**Abstract:**

Given the similarity of ring molecules and their corresponding linear mRNAs, reconstructing circular RNA sequences (loops) from short RNA sequence reads has proven challenging. High-throughput full-length ring circuits cannot be detected using previous sequencing methods. Compared with previous methods, circular reverse transcription and size selection achieve 20-fold higher enrichment of RNA than total RNA. To reconstruct the sequence of looped molecules, we generated an algorithm called the siRNA identifier using long-read (CIRI-long) sequence data. A new form of intronic self-linked RNA. Our method uses long readouts from the nanopore to reconstruct full-length CircRNA sequences.

**Introduction:**

CircRNA has been associated with disease progression and prediction.Circular RNAs (circRNAs) are a broad class of RNAs with a covalent circular structure that regulate a variety of biological processes. The majority of circRNAs studied so far have been proposed to

* act as microRNA (miRNA) sponges
* act as RNA-binding protein (RBP) sponges
* enhance protein function & encode peptides
* form RNA duplex structures

The majority of current methods depend on the alignment of Illumina short RNA sequence reads (RNA-seq), and their detection ability is severely restricted due to the Illumina sequence reads' short duration. so we found long-read sequencing techniques such as Oxford and nanopore. PacBio sequencing was recently used to assess the full-length loop sequences of reverse transcription products for polymerase chain reaction in a recent report (RT-PCR). Specific PCR primers, on the other hand, were designed to target a subset of candidate ring loops and were only able to detect full-length sequences of selected ring molecules one at a time. Using nanoscale sequencing technology, we present an experimental and computational method (CIRI-long) for mass profiling of full-length looped rings. In contrast to currently available approaches, to discover and recreate ring molecules, and to provide new insights into the diversity of ring molecules and their biosynthesis.

**Related Work:**

Using nanopore sequencing, we present an experimental and computational protocol for detecting full-length circRNA isoforms. To provide clear evidence of the circular structure of the transcripts, several rounds of circular amplification of the same circRNA molecule are reverse transcribed. CIRI-long has been thoroughly tested using simulated datasets and Illumina RNA-seq data. The reliability of our method in decoding the complex array of circular transcripts and splicing events that have not been studied previously has been demonstrated using virtual datasets, Illumina RNA-seq data, and experimental validation. It was discovered that the selection of section length is the main factor affecting RNA detection efficiency, When opposed to Illumina-based traditional sequencing, choosing the right length ( 1 kb) and optimising other processing conditions resulted in a 20-fold rise in circular readings. Many species of vertebrates have been shown to have an excessive amount of ring particles, according to recent research. CircAtlas, for example, gathered a total of 252,811 electrical circuits found in mice, while the GENCODE project only explained 138,835 versions. Check the diversity of the fields specified in the specified fields; unlike other existing methods that extend through BSJs, this approach checks the diversity of the fields specified in the specified fields. This method demonstrated that, each cloned cDNA molecule contained multiple copies of the corresponding full-length CircRNA sequence, and each long reading provided direct evidence of the existence and sequence of CircRNA. CIRI-long can accurately quantify additional alternate generalization and alternative splicing events and provide a more than five-fold increase in alternate generalization events compared to results obtained with Illumina RNA-seq. In addition, nanopore sequence libraries have detected multiple cirexons and AS events Compared to those detected with Illumina data.

**Results:**

### A modified nanopore sequencing protocol for efficient identification of circRNAs:

### To efficiently enrich circRNAs, a customized approach for RNA-seq library preparation was modified .In brief, ribosomal RNA (rRNA) was removed from extracted total RNA.Then additional poly(A)-tailing treatment was applied before RNase R .then Afterwards, reverse transcription was performed with random primers and SMARTer reverse transcriptase to amplify circRNAs by producing long cDNA molecules containing multiple copies of full-length circRNA sequences To obtain clean sequences that contain only circRNAs, reads were first base-called and de-multiplexed using ont-guppy. Then barcodes and primers were trimmed using Porechop with a modified primer library.

### Experimental optimization for capture of full-length circRNAs:

32 nanoscale libraries were established to examine the effects of different experimental conditions on our methodologies in order to establish the best technique (Fig. 2a). In summary, total RNA was isolated from two biological copies of the mouse brains and separated into two batches. Several reverse transcription techniques (with or without A-tailing) ​​were used in each batch, as shown inFigure 2.(“<https://www.nature.com/articles/s41587-021-00842-6/figures/2>”)

### Validation of circRNAs using simulated and Illumina RNA-seq datasets:

NanoSim26 was used to produce simulated nanopore sequencing datasets and then CIRI-long was used to describe circRNAs in the simulated data to assess the dependability of CIRI-long. CIRI-long received an F1 score of 0.92 for read-level analysis, suggesting that circular read detection is very sensitive and accurate (Supplementary Table 2).The simulated length of CCSs detected by CIRI-long was likewise substantially linked with the length of CCSs observed by CIRI-long (Fig. 3a). To compare circRNA levels, the anticipated and simulated coordinates of each circRNA were examined. CIRI-long was able to reliably detect the BSJ site for the great majority(96.57%)detectedcircRNAs,as..showninFig.3b.(https://www.nature.com/articles/s41587-021-00842-6/figures/3). We evaluated the accuracy of raw sequencing reads and CCSs, respectively, due to the relatively high error rate in nanopore sequencing data. The coverage of full-length CCS grew in lockstep with read duration, resulting in CCS accuracy ranging from 92.6 percent to 98.1 percent (5 copies of full-length CCS; Supplementary Figs. 7 and 8).

### A new type of intronic self-ligated circRNAs:

To investigate the variety of circRNA biogenesis, all circRNAs were classified into two groups based on the start and end coordinates of back-splicing processes. CircRNAs were considered annotated if their start and end coordinates matched the annotated splice sites in the GENCODE vM20 annotation; otherwise, they were classified as new.In accordance with past research 90% of exonic circRNAs were formed from annotated splice signals, but the majority of intronic and intergenic circRNAs were categorised as new circRNAs (Fig. 5a).Notably, only a tiny percentage of intronic circRNAs corresponded to identified splice sites. Intronic circRNAs were divided into three groups to determine putative biogenesis processes of these circRNAs: Intronic exonic circRNAs (intronic circRNAs with flanking conventional AG/GT splice signals), intronic self-ligated circRNAs, and lariat intronic circRNAs (Fig. 5b ).Intronic exonic circRNAs employ the typical AG/GT splice signals and have a back-splicing pattern similar to exonic circRNAs. Intronic self-ligated circRNAs, on the other hand, are produced by direct ligation of the host intron's 3′ and 5′ splice sites, which were previously unknown. The 2′–5′ branch of introniclariats..yieldslariat..intronic..circRNAs.we..canknow..more..from(<https://www.nature.com/articles/s41587-021-00842-6/figures/5>).

**Methods:**

### Nanopore library preparation:

Total RNA was obtained from two healthy adult mice, and its quality was tested using an Agilent fragment analyzer system. Total RNA was then separated into two groups for the development of RNase R and A-tailing/RNase R libraries. RNA was initially incubated at 37 °C with 20 U l1 of RNase R, followed by 5 U l1 of poly(A) polymerase (NEB) to add an extra poly(A) tail .Next, SMARTer or Maxima reverse transcriptase was used to reverse transcribe RNase R-treatedRNA. After replacing the 3′ SMART CDS primer II A 5′-AAGCAGTGGTATCAACGCAGAGTACT(30)N-1N-3′with5′-AAGCAGTGGTATCAACGCAGAGTACNNNNNN-3′, 3.5 l of RNA and 1 l of SMARTer CDS random primer (12 M) were combined and incubated at 72 °C for 3 min and 25 °C for 10 min and maintained at 42 °C And operations are continuing to know more visit the link (<https://ciri-cookbook.readthedocs.io/en/latest/CIRI-long_sequencing.html>).

### Nanopore sequencing and base-calling:

### cDNA libraries were generated using the ONT technique SQK-LSK109 and barcoded using EXP-NBD104 and EXP-NBD114 kits, and nanopore sequencing was done on the MinION (MN26543) platform using a FLO-MIN106 flow cell.You can take a peek at (<https://github.com/rrwick/MinION-desktop>)

### Adapter trimming and alignment:

### pycoQC44 was used to do quality control on base-called raw readings (v2.5.0.14). To eliminate barcodes and adapters, reads were trimmed with Porechop v2.0.4 (https://github.com/rrwick/Porechop) using manually selected SMARTer adapter sequences (AAGCAGTGGTATCAACGCAGAGTAC and GTACTCTGCGTTGATACCACTGCTT).

### Northern blotting:

### TRIzol was used to separate RNA from adult mouse brain tissues, which was then treated with or without RNase R. RNA samples were electrophoresed in 8 percent denaturing urea polyacrylamide gel and transferred to Hybond-N+ nylon membranes (GE). The RNA signal was finally detected utilising the Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher Scientific).

### Quantitative PCR validation:

### To assess the reliability of circRNA quantification using nanopore sequencing data, outward divergent primer sets (Supplemental Table 3) were created to measure the expression of 16 randomly selected circRNAs with varying numbers of supported reads ranging from 200 to 30,000.RNase R and the reverse transcription kit mentioned above were used to produce cDNAs for quantitative real-time PCR. The Sanger sequencing method was used to determine the sequences of the PCR products (Supplementary Table 4). In the BSJ sections of five circRNAs, dual peaks or noises caused by homopolymers have been discovered (denoted as ‘ambiguous borders').