characterization of full-length circRNAs, providing useful resources for circRNA functional prioritization and benchmarking short-read-based computational methods[23,24].

Development of the protocol

Recent studies have developed many computational tools for detecting full-length circRNA isoforms from Illumina sequencing datasets[25–30]. Previously, we developed the CIRI-AS[19] algorithm for detecting the circular exon structure using splice-aware alignment from the back-spliced junction

[1]Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou, China. [4]These authors contributed equally: Lingling Hou, Jinyang Zhang. ✉e-mail: zhangjinyang@biols.ac.cn; zhfq@biols.ac.cn

(BSJ) read pairs. Subsequently, our CIRI-full[20] algorithm has incorporated the reverse overlap features, where the 5′ and 3′ ends on both paired reads are reversely overlapped with each other, to assemble full-length circRNAs in paired-end 250 bp data. At the same time, other methods[21,25–27] have also employed alignment- and assembly-based strategies for reconstructing full-length circRNA sequences. However, these methods were limited by the relatively short read length of Illumina sequencing, which only recovered circRNAs under a limited length of 500 bp[31]. For longer circRNAs, the BSJ and internal exon sequences cannot be fully covered in the same read pair, which is the largest limitation in reconstructing circRNAs in these methods. To this end, we describe the CIRI-long protocol, which implements a long-read nanopore sequencing method for effective detection of the full-length isoforms of circRNA with various lengths[22,32].

Unlike their mRNA counterparts, circRNAs lack a poly(A) tail. To effectively capture the full-length circRNAs without poly(A) tails, CIRI-long adopts several key modifications to the standard circRNA sequencing protocol that involves RNase R treatment and Illumina sequencing platform[33] (Fig. 1). Firstly, total RNA is extracted, and ribosomal RNAs (rRNAs) are removed using the RiboErase kit. Then, additional poly(A) tails are added to the 3′ end of ribosomal-depleted RNAs, which can increase the efficiency of RNase R digestion of structured linear RNAs[34,35]. Afterward, circRNAs are captured using rolling circular RT and Switching Mechanism at the 5′ end of RNA Template (SMART) technology. The long complementary DNA product may consist of multiple copies of the full-length circRNA sequence. Finally, long cDNA products from circRNAs are further enriched using fragment size selection, and the constructed sequencing library is sequenced using the nanopore MinION device. Here the long cDNA molecules provide both the full-length sequences of circRNAs and direct evidence of the existence of these circular structures. CIRI-long can therefore effectively detect circRNAs with canonical GT/AG splice sites, as well as other novel circRNA types that are derived from noncanonical splicing signals.

Applications of the CIRI-long protocol

The CIRI-long protocol can be used to study the isoform level changes of circRNAs in different tissues, samples or cell lines. Besides, as CIRI-long can provide strong evidence of back splicing, this protocol can also be utilized for discovering less well-studied circRNA types including mitochondrial-derived circRNAs and intron self-ligated circRNAs[22,36] (Fig. 2). In addition, the full-length circRNA sequences are essential for downstream analyses such as functional prediction[24]. The CIRI-long result can be used as input for sequence-based miRNA or RNA-binding protein (RBP) binding site prediction tools, providing an important basis for circRNA functional prioritization. Moreover, the

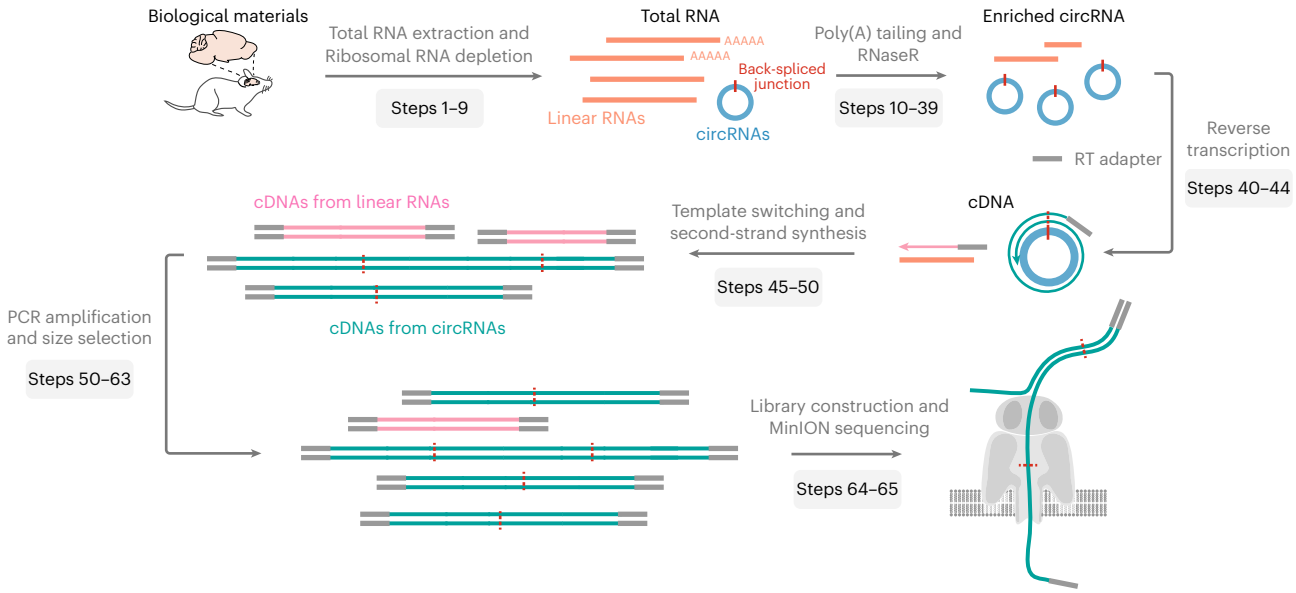


Fig. 1 | Overview of the CIRI-long protocol. Schematic diagram of the CIRI-long library construction procedures. The extracted RNAs are treated using steps including rRNA depletion, poly(A) tailing and RNase R treatment for circRNA enrichment. Afterward, rolling circular RT is performed to amplify circRNAs without exposed 3′ and 5′ ends. Then, PCR amplification is performed to acquire enough library for sequencing. Amplified cDNAs are size selected to ~1 kb, and sequencing libraries are constructed using the nanopore ligation kit.

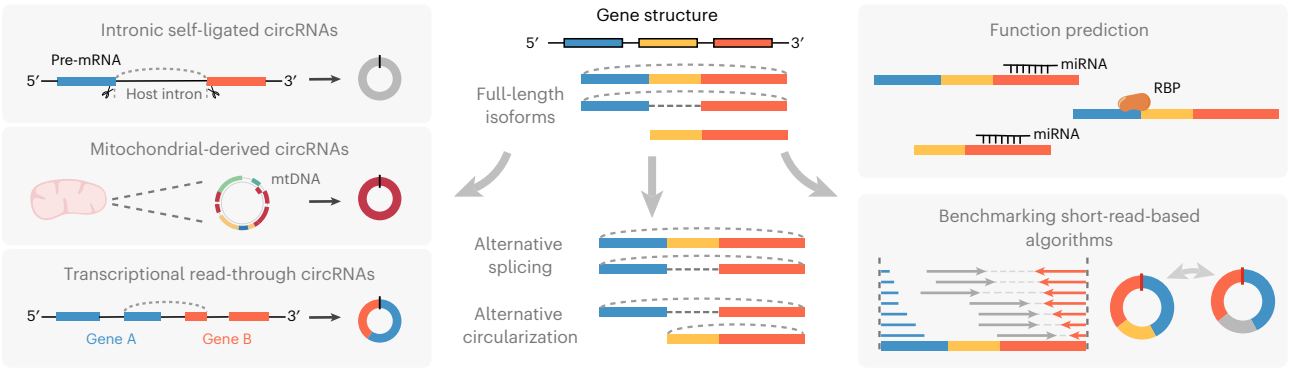


Fig. 2 | Applications of the CIRI-long protocol. Left: the CIRI-long protocol can be used to characterize circRNAs with both canonical and noncanonical splice signals, enabling the effective identification of rare circRNAs including mitochondrial-derived, transcriptional read-through and intronic self-ligated circRNAs. Middle: CIRI-long provides full-length circRNAs isoform sequences, which is useful for alternative splicing and alternative circularization analysis. Right: the assembled full-length circRNAs can support downstream analyses such as function prediction and benchmarking short-read-based circRNA sequencing strategies.

constructed full-length sequences can also serve as a benchmarking dataset for evaluating the performance of short-read-based circRNA assembly algorithms[23]. Considering that most studies still use the canonical Illumina-based circRNA profiling methods[33–35,37], the full-length circRNA sequences can provide important insight into improving these short-read-based computational tools.

Comparison with other methods

CIRI-long combines rolling circular RT and nanopore sequencing to characterize the full-length sequences of circular templates. Similarly, various approaches[38–40] have also been developed to detect full-length circRNAs. Through enzymatic fragmentation and polyadenylation, circNick-LRS[40] generates linear 'nicked' circRNA fragments that are compatible with the nanopore cDNA library preparation protocol. Although circNick-LRS can effectively characterize circRNAs using long-read sequencing, the fragmentation step could break circRNAs molecules into multiple fragments, which might prevent the sequencing of full-length sequences and even cause the identification of falsely truncated isoforms.

In contrast, isoCirc[38] and circFL-seq[39] can both detect full-length circRNA copies using rolling circle amplification and rolling circular RT, respectively. Specifically, isoCirc uses Mung Bean nuclease to digest 3′ cDNA overhangs, and SplintR ligase to create circular DNA templates. Afterward, rolling circle amplification is performed for effective amplification of the circular templates. The isoCirc method requires more complicated experimental steps, including circular ligation and rolling circle amplification for detecting circRNAs. Moreover, the overhang removal and circular ligation also rely on rolling circle RT products, and the circular ligation step could induce false-positive circRNAs by creating artificially ligated concatemers. Meanwhile, like CIRI-long, circFL-seq uses poly(A) tailing, RNase R treatment and rolling circular RT. The main difference between circFL-seq and CIRI-long is that circFL-seq uses poly(A) tailing treatment to add additional poly(A) tails to the first strand, while CIRI-long uses a simpler template switching (TS) mechanism to enable second-strand synthesis.

Experimental design

Starting materials

The CIRI-long protocol can start with total RNA extracted from most cell lines, tissue samples or in vitro-transcribed RNAs. Here we described the protocol for constructing CIRI-long libraries using mouse brain samples and human HeLa cells as examples. When applying the protocol to other sample types, we recommend adjusting the amount of starting material based on the circRNA expression levels in your samples. We found that starting with 1 μg of mouse brain total RNA produces just enough library for nanopore sequencing, thus the amount of starting material should be increased for other tissues with lower circRNA abundance.

Enrichment of circRNAs

Considering that circRNAs are a small percentage of total RNA, CIRI-long implements a modified RNase R treatment step to effectively digest linear RNAs. Although RNase R is widely used to digest

structured linear RNAs in circRNA-related studies[41], studies have demonstrated that a single-stranded overhang of at least seven nucleotides is required for effective digestion of duplex RNAs[42], where highly structured RNAs with short 3′ overhangs could be resistant to RNase R digestion, and thus affect the enrichment efficiency of circRNAs[35]. Thus, additional poly(A)-tailing treatment with *Escherichia coli* Poly(A) Polymerase is added before RNase R digestion, which could increase the length of the 3′ overhang and promote the degradation of RNase R-resistant RNAs[43].

Full-length capture of circRNAs

To capture the full-length circRNA sequences, CIRI-long uses the SMARTscribe Reverse Transcriptase with high RT ability to generate long cDNAs containing multiple copies of the full-length circRNA sequence. The SMART technology is widely used in RNA sequencing due to its ability to generate full-length cDNA products and amplify nanogram amounts of RNA. Here, the TS ability of SMARTscribe Reverse Transcriptase is required for adding a TS oligo to the 3′ end of first-strand cDNA and enabling second-strand synthesis. Interestingly, the TS activity of reverse transcriptase relies on the 5′ end of mRNA, which does not exist in the circular template. However, the TS products have been detected in the constructed library, which could be explained by RNA hydrolysis during RT (Fig. 3a). To validate this theory, we tested the TS mechanism using an in vitro transcription and RT experiment (Fig. 3b). As shown in Fig. 3c, denaturing urea polyacrylamide gel electrophoresis

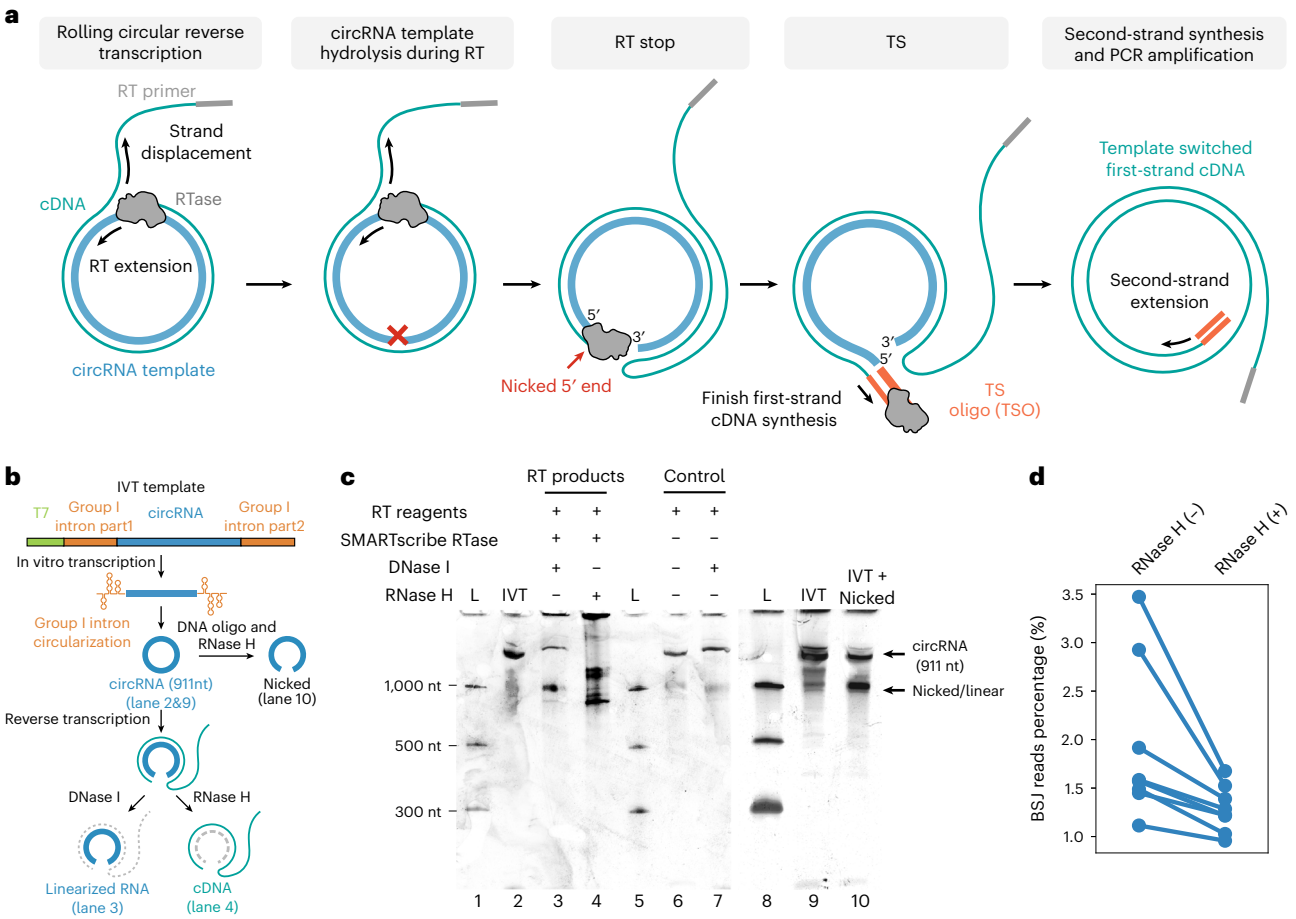


Fig. 3 | RNA hydrolysis and TS during circular RT. a, The hypothesized mechanism of TS during circular RT. CircRNA templates are hydrolyzed during RT, providing a 5′ end necessary for TS. **b**, The schematic diagram of in vitro transcription and RT experiment. We employed the group I intron autocatalysis strategy to produce a 911 nt circRNA using in vitro transcription (IVT), and RT was performed as described in the protocol. Then, DNase I or RNase H were added to the RT products to digest first-strand cDNA or RNA, respectively. **c**, Denaturing urea PAGE analysis of the digested products. Here, a large proportion of linearized circRNAs can be observed after RT, and the nicked/linear RNA circRNA band is validated using both DNase I (lane 3) and RNase H digestion (lane 4). Meanwhile, no linearized RNA band is observed in control samples without SMARTscribe Reverse Transcriptase (lanes 6 and 7), which indicates that the circRNAs are specifically linearized during RT rather than hydrolyzed by RT buffer. Lane 10: the circRNA was nicked using DNA oligo and RNase H digestion as a control. L: ladder lanes. **d**, Effect of adding additional RNase H enzyme to cleave RNA templates during Step 44 on circRNA detection by CIRI-long. Each line represents a pair of RNase H (−) and RNase H (+) samples. Data are obtained from the original CIRI-long article[22].

(PAGE) analysis of RT products from the in vitro-transcribed circRNA suggested that circRNA templates are mostly hydrolyzed during RT, providing the necessary 5′ end for TS activity. Moreover, we have demonstrated in the original CIRI-long publication[22] that additional RNase H treatment to nick circRNAs after RT does not improve circRNA detection efficiency (Fig. 3d), which is consistent with the high percentage of linearized circRNAs in the in vitro experiment. Thus, we believe that RNA template hydrolysis can promote TS, which enables the downstream amplification of cDNAs containing multiple copies of the full-length circRNA sequence.

Size selection and sequencing library construction

To obtain sufficient cDNA products for sequencing, polymerase chain reaction (PCR) amplification is performed on the reverse-transcribed cDNA. Furthermore, the amplified cDNA is selected to the length of >1 kb using AMPure XP beads to enrich the long cDNA molecules generated from multiple rounds of RT on circRNAs. Afterward, the size-selected cDNA from multiple samples could be multiplexed using nanopore barcoding kits, and finally pooled together for nanopore sequencing. Empirically, a canonical CIRI-long library requires ~2 GB data for effective detection of circRNAs. Thus, we recommend pooling no more than six samples in a single MinION flowcell (8–15 GB output per flowcell depending on the library quality and flowcell status). For sequencing with the PromethION flowcell, the number could be increased to 24 accordingly.

Data analysis

The CIRI-long algorithm implements a kmer-based strategy to detect circRNAs from long-read sequencing products. As shown in Fig. 4, the long reads are first trimmed and split into *k*-mers for scanning repetitive sequence patterns. Then, the cyclic consensus sequence (CCS) for each read is constructed using the partial order alignment algorithm. Each CCS sequence represents a full-length circRNA, and is subsequently aligned to the reference genome to determine the BSJ sites along with the circular exon structure. Then, results from multiple samples are aggregated together. The circRNA candidates are clustered based on genomic coordinates, and the splice junction sequences are further corrected using dynamic programming algorithm that combines raw reads from multiple samples and keeps the final junction sites that are supported by the largest number of reads. Finally, the expression matrix of circRNAs in each sample is calculated. A detailed version of CIRI-long installation and usage documentation can be found at https://ciri-cookbook.readthedocs.io/en/latest/CIRI-long_0_home.html.

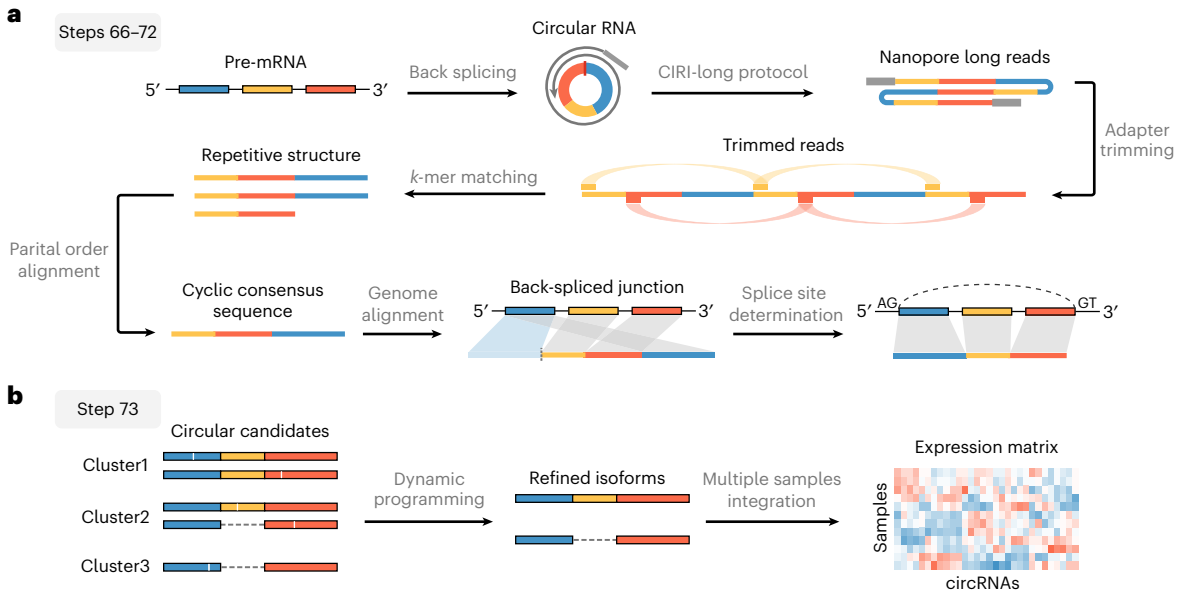


Fig. 4 | Overview of the CIRI-long data analysis workflow. a, The identification of circRNAs from long nanopore reads. The CIRI-long algorithm includes *k*-mer based repetitive identification, consensus sequencing calculation and BSJ determination. **b**, Aggregation of circRNA identification results from multiple samples. All candidate BSJ reads are clustered based on their genomic coordinates, and the BSJ sequences for each cluster are realigned and corrected using dynamic programming algorithms. Finally, the expression matrix of corrected circRNAs in each sample is calculated.

Expertise needed to implement the protocol

The CIRI-long protocol only requires basic molecular biology experimental techniques, and special attention (bleaching the workbench with RNaseZap and changing gloves frequently) is needed to avoid RNA degradation during the experimental procedures. Additionally, users should have access to a high-performance computing cluster or cloud services for data analysis. Users should also have basic knowledge of how to install and run software under the Linux operating system.

Limitations

The CIRI-long protocol uses RNase R treatment to enrich circRNAs which lack 3′ tails (and are thus resistant to the treatment) by degrading linear RNAs, and the constructed library can only be used for circRNA analysis. The most crucial step in the CIRI-long library construction is TS during circular RT. Although experimental results have suggested the hydrolysis of circRNA template during RT can provide the 5′ end to activate the terminal transferase activity of reverse transcriptase, the efficiency and underlying mechanism of this process still needs further assessment. In the CIRI-long protocol, the fragment size selection conditions have been optimized for circRNA detection efficiency, but this selection may still result in a tendency for capturing longer circRNAs. Finally, like most circRNA characterization methods, CIRI-long does not use a unique molecular identifier during RT, meaning that the level of PCR duplicates in the experimental procedure might induce bias in quantification results.

Materials

Biological materials

- Whole brain sample of adult mice (C57BL/6, female, 17 weeks) ▲ **CRITICAL** Our procedure is suitable for most tissue samples and cell lines. However, users may need to adjust the amount of starting tissue based on the average yield of total RNA and the relative mass of circRNAs in each tissue. All procedures were approved by the Animal Ethics Committee at the Institute of Zoology, Chinese Academy of Sciences. **! CAUTION** Animal experiments should be performed according to the relevant institutional and national guidelines and regulations.
- HeLa cell line (American Type Culture Collection, cat. no. CCL-2; RRID: CVCL_0030) **! CAUTION** The cell lines should be regularly checked to ensure their authenticity and avoid contamination.

Reagents

- Phosphate-buffered saline (PBS) buffer (HyClone, cat. no. SH30256.01)
- RNaseZap (Thermo Fisher Scientific, cat. no. AM9782)
- AMPure XP (Beckman, cat. no. A63882)
- RNAClean XP (Beckman, cat. no. A63987)
- Nuclease-free water (UltraPure DNase/RNase-free water, Thermo Fisher Scientific, cat. no. 10977015)
- Ethanol, pure 200 proof, for molecular biology (Sigma-Aldrich, cat. no. E7023-500mL)
- Qubit double-stranded DNA (dsDNA) HS assay kit (Thermo Fisher Scientific, cat. no. Q32854)
- Qubit RNA HS assay kit (Thermo Fisher Scientific, cat. no. Q32852)
- HS NGS Fragment Kit (1–6,000 bp) (Agilent, cat. no. DNF-474-0500)
- HS RNA Kit (15NT) (Agilent, cat. no. DNF-472-0500)
- TRIzol reagent (Thermo Fisher Scientific, cat. no. 15596018) **! CAUTION** TRIzol is a highly corrosive and toxic chemical, and exposure can lead to severe chemical burns and the risk of permanent scarring. Ensure skin and eye protection are used when handling this reagent.
- Liquid nitrogen **! CAUTION** Never allow your unprotected body parts to be in direct contact with liquid nitrogen or objects cooled by liquid nitrogen due to its extremely low temperature. Contact of liquid nitrogen with skin or eyes can cause severe frostbite. To protect your hands when grinding, always wear face/eye protection and use anti-freeze gloves.
- 2-Propanol or isopropanol, for molecular biology (Sigma-Aldrich, cat. no. I9516-500 ML)
- Trichloromethane or chloroform, for molecular biology (Sigma-Aldrich, cat. no. I9516-500 ML) **! CAUTION** Chloroform can be toxic to humans if inhaled or swallowed. Exposure to chloroform could cause cancer. Protect your skin and eyes when handling this reagent and use in a chemical fume hood.
- RiboErase kit (human/mouse/rat) (KAPA Biosystems, cat. nos. KK8481 and KK8482)
- *E. coli* Poly(A) Polymerase (New England Biolabs, cat. no. M0276S)
- Ribonuclease R (RNase R) (Lucigen, cat. no. RNR07250)

- SMARTer PCR cDNA Synthesis Kit (Clonetech, cat. nos. 634925 and 634926) **! CAUTION** The SMARTscribe Reverse Transcriptase has better performance than Maxima H Minus reverse transcriptase. Theoretically, other reverse transcriptases with TS activity should work, but their performance needs further assessment.
- RNasin Plus RNase Inhibitor (40 U/μl) (Promega, cat. no. N2611)
- RNase H (New England Biolabs, cat. no. M0297L)
- Deoxynucleotide (dNTP) Solution Mix (10 mM) (New England Biolabs, cat. no. N0447L)
- LongAmp Taq 2× Master Mix (New England Biolabs, cat. no. M0287L)
- SMARTer CDS random primer: 5′-AAGCAGTGGTATCAACGCAGAGTACNNNNNN-3′ (Order 100 nmol of DNA oligo, PAGE purification; Dissolve the primer powder in nuclease-free water to a final concentration of 50 uM. The dissolved primer can be stored at −20 °C for months.)

Equipment

- 25 ml reagent reservoir for eight-channel pipettors (VWR, cat. no. BITXSR-0025-BWM)
- Mortar and pestle
- 1.5 ml DNA LoBind tubes (Eppendorf, cat. nos. 0030108051)
- 1.5 ml tubes (Axygen, cat. no. MCT-150-C)
- 0.2 ml PCR tubes (Axygen, cat. no. PCR-02-C)
- 0.2 ml eight-strip PCR Tubes and Domed Cap Strips (Axygen, cat. no. PCR-0208-CP-C)
- 15 ml clear polypropylene centrifuge tubes (Corning, cat. no. 430790)
- 50 ml clear polypropylene centrifuge tubes (Corning, cat. no. 430828)
- Qubit assay tubes (Thermo Fisher Scientific, cat. no. Q32856)
- 10-, 100-, 200- and 1,000 μl filter barrier tips pipette tip, filter, sterile (Axygen, mod. nos. TF-300-R-S, TF-100-R-S, TF-200-R-S, TF-1000-L-R-S)
- 5200 Fragment Analyzer systems (Agilent, mod. no. M5310AA), Bioanalyzer 2100 (Agilent, mod. no. G2939BA), Tapestation (Agilent, mod. no. G2992AA), or other equivalent systems
- Qubit 3.0 fluorometer (Thermo Fisher Scientific, mod. no. Q33216) or Qubit 4.0 fluorometer, Thermo Fisher Scientific, mod. no. Q33238)
- Nanodrop 2000 (Thermo Fisher Scientific, mod. no. ND2000)
- Eppendorf ThermoMixer C with Thermo top (Eppendorf, mod. no. 5382000023)
- Tube Rotator (MACSmix, mod. no. 130-090-753)
- DynaMag-2 magnet for 1.5 ml microtube (Thermo Fisher Scientific, mod. no. 12321D)
- DynaMag-96 magnets (Thermo Fisher Scientific, mod no. 12331D)
- Refrigerated centrifuge (e.g., Eppendorf, mod. no. 5430R)
- Bench top centrifuge (e.g., Eppendorf, mod. no. minispin)
- Thermocycler (e.g., Biometra mod. no. TRIO 48)
- Minicentrifuge (e.g., Kelyn-Bell, mod. no. LX-200)
- Nanopore sequencer (Oxford Nanopore Technologies, MinION/GridION/PromethION)

Software

- MinKNOW (version v4.1.2 or later): https://community.nanoporetech.com/downloads
- Guppy (version v3.4.5 or later): https://community.nanoporetech.com/downloads **▲ CRITICAL** CIRI-long has been tested on reads base called with Guppy v3.4.5, and more recent versions are also likely to be applicable. The latest version of Guppy is always recommended considering the base calling accuracy is continuously improving in the recent Guppy releases.
- LongQC[44] (version 1.2.0c): https://github.com/yfukasawa/LongQC
- Porechop (modified version): https://github.com/Kevinzjy/Porechop
- bwa[45] (version 0.7.17): https://github.com/lh3/bwa
- CIRI-long[22] (version 1.1.0 or later): https://github.com/bioinfo-biols/CIRI-long

Equipment setup

Example datasets

For mouse brain samples, download the example CIRI-long sequencing dataset generated in our original CIRI-long[22] publication from the GSA database (https://download.cncb.ac.cn/gsa/CRA003317/CRR194209/CRR194209.fq.gz)

Box 1 | Software setup for CIRI-long analysis ● Timing 1 h

Procedure

CIRI-long installation

▲ **CRITICAL** The CIRI-long software can be run only in a UNIX/Linux environment. Users should have access to a high-performance computing cluster or Linux workstation. CIRI-long requires python version 3.7 or higher, and installation using the python package 'virtualenv' is highly recommended:

Step 1. Install pip and virtualenv

```
python3 -m pip install --user --upgrade pip

python3 -m pip install --user virtualenv
```

Step 2. Create and activate the CIRI-long virtual environment

```
python3 -m venv CIRI-long

source CIRI-long/bin/activate
```

Step 3. Install CIRI-long and verify installation was successful

```
pip install CIRI-long

which CIRI-long
```

Following the above steps, the 'CIRI-long' software should be successfully installed into the virtual environment. The tutorial for installing and running CIRI-long is also available at https://ciri-cookbook.readthedocs.io/en/latest/CIRI-long_0_home.html.

Other programs for CIRI-long data analysis

Other programs need to be installed according to the manufacturer's instructions.

Download and install MinKNOW from https://community.nanoporetech.com/downloads, tutorial: https://community.nanoporetech.com/docs/prepare/library_prep_protocols/experiment-companion-minknow/v/mke_1013_v1_revcl_11apr2016.

Download and install Guppy (v3.4.5 or later) from https://community.nanoporetech.com/downloads, tutorial: https://community.nanoporetech.com/docs/prepare/library_prep_protocols/Guppy-protocol/v/gpb_2003_v1_revaj_14dec2018/linux-guppy.

Download and install LongQC (v1.2.0c) from https://github.com/yfukasawa/LongQC. Download and install Porechop (modified version) from https://github.com/Kevinzjy/Porechop. Download and install bwa (v0.7.17) from https://github.com/lh3/bwa.

Software installation

The CIRI-long pipeline is designed to run in the UNIX/Linux environment. To install and configure the necessary software, follow the instructions in Box 1.

Procedure

Total RNA extraction ● Timing 2 h

▲ **CRITICAL** All reagents must be kept nuclease free. All experimental procedures including reagent preparation should be performed under RNase- and DNase-free conditions. Bleach the workbench with RNaseZap before starting the experiment. Change gloves frequently to avoid RNase contamination. RNA samples and enzymes should be kept on ice when not used or stored. We recommend RNA extraction using TRIzol, but other RNA extraction methods can also be used on appropriate sample types.

1 Grind a whole brain of an adult mouse (~450 mg) in liquid nitrogen with mortar and pestle immediately after the tissue was dissected, then separate the ground sample into six parts. Add each part into a 1.5 ml tube containing 1ml TRIzol (Invitrogen). For adherent cells, add 1ml of TRIzol to a 10 cm dish to digest the cells after washing the cells once with PBS, and transfer to a 1.5 ml tube. Extract total RNA according to the manufacturer's instructions.

2 Elute RNA in 50–100 μl of nuclease-free water.

3 Quantify the extracted RNA using a Qubit fluorometer, and assess the purity of RNA using a Nanodrop 2000. The expected yield of RNA from one adult mouse brain should be 200–400 μg. If using a smaller amount of tissue, reduce the elution volume to obtain RNA with a concentration above 100 ng/μl.

4 Check the quality of RNA with the 5200 Fragment Analyzer system (Agilent). RNA with an RNA Quality Number ≥7.0 was used for the subsequent procedure.

■ **PAUSE POINT** The total RNA can be stored at −80 °C for months.

rRNA depletion ● Timing 2.5–3 h

5 Equilibrate the RNAClean XP beads for at least 30 min at room temperature (25 °C), and vortex thoroughly to resuspend the beads.

6 Use 1 μg total RNA as starting material.
▲ **CRITICAL STEP** The amount of starting RNA will affect library yield. For cell lines or tissues with lower circRNA expression levels, increase the amount of starting RNA accordingly.
7 Remove rRNA using an RNase H-based commercial kit (RiboErase kit (human/mouse/rat), KAPA Biosystems) according to the manufacturer's user guide, and elute RNA in 18 μl nuclease-free water.
8 Aspirate 2 μl RNA for quantification with a Qubit fluorometer. The expected RNA yield is in the range of 30–60 ng.
9 Aspirate another 1 μl RNA for detecting whether mRNA is degraded or not with the 5200 Fragment Analyzer systems (Agilent).
▲ **CRITICAL STEP** The expected RNA fragment size distribution should have a peak between 1 and 2 kb.
■ **PAUSE POINT** The purified rRNA-removed RNA can be stored at −80 °C for months.

? TROUBLESHOOTING

Poly(A) Tailing ● Timing 1 h

10 Prepare the following reagents in a 0.2 ml RNase/DNase-free PCR tube:

Reagent	Volume (μl)	Final concentration
rRNA-depleted RNA	15	—
10× *E. coli* Poly(A) Polymerase Reaction Buffer	2	1×
Adenosine triphosphate (10 mM)	2	1 mM
E. coli Poly(A) Polymerase (5 U/μl)	1	0.25 U/μl
Total volume	20	—

11 Incubate at 37 °C for 30 min (hot lid temperature 65 °C). Equilibrate the RNAClean XP beads during this step for at least 30 min at room temperature and vortex thoroughly to resuspend the beads before use. Stop the reaction by proceeding to the clean-up step.
12 Add 44 μl (2.2× volumes) of RNAClean XP (Beckman) beads to the 20 μl polyadenylated RNA sample and mix the sample well by pipetting up and down eight times and spin down briefly in a minicentrifuge (Kelyn-Bell).
13 Incubate at room temperature for 5 min to bind RNA to the beads.
14 Place the 0.2 ml tube on the magnetic stand until the solution is clear (~2 min).
15 Keep the tube on the magnetic stand. Remove 60 μl supernatant carefully, taking care not to disturb the beads.
16 Add 200 μl of freshly prepared 75% (vol/vol) ethanol to the tube. Wait for 30 s, then discard all the supernatant.
▲ **CRITICAL STEP** Beads should always be kept on the magnetic stand while washing with ethanol, and should not be resuspended.
17 Wash beads once more with 75% (vol/vol) ethanol by repeating Step 16.
18 Quickly spin the tubes in a minicentrifuge and remove all residual liquid.
19 Air-dry the beads on the magnetic rack for 30 s.
▲ **CRITICAL STEP** The beads should not be over-dried during the process, because cracked beads will decrease the RNA recovery.
20 Resuspend beads with 20 μl nuclease-free water.
21 Incubate at room temperature for 2 min to elute RNA from the beads.
22 Put the tube back on the magnetic rack for 1 min to pellet the beads.
23 Transfer 18.5 μl supernatant to a new 0.2 ml tube. Take care not to disturb the bead pellet.
24 Take 1 μl of poly(A)-tailed RNA and measure the concentration with a Qubit fluorometer. The expected yield for each sample should be in the range of 1–3 ng/μl.
■ **PAUSE POINT** The purified polyadenylated RNA can be stored at −80 °C for months.

RNase R treatment ● Timing 0.8 h

25 Equilibrate the RNAClean XP beads for at least 30 min at room temperature and vortex thoroughly to resuspend the beads.

26 Add the following reaction components in a 0.2ml RNase/DNase-free PCR tube in the order specified below:

Reagent	Volume (µl)	Final concentration
Polyadenylated RNA	17.5	—
RNase R Buffer	2	—
RNase R (20 U/µl)	0.5	0.5 U/µl
Total volume	20	—

27 Incubate at 37 °C for 15 min (hot lid temperature 65 °C). Stop the reaction by proceeding to the clean-up step.
28 Vortex thoroughly to resuspend the RNAClean XP beads before use.
29 Add 44 µl (2.2× volumes) of RNAClean XP (Beckman) beads to the 20 µl RNase R-treated RNA sample from Step 27 and mix the sample well by pipetting up and down eight times and spinning down briefly.
30 Incubate at room temperature for 5 min to bind RNA to the beads.
31 Place the 0.2 ml tube on the magnetic stand until the solution is clear (~2 min). Keep the tube on the magnetic stand. Remove 60 µl supernatant carefully, taking care not to disturb the beads.
32 Add 200 µl of freshly prepared 75% (vol/vol) ethanol to the tube. Wait for 30 s, then discard the entire supernatant.
▲ **CRITICAL STEP** Beads should always be kept on the magnetic stand while washing with ethanol, and should not be resuspended.
33 Wash beads once more with 75% (vol/vol) ethanol by repeating Step 32.
34 Quickly spin the tubes in a minicentrifuge and remove all residual liquid.
35 Air dry the beads on the magnetic rack for 30 s.
▲ **CRITICAL STEP** The beads should not be over-dried during the process, because cracked beads will decrease the RNA recovery.
36 Resuspend beads with 5 µl nuclease-free water.
37 Incubate at room temperature for 2 min to elute RNA from the beads.
38 Put the tube back on the magnetic rack for 1 min to pellet the beads.
39 Transfer 3.5 µl supernatant to a new 0.2 ml tube. Take care not to disturb the bead pellet.
▲ **CRITICAL STEP** The concentration of RNase R-treated RNA is too low to be detected, so the quality control step can be omitted.
■ **PAUSE POINT** The purified RNase R-treated RNA can be stored at −80 °C for months.

RT ● Timing 2 h

▲ **CRITICAL** The initial Steps 43–44 for first-strand synthesis are critical and should not be delayed when the temperature reaches 42 °C in Step 41. The master mix in Step 42 needs to be prepared while your tubes are incubating at Step 41 to quickly start the cDNA synthesis.

40 Prepare the following mix containing the components listed below in a 0.2 ml RNase/DNase-free PCR tube. Mix by pipetting, and spin the tubes briefly in a microcentrifuge.

Reagent	Volume (µl)	Final concentration
RNase R treated RNA	3.5	—
SMARTer CDS random primer (50 µM)	1	11 µM
Total volume	4.5	—

41 Incubate the tubes at 72 °C in a hot-lid thermal cycler for 3 min, then reduce the temperature to 25 °C for 10 min, then hold at 42 °C for no more than 5 min. (hot lid temperature 85 °C). Prepare the RT master mix in Step 42 when the temperature reaches 25 °C.
42 Prepare the following mix containing the components listed below in a 0.2 ml RNase/DNase-free PCR tube at the same time. Mix well by pipetting and spin the tube briefly in a microcentrifuge.

Reagent	Volume (μl)	Final concentration
5× First Strand Buffer (RNase free)	2	1.8×
Dithiothreitol (100 mM)	0.25	2.5 mM
dNTP mix (10 mM)	1	1 mM
SMARTer II A Oligonucleotide (12 μM)	1	1.2 μM
RNase Inhibitor (40 U/μl)	0.25	1 U/μl
SMARTScribe Reverse Transcriptase (100 U/μl)	1	10 U/μl
Total volume	5.5	—

▲ **CRITICAL STEP** Add the reverse transcriptase to the mix just before use. If you are processing multiple samples at the same time, prepare a master mix to avoid inaccurate pipetting caused by adding too small volumes of reagents.

43 Transfer 5.5 μl of the Master Mix into the tube containing RNA from Step 41 above. Mix the contents of the tube by gently pipetting, and spin the tube briefly to collect the contents at the bottom.

44 Put the tube back into the thermal cycler, and incubate at 42 °C for 90 min to obtain the full-length cDNAs (hot lid temperature 85 °C).

45 Terminate the reaction by heating the tube at 70 °C for 10 min (hot lid temperature 85 °C).

46 Dilute the 10 μl first-stranded cDNA product fivefold by adding 40 μl TE buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM ethylenediaminetetraacetic acid), to reduce the influence of salt ions in the RT reaction on the subsequent PCR reaction.

■ **PAUSE POINT** cDNA can be stored at 4 °C overnight or −20 °C for months.

cDNA PCR amplification ● Timing 1.5 h

▲ **CRITICAL** To acquire enough library for nanopore sequencing after the final size selection, cDNA was amplified in eight tubes.

47 Equilibrate the Agencourt AMPure XP beads for Step 52 for at least 30 min at room temperature and vortex thoroughly to resuspend the beads.

48 Thaw the components below on ice and mix well by inverting several times before use. To prepare the master mix, add the following components to a 1.5 ml tube on ice:

Reagent	Volume (μl)	Volume (μl) ×8.8	Final concentration
LongAmp Taq 2× Master Mix	25	220	1×
5′ PCR Primer II A (12 μM)	2	17.6	0.48 μM
Diluted cDNA[a]	5	44	—
Nuclease-free water	18	158.4	—
Total volume	50	440	—

[a]The volume of the cDNA products should not exceed 10% of the total reaction volume.

49 Mix well by pipetting up and down eight times. Quickly spin to collect all liquid to the bottom of the tube if necessary. Aliquot 50 μl master mix to the wells of an eight-strip PCR tube.

▲ **CRITICAL STEP** Process eight samples simultaneously at Steps 50–58, and combine the eight tubes into one tube at Step 59.

50 Transfer the PCR tubes to a Thermocycler and begin the program below:

Step	Number of cycles	Denature	Anneal	Extend
1	1	95 °C, 30 s	—	—
2	19–21[a]	95 °C, 15 s	62 °C, 15 s	65 °C, 2 min
3	1	—	—	65 °C, 5 min

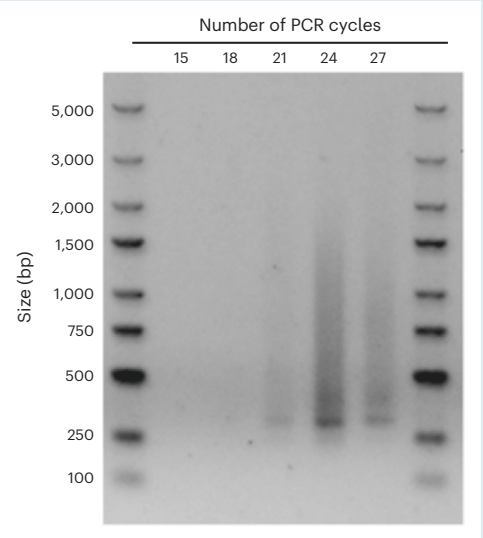
[a]Use an optimal number of cycles. In our CIRI-long experiments, we used 19 cycles for library generation.

Box 2 | Optimization of PCR cycle number ● Timing 2 h

For the best results, you should optimize the PCR cycling parameters for your experiment when applying the protocol to different species or tissues. Over-amplification of cDNA can lead to more duplicates and less complexity of the library. On the other hand, insufficient amplification will result in insufficient amount of cDNA for sequencing. The optimization for PCR cycling parameters is as follows:

1 Prepare 100 µl PCR reaction mix as outlined in Step 48, amplify for 15 cycles first, then pause the program. Transfer 30 µl mix from the tube to a new PCR tube labeled 'Optimization'. Store the remaining 70 µl reaction tube labeled 'Experimental' at 4 °C.
2 Transfer 5 µl from the 'Optimization' tube above to a clean tube. This is the '15 cycle' tube.
3 Return the 'Optimization' tube (remaining 25 µl mix) to the thermal cycler, and run additional three cycles. Transfer 5 µl from the 'Optimization' tube above to a clean tube.
4 Repeat the above step three more times.
5 Detect the amount of amplified products for 15 cycles, 18 cycles, 21 cycles, 24 cycles and 27 cycles by running 5 µl of each reaction on an agarose gel or 1 µl amplified products on a Bioanalyzer.
6 A peak distribution smaller than 1 kb observed on the gel indicates that the cDNA is over-amplified. No smear appearing on the gel indicates that the cDNA is under-amplified. The number of cycles between under-amplified and over-amplified can be chosen as the optimized number of PCR cycles to amplify the remaining 70 µl reaction tube labeled 'Experimental'.

The image shows gel electrophoresis of PCR products amplified for 15, 18, 21, 24 and 27 cycles. No smear is found when fewer than 18 cycles are used, while 21 cycles can generate a clear smear. Thus, the optimal number of PCR cycles is determined to be 19–21 cycles.



▲ **CRITICAL STEP** We typically amplify cDNA libraries for 19 cycles. Users may increase the cycle number to ≥21 to increase the yield if necessary. The PCR cycle number should be optimized (for further instructions, see Box 2).

? TROUBLESHOOTING

51 Hold the PCR products at 4–10 °C

■ **PAUSE POINT** Amplified cDNA can be stored at 4 °C overnight or −20 °C for at least a few days.

Fragment size selection and quality control ● Timing 0.5 h

52 Add 25 µl (0.5× volumes) of AMPure XP beads to the 50 µl PCR products and mix the sample well by pipetting up and down eight times and spinning down briefly.
53 Incubate at room temperature for 5 min to bind amplified cDNA to the beads.
54 Place the 0.2 ml tube on the magnetic stand until the solution is clear (~2 min). Keep the tube on the magnetic stand. Remove 60 µl supernatant carefully, taking care not to disturb the beads
55 Add 200 µl of freshly prepared 75% (vol/vol) ethanol to the tube. Wait for 30 s, then discard all the supernatant.

▲ **CRITICAL STEP** Beads should always be kept on the magnetic stand while washing with ethanol, and should not be resuspended.

56 Wash beads once more with 75% (vol/vol) ethanol by repeating Step 55.
57 Quickly spin the tubes in a minicentrifuge and remove all residual liquid.

58 Air dry the beads on the magnetic rack for 30 s.
▲ CRITICAL STEP The beads should not be over-dried during the process, because cracked beads will decrease the RNA recovery.
59 Resuspend beads with 8 µl nuclease-free water. Pool eight tubes of 8 µl samples to get a total volume of 64 µl. Incubate at room temperature for 2 min to elute the cDNA library from the beads.
60 Put the tube back on the magnetic rack for 1 min to pellet the beads.
61 Transfer 62 µl supernatant to a new 1.5 ml tube. Take care not to disturb the bead pellet.
62 Aspirate 1 µl DNA for quantification with a Qubit fluorometer. The expected DNA yield is at least 50 ng.
63 Aspirate another 1 µl DNA for detecting DNA fragment distribution with the 5200 Fragment Analyzer system (Agilent). The expected DNA size distribution should have a peak between 1 and 2 kb.
■ PAUSE POINT The purified second-strand cDNAs can be stored at −20 °C for at least 1 year.

? TROUBLESHOOTING

Nanopore sequencing ● Timing 2–3 h for library preparation and 48 h for sequencing

64 Prepare the cDNA libraries following the Oxford Nanopore protocol 'Ligation sequencing amplicons —native barcoding (SQK-LSK109 with EXP-NBD104 and EXP-NBD114)'. The specific steps include end prep, native barcode ligation and nanopore adapter ligation, and flow cell priming and loading. The updated barcoding kit (SQK-LSK110) can replace SQK-LSK109 for this experiment.
▲ CRITICAL STEP Make sure the DNA meets the quantity and quality requirements specified by the manufacturers. Using too little or too much DNA will affect your library preparation. 100–200 fmol input DNA is required for each sample when using R9.4.1 flow cells. The online tool NEBioCalculator (https://nebiocalculator.neb.com/#!/dsdnaamt) is recommended to convert dsDNA mass to moles. Using different barcodes for different samples is critical for demultiplexing pooled sequencing reads. A 0.5:1 AMPure beads-to-sample ratio is recommended to select against short fragments after sequencing adapter ligation. The wash buffer SFB, not LFB, which is designed to enrich DNA fragments >3 kb, must be used to purify all fragments equally. Load 30–50 fmol DNA library for R9.4.1 flow cells.
65 Sequence the constructed cDNA library using MinKNOW on a MinION Mk1B device. The protocol is also applicable to other nanopore sequencing devices (i.e., MinION Mk1C, GridION or PromethION).
▲ CRITICAL STEP To ensure the accuracy of sequencing reads, the high accuracy or super accuracy mode should be used for base calling in the MinKNOW graphical user interface.

? TROUBLESHOOTING

Data analysis ● Timing 6 h

▲ CRITICAL Ensure your machine has 64 GB of memory. Here an example dataset of mouse brain samples from the original CIRI-long manuscript was used for illustration. Replace the input and output file and directory name with your dataset.
66 Download the mouse GRCm38.p6 reference genome and GENCODE Release M25 annotation:

```
wget https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M25/gencode.vM25.annotation.gtf.gz

wget https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M25/GRCm38.p6.genome.fa.gz
```

▲ CRITICAL STEP For human and mouse samples, the GENCODE project is recommended. For data with other species, replace the reference genome and annotation gtf with your customized reference.
67 Build the bwa index of the reference genome for sequence alignment:

```
gunzip GRCm38.p6.genome.fa.gz

bwa index -a GRCm38.p6.genome.fa
```

68 Obtain the basecalled sequencing reads from MinKNOW software. The fastq file should be demultiplexed using MinKNOW or other demultiplexing tools. Download the example dataset and rename it to reads.fastq.gz

```
wget https://download.cncb.ac.cn/gsa/CRA003317/CRR194209/CRR194209.fq.gz

mv CRR194209.fastq.gz reads.fastq.gz
```

69 Use the LongQC program to evaluate the quality of sequencing reads:

```
python longQC.py sampleqc -x ont-ligation -o longqc_out -p 16 reads.fastq.gz
```

Check the read length distribution of your dataset. The optimal length for detecting circRNAs is 1,000–1,500 bp.

? TROUBLESHOOTING

70 Trim the sequencing adapter and SMARTer primers using Porechop:

```
porechop -i reads.fastq.gz -o reads.trimmed.fastq.gz --threads 16
```

▲ **CRITICAL STEP** Be sure to use the modified version of Porechop with the SMARTer primer sequence for adapter trimming. Otherwise, please add the following lines to line 80 of the original version of Porechop/adapters.py:

```
Adapter('SMARTer primer', start_sequence=('CIRI-long', 'AAGCAGTGGTATCAACGCAGAGTAC'), end_sequence=('CIRI-long_rev', 'GTACTCTGCGTTGATACCACTGCTT')),
```

71 Detect circRNA reads from nanopore sequencing data using the 'CIRI-long call' command:

```
CIRI-long call -i reads.trimmed.fastq.gz -o call_output -r GRCm38.p6.genome.fa -p sample -a gencode.vM25.annotation.gtf -t 16 -c mouse_bed_v2.0.bed
```

▲ **CRITICAL STEP** The mouse_bed_v2.0.bed is downloaded from the circAtlas v2.0 database and must have either a 'gtf' or 'bed' suffix. CIRI-long will use the provided splice sites in the annotation bed/gtf file to adjust BSJ junctions more accurately.

? TROUBLESHOOTING

72 Prepare the sample sheet in the following format, save it to 'input.txt':

```
SMARTer_H-_rep1 call_output/sample/sample.cand_circ.fa
```

▲ **CRITICAL STEP** The sample sheet should have two columns separated by a space. The first column contains the sample name, and the second column contains the path of output circular reads fasta of each sample from Step 71. Each line represents one sample.

73 Use the CIRI-long collapse command to integrate results from multiple samples and perform cluster analysis:

```
CIRI-long collapse -i input.txt -o./collapse_out -p sample -r GRCm38.p6.genome.fa -a gencode.vM25.annotation.gtf -t 16 -c mouse_bed_v2.0.bed
```

The output directory (collapse_out) created in this step will include several text files:

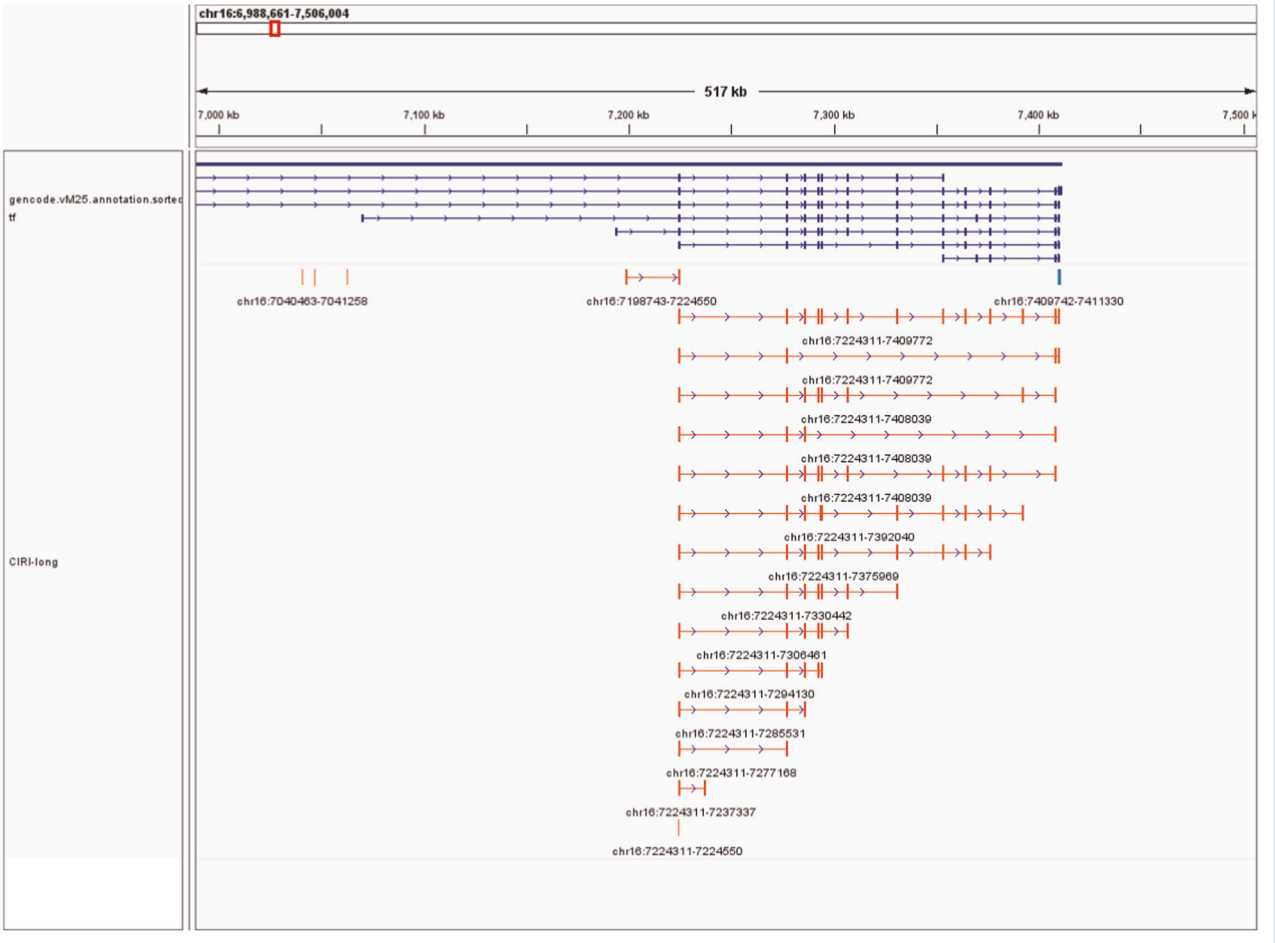
- sample.info: the main output in GTF format. Columns represent the genomic position of assembled circRNA transcripts. The attribute columns consist of additional information including host gene and predicted exons of circRNAs

Box 3 | Visualization of CIRI-long output ● Timing 0.1 h

To visualize the assembled circRNAs from CIRI-long results, convert the sample.info output to bed format:

```
python3 misc/convert_bed.py collapse_out/sample.info sample_circ.bed
```

Then, visualize the bed-formatted output using UCSC Genome Browser, IGV[48] or Jbrowse2 (ref. [49]):
The image shows the 19 full-length circRNA isoforms detected from the Rbfox1 gene locus.



- sample.expression: the expression matrix of circRNAs at BSJ level. Each row represents an individual BSJ, and one BSJ can be derived from multiple circRNA isoforms with different alternative splicing structures
- sample.isoform: the isoform usage matrix of the circRNA isoform. Each row represents an assembled isoform, and values are calculated as the number of isoform reads divided by the total number of reads from the same BSJ site
- sample.reads: predicted exon structure for each CCS read

▲ **CRITICAL STEP** The reference genome, annotation.gtf and circRNA.bed files should be the same as in Step 71.

▲ **CRITICAL STEP** Users can convert the GTF-formatted 'sample.info' to.bed format for downstream visualization using scripts in the CIRI-long package (Box 3).

? TROUBLESHOOTING

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
9	RNA fragments <1 kb	RNA degradation	All experimental procedures including reagent preparation should be performed under RNase- and DNase-free conditions. Change gloves frequently to avoid RNase contamination
50	Low quantity of amplified cDNA products	Insufficient PCR cycle number	Increase the number of PCR cycles to obtain enough library for sequencing
		Poor sample clean-up during library prep	Completely remove the residual ethanol and do not over dry beads during clean-up
63	Presence of peaks smaller than 1 kb after amplification	Over-amplification of reverse transcribed cDNA	Decrease the number of PCR cycles
	Low quantity of size-selected cDNA products	Low percentage of circRNAs in your sample	Increase the amount of starting RNA or increase the beads-to-sample ratio in Step 52 for less stringent size selection
65	High percentage (>15%) of adapter in sequencing run	Insufficient PCR cycle number	Increase the number of PCR cycles to obtain enough library for sequencing or increase the amount of cDNA for ligation
		Poor sample clean-up during library prep	Perform an additional clean-up before Step 64 to remove free adapters
69	Length of sequencing reads mostly too short (<1 kb)	Failed to enrich long cDNA fragments	Perform size selection carefully following the instructions in Steps 52–63
	Length of sequencing reads mostly too long (> 1.5 kb)	Excessive size selection of cDNA library	Increase the beads-to-sample ratio according to your sample
71	Failed to run CIRI-long	Invalid parameters specified when running CIRI-long	Install and run CIRI-long following the protocol and instructions from https://ciri-cookbook.readthedocs.io/en/latest/CIRI-long_2_usage.html
	Low percentage (<1%) of cyclic consensus reads	Failed to enrich long cDNA fragments	Decrease the sample-to-beads ratio for selecting longer cDNA products
73	Low number of detected circRNAs	Insufficient sequencing depth	Decrease the number of multiplexed libraries on a single flowcell. Use a platform with higher throughput (e.g. PromethION)
		Low enrichment of circRNAs in the procedure	Increase the amount or prolong the processing time of RNase R

Timing

Steps 1–4, total RNA extraction: 2 h
Steps 5–9, rRNA depletion: 2.5–3 h
Steps 10–24, poly(A) tailing: 1 h
Steps 25–39, RNase R treatment: 0.8 h
Steps 40–46, RT: 2 h
Steps 47–51, cDNA PCR amplification: 1.5 h
Steps 52–63, fragment size selection: 0.5 h
Steps 64–65, nanopore sequencing: 50 h
Steps 66–73, data analysis: 6 h
Box 1, Software setup, 1h
Box 2, PCR cycle optimization: 2 h
Box 3, Visualization of CIRI-long output: 0.1 h

Anticipated results

For a canonical CIRI-long library, the expected amount of rRNA-depleted RNA in Steps 5–9 is in the range of 30–60 ng, with a length of 1–2 kb. In Steps 10–24, the amount of polyadenylated RNA should be 1–3 ng/μl. After RNase R digestion steps 25–39, the amount of RNase R-treated RNA should be below the detection threshold, so no quality control needs to be done after this step. However, after RT, PCR amplification and size selection (Steps 40–63), the expected DNA yield is ~50 ng and the selected library size should be ~1 kb, as shown in Fig. 5.

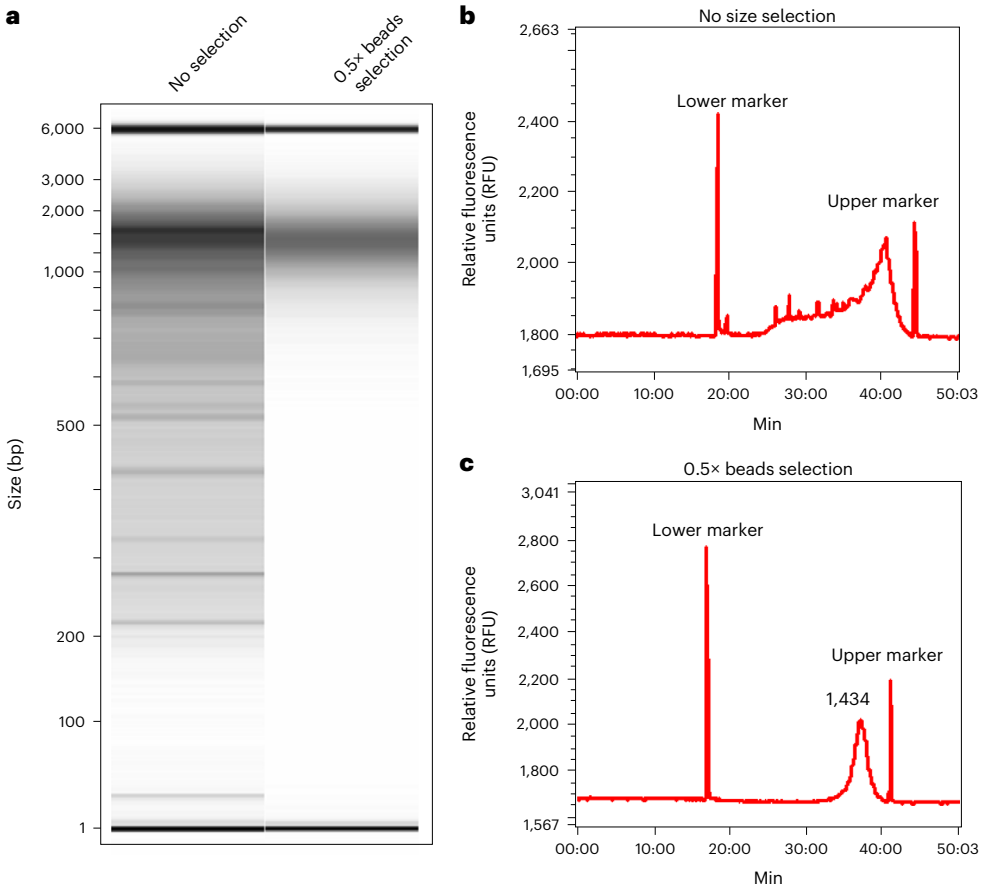


Fig. 5 | Anticipated library size after beads selection. a, Bioanalyzer results of unselected and 0.5× beads selected libraries from the original CIRI-long article[22]. A successful CIRI-long library should be >1 kb in length after size selection. **b**, The size distribution of the unselected library. **c**, The size distribution of library with ideal size selection.

For quality validation of the CIRI-long sequencing data, the length of raw nanopore reads should be in the range of 1,000–1,500 bp. The genomic mapping rate should be >90%, and the percentage of rRNA should be lower than 5%. The number of BSJ reads should be different depending on the percentage of circRNAs in your tissue sample. In this example, ~6% and ~4% reads should be identified as BSJ reads for mouse brain samples and HeLa cells, respectively. For tissues or cell line samples with lower circRNA expression levels, the percentage of BSJ reads should be lower. We recommend sequencing at least 1–3 million reads (1–3 GB bases) for each sample to ensure the effective detection of most highly expressed circRNAs. We expect to get ~10 k circRNAs using 1–3 million nanopore reads for mouse brain samples.

Data availability

The CIRI-long dataset used in this manuscript can be downloaded from the National Genomics Data Center[46] (China National Center for Bioinformation: https://bigd.big.ac.cn/gsa) with accession number CRR194209.

Code availability

The CIRI-long software can be downloaded from GitHub at https://github.com/bioinfo-biols/CIRI-long[47].

References

1. Kristensen, L. S. et al. The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* **20**, 675–691 (2019).
2. Liu, C.-X. & Chen, L.-L. Circular RNAs: characterization, cellular roles, and applications. *Cell* **185**, 2016–2034 (2022).

3. Wu, W., Ji, P. & Zhao, F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.* **21**, 101 (2020).
4. Gao, Y., Wang, J. & Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* **16**, 4 (2015).
5. Zhang, J., Chen, S., Yang, J. & Zhao, F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat. Commun.* **11**, 90 (2020).
6. Wu, W., Zhang, J., Cao, X., Cai, Z. & Zhao, F. Exploring the cellular landscape of circular RNAs using full-length single-cell RNA sequencing. *Nat. Commun.* **13**, 3242 (2022).
7. Ji, P. et al. Expanded expression landscape and prioritization of circular RNAs in mammals. *Cell Rep.* **26**, 3444–3460 e3445 (2019).
8. Dong, R., Ma, X. K., Li, G. W. & Yang, L. CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genom. Proteom. Bioinform.* **16**, 226–233 (2018).
9. Gan, X. et al. CircMUC16 promotes autophagy of epithelial ovarian cancer via interaction with ATG13 and miR-199a. *Mol. Cancer* **19**, 45 (2020).
10. Kristensen, L. S. et al. Spatial expression analyses of the putative oncogene ciRS-7 in cancer reshape the microRNA sponge theory. *Nat. Commun.* **11**, 4551 (2020).
11. Chen, S. et al. circVAMP3 drives CAPRIN1 phase separation and inhibits hepatocellular carcinoma by suppressing c-Myc translation. *Adv. Sci.* **9**, e2103817 (2022).
12. Liu, C. X. et al. RNA circles with minimized immunogenicity as potent PKR inhibitors. *Mol. Cell.* **82**, 420–434 e426 (2022).
13. Chen, S., Zhang, J. & Zhao, F. Screening linear and circular RNA transcripts from stress granules. *Genom. Proteom. Bioinform.* https://doi.org/10.1016/j.gpb.2022.01.003 (2022).
14. Yang, Y. et al. Extensive translation of circular RNAs driven by *N*(6)-methyladenosine. *Cell Res.* **27**, 626–641 (2017).
15. Li, Y. et al. HNRNPL circularizes ARHGAP35 to produce an oncogenic protein. *Adv. Sci.* **8**, 2001701 (2021).
16. Zhang, M. et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat. Commun.* **9**, 4475 (2018).
17. Dal Molin, A. et al. CRAFT: a bioinformatics software for custom prediction of circular RNA functions. *Brief. Bioinform.* https://doi.org/10.1093/bib/bbab601 (2022).
18. Gao, Y., Zhang, J. & Zhao, F. Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* **19**, 803–810 (2018).
19. Gao, Y. et al. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.* **7**, 12060 (2016).
20. Zheng, Y., Ji, P., Chen, S., Hou, L. & Zhao, F. Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Med.* **11**, 2 (2019).
21. Zhang, X. O. et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).
22. Zhang, J. et al. Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat. Biotechnol.* **39**, 836–845 (2021).
23. Zhang, J. et al. Evaluation of circRNA sequence assembly methods using long reads. *Front. Genet.* **13**, 816825 (2022).
24. Soutschek, M., Gross, F., Schratt, G. & Germain, P.-L. scanMiR: a biochemically based toolkit for versatile and efficient microRNA target prediction. *Bioinformatics* **38**, 2466–2473 (2022).
25. Wu, J. et al. CircAST: full-length assembly and quantification of alternatively spliced isoforms in circular RNAs. *Genom. Proteom. Bioinform.* **17**, 522–534 (2019).
26. Ye, C. Y. et al. Full-length sequence assembly reveals circular RNAs with diverse non-GT/AG splicing signals in rice. *RNA Biol.* **14**, 1055–1063 (2017).
27. Metge, F., Czaja-Hasse, L. F., Reinhardt, R. & Dieterich, C. FUCHS-towards full circular RNA characterization using RNAseq. *PeerJ* **5**, e2934 (2017).
28. Hossain, M. T., Peng, Y., Feng, S. & Wei, Y. FcircSEC: an R package for full length circRNA sequence extraction and classification. *Int. J. Genom.* **2020**, 9084901 (2020).
29. Qin, Y. et al. Reference-free and de novo identification of circular RNAs. Preprint at *bioRxiv* https://doi.org/10.1101/2020.04.21.050617 (2020).
30. Zheng, Y. & Zhao, F. Visualization of circular RNAs and their internal splicing events from transcriptomic data. *Bioinformatics* **36**, 2934–2935 (2020).
31. Zhang, J. & Zhao, F. Reconstruction of circular RNAs using Illumina and Nanopore RNA-seq datasets. *Methods* **196**, 17–22 (2021).
32. Zhang, J. & Zhao, F. Characterizing circular RNAs using Nanopore sequencing. *Trends Biochem. Sci.* **46**, 785–786 (2021).
33. Jeck, W. R. et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
34. Pandey, P. R., Rout, P. K., Das, A., Gorospe, M. & Panda, A. C. RPAD (RNase R treatment, polyadenylation, and poly(A)+ RNA depletion) method to isolate highly pure circular RNA. *Methods* **155**, 41–48 (2019).
35. Panda, A. C. et al. High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic Acids Res.* **45**, e116 (2017).
36. Zhang, Z. & Han, L. Circular RNAs sequenced at last. *Nat. Biotechnol.* **39**, 811–812 (2021).

37. Vo, J. N. et al. The landscape of circular RNA in cancer. *Cell* **176**, 869–881 e813 (2019).
38. Xin, R. et al. isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nat. Commun.* **12**, 266 (2021).
39. Liu, Z. et al. circFL-seq reveals full-length circular RNAs with rolling circular reverse transcription and nanopore sequencing. *eLife* https://doi.org/10.7554/eLife.69457 (2021).
40. Rahimi, K., Veno, M. T., Dupont, D. M. & Kjems, J. Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nat. Commun.* **12**, 4825 (2021).
41. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* **32**, 453–461 (2014).
42. Vincent, H. A. & Deutscher, M. P. Substrate recognition and catalysis by the exoribonuclease RNase R. *J. Biol. Chem.* **281**, 29769–29775 (2006).
43. Xiao, M. S. & Wilusz, J. E. 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**, 8755–8769 (2019).
44. Fukasawa, Y., Ermini, L., Wang, H., Carty, K. & Cheung, M. S. LongQC: a quality control tool for third generation sequencing long read data. *G3 (Bethesda)* **10**, 1193–1196 (2020).
45. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
46. Wang, Y. et al. GSA: Genome Sequence Archive. *Genom. Proteom. Bioinform.* **15**, 14–18 (2017).
47. Zhang, J. Full-length circular RNA profiling by nannopore sequencing with CIRI-long. https://github.com/bioinfo-biols/CIRI-long, https://doi.org/10.5281/zenodo.7399826 (2022).
48. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
49. Diesh, C. et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. Preprint at https://doi.org/10.1101/2022.07.28.501447 (2022).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (32130020, 32025009, 91940306, 32200530) and the National Key R&D Project (2021YFA1300500, 2021YFA1302000).

Author contributions

L.H. and J.Z. developed the experimental protocol for CIRI-long. L.H., J.Z. and F.Z. wrote the manuscript. F.Z. conceived the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-023-00815-w.

Correspondence and requests for materials should be addressed to Jinyang Zhang or Fangqing Zhao.

Peer review information *Nature Protocols* thanks Michael Clark and Leng Han for their contribution to the peer review of this work.

Received: 14 June 2022; Accepted: 17 January 2023;
Published online: 12 April 2023

Related links
Key reference using this protocol
Zhang, J. et al. *Nat. Biotechnol.* **39**, 836–845 (2021): https://doi.org/10.1038/s41587-021-00842-6