SEQUENCING

Circular RNAs sequenced at last

Full-length circular RNAs are sequenced at high throughput using nanopores.

Zhao Zhang and Leng Han

Circular RNAs (circRNAs) in eukaryotes number in the millions, and the list of their biological functions continues to grow[1]. Yet the full-length sequences of most circular RNAs are unknown. Short-read sequencing can identify circRNAs from signatures of the back-splicing reactions that create them, but it usually cannot recover their full-length sequences[2]. Long-read sequencing, which can more easily distinguish circular RNAs from similar linear RNAs, promises to overcome this limitation. Writing in *Nature Biotechnology*, Zhang et al.[3] use optimized nanopore sequencing and a new algorithm, CIRI-long, to reconstruct the complete sequences of circRNAs at high throughput. This state-of-the-art approach, together with another recent method[4], can sequence full-length circRNAs with high efficiency and accuracy, opening new opportunities to understand their biogenesis and functions.

The expanding list of biological functions for circRNAs includes serving as sponges for miRNA and RNA binding proteins[1], as templates for translation, and as regulators of macromolecules[5]. Although knowledge of their full-length sequences would be invaluable to deciphering their functions, sequencing of circRNAs has been hampered by the inability of short-read sequencing to distinguish circular from similar linear RNAs, to detect low-abundance circRNAs, and to accurately quantify circRNA abundance[2]. Recent computational approaches can partially address these issues but are still limited by the length of short reads[2]; for example, the maximum length of reconstructed circRNAs with one short-read method is ~500 bp[3]. Long-read sequencing technologies, such as those from Pacific Biosciences and Oxford Nanopore Technologies, have been applied to address the most challenging problems in genome and transcriptome sequencing, including copy number variation, RNA alternative splicing[6] and, very recently, circRNAs[4].

To develop their optimized protocol for nanopore circRNA sequencing, Zhang et al. began by comprehensively evaluating various experimental strategies

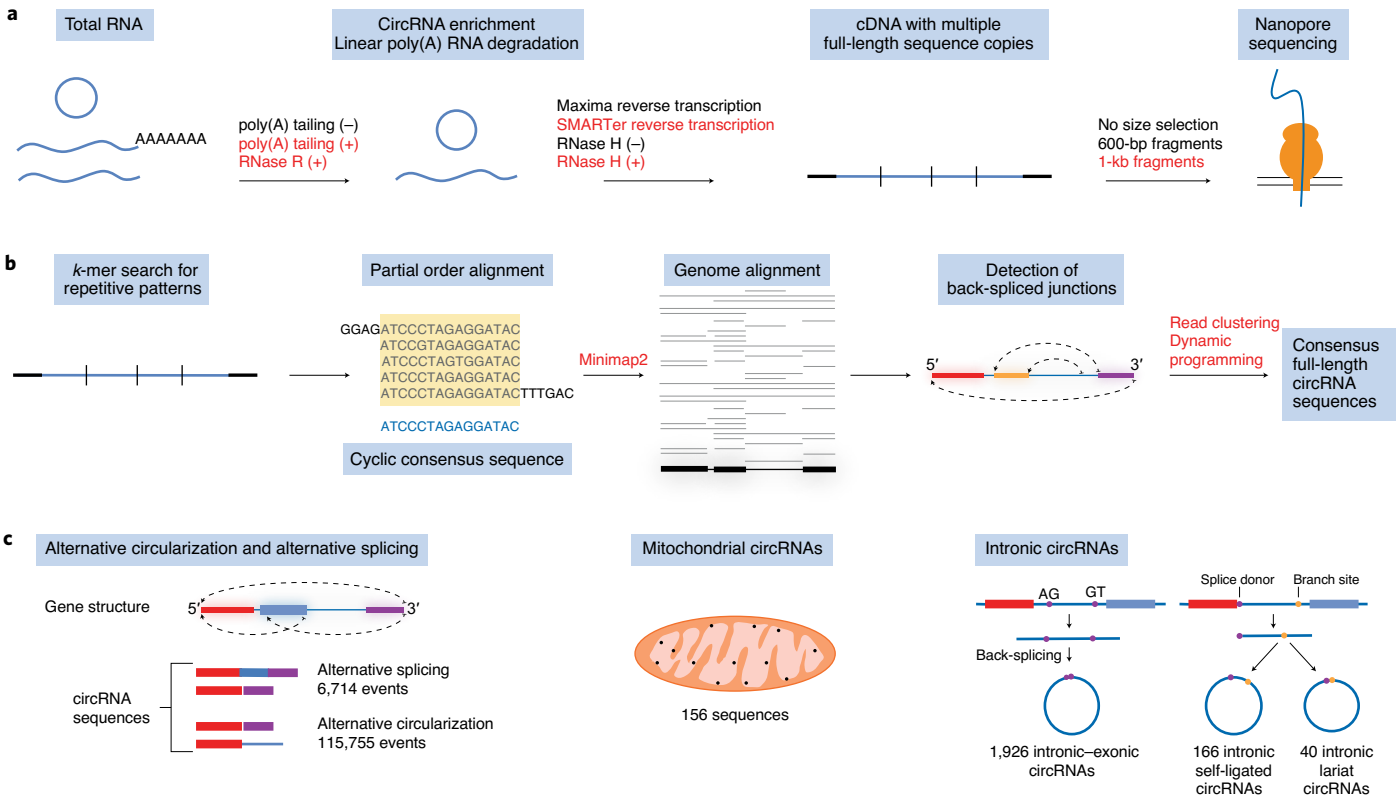

Fig. 1 | Characterization of circRNAs through nanopore sequencing. a, Optimized experimental protocol for circRNA sequencing. **b**, Computational algorithm, CIRI-long, for full-length circRNA detection. **c**, Discovery of alternative circularization and alternative splicing events (left), mitochondrial circRNAs (middle) and intronic self-ligated circRNAs (right).

and incorporating those that were most successful. The resulting protocol combines (i) a poly(A)-tailing treatment to increase the efficiency of linear RNA degradation; (ii) random primers and SMARTer reverse transcriptase to reverse transcribe multiple copies of a circRNA to long cDNA molecules for amplification; and (iii) fragment size selection of cDNAs before nanopore sequencing to further increase efficiency (Fig. 1a). This protocol achieved 20-fold higher enrichment efficiency compared with previous approaches and detected fourfold more circRNAs than short-read sequencing methods (140,588 versus 32,223). The detection of low-abundance circRNAs was improved, with >50% low-abundance circRNAs with fewer than five supporting reads that were not identified through Illumina RNA-seq reads nor archived in the CircAtlas. The advantages of this protocol include the straightforward library preparation with an additional poly(A)-tailing treatment only and the requirement for minimal input total RNA (1 μg) compared to another recent long-read method[4].

To address the challenges of analyzing their sequencing data, Zhang et al. developed a computational algorithm, CIRI-long, that uses *k*-mers to screen repetitive patterns and identify the boundaries of circRNAs (Fig. 1b). It aligns cyclic consensus sequences to correct sequencing errors. Specifically, cyclic consensus sequences are aligned to a reference genome and flanking splicing signals are used to identify high-confidence candidate circRNAs. The researchers aggregated all detectable circRNAs to generate the putative circRNA locus and removed loci with low reliability to generate the final expression matrix. They validated a number of circRNAs by PCR and Sanger sequencing and found highly correlated expression levels, suggesting that the protocol is powerful not only for detecting circRNAs but also for accurately quantifying their expression. CIRI-long is designed to detect circRNAs from long reads, so it should work for long reads generated by other platforms (for example, single-molecule real time sequencing) with appropriate adjustments (for example, removal of certain adaptors from different platforms).

The approach of Zhang et al. has an accuracy of 98.1%, which largely overcomes the limitation introduced by the relatively high error rate of nanopore sequencing. Furthermore, it reduces the amount of input total RNA amount to 1 μg, which is much lower than in previous methods (5 μg (ref. [7]) and 20 μg (ref. [4])). This advantage should substantially increase sequencing throughput, especially for clinical samples.

Zhang et al. applied their protocol to characterize circRNAs in the mouse brain, identifying 115,755 alternative circularization events and 6,714 alternative splicing events (Fig. 1c). These numbers are respectively fivefold and twofold higher than what they achieved by applying a previous short-read sequencing method[7] to the same dataset.

CircRNAs of apparent mitochondrial origin have usually been thought to represent noise arising from sequencing misalignments because mitochondria lack the canonical spliceosome machinery. However, accumulating evidence has demonstrated the possibility of circRNAs derived from mitochondrial genes[8]. Zhang et al. add to this evidence, identifying 156 mitochondrial circRNAs with full-length sequences. Moreover, they identified a novel type of circRNA—intronic self-ligated circRNA—that is circularized from the 5′ splicing site and 3′ splicing site of a whole intron. This type of intronic self-ligated circRNA is enriched in the central nervous system and may exert specific functions in neural processes.

The protocol of Zhang et al. is likely to be further optimized by the research community. Future efforts are needed to reduce the error rate, to increase the sequencing depth, to estimate the expression levels of linear RNA and circRNAs simultaneously, and to avoid potential bias introduced during library preparation (for example, to capture circRNAs that are sensitive to RNase R).

Sequencing circRNAs will enable deeper understanding of their biogenesis[5] via processes such as alternative circulation, alternative splicing and intronic self-ligation. It will also elucidate their biological functions. For example, a recent study[9] used a full-length sequence to discover that a circRNA was translated, as well as the function in myogenesis of the protein product. Another study identified the role of circRNAs in kinase regulation in the innate immune response[10]. The work of Zhang et al., as well as the recent isoCirc method[4], show the advantages of long reads over short reads for this application, demonstrating substantial improvements in efficiency, sensitivity and detection of low-abundance circRNAs. ❐

Zhao Zhang [1,2] and Leng Han [2,3] ✉

[1]MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Fudan University, Shanghai, China. [2]Department of Biochemistry and Molecular Biology, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX, USA. [3]Center for Epigenetics and Disease Prevention, Institute of Biosciences and Technology, Texas A&M University, Houston, TX, USA.

✉*e-mail: leng.han@tamu.edu*

Published online: 04 May 2021

https://doi.org/10.1038/s41587-021-00916-5

References

1. Wu, W., Ji, P. & Zhao, F. *Genome Biol.* **21**, 101 (2020).
2. Zheng, Y., Ji, P., Chen, S., Hou, L. & Zhao, F. *Genome Med.* **11**, 2 (2019).
3. Zhang, J. et al. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00842-6 (2021).
4. Xin, R. et al. *Nat. Commun.* **12**, 266 (2021).
5. Chen, L. L. *Nat. Rev. Mol. Cell Biol.* **21**, 475–490 (2020).
6. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. *Nat. Rev. Genet.* **21**, 597–614 (2020).
7. Gao, Y. et al. *Nat. Commun.* **7**, 12060 (2016).
8. Zhao, Q. et al. *Cell* **183**, 76–93.e22 (2020).
9. Legnini, I. et al. *Mol. Cell* **66**, 22–37.e9 (2017).
10. Liu, C.-X. et al. *Cell* **177**, 865–880.e21 (2019).

Competing interests

The authors declare no competing interests.