Accurate assembly of circular RNAs with TERRACE

Tasfia Zahin[1], Qian Shi[1,†], Xiaofei Carl Zang[2,†], and Mingfu Shao[1,2,*]

[1] Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

[2] Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

Abstract. Circular RNA (circRNA) is a class of RNA molecules that forms a closed loop with its 5' and 3' ends covalently bonded. CircRNAs are known to be more stable than linear RNAs, admit distinct properties and functions, and have been proven to be promising biomarkers. Existing methods for assembling circRNAs heavily rely on the annotated transcriptomes, hence exhibiting unsatisfactory accuracy without a high-quality transcriptome. We present TERRACE, a new algorithm for full-length assembly of circRNAs from paired-end total RNA-seq data. TERRACE uses the splice graph as the underlying data structure that organizes the splicing and coverage information. We transform the problem of assembling circRNAs into finding paths that "bridge" the three fragments in the splice graph induced by back-spliced reads. We adopt a definition for optimal bridging paths and a dynamic programming algorithm to calculate such optimal paths. TERRACE features an efficient algorithm to detect back-spliced reads missed by RNA-seq aligners, contributing to its much improved sensitivity. It also incorporates a new machine-learning approach trained to assign a confidence score to each assembled circRNA, which is shown superior to using abundance for scoring. On both simulations and biological datasets TERRACE consistently outperforms existing methods by a large margin in sensitivity while maintaining better or comparable precision. In particular, when the annotations are not provided, TERRACE assembles 123%-413% more correct circRNAs than state-of-the-art methods. TERRACE presents a major leap on assembling full-length circRNAs from RNA-seq data, and we expect it to be widely used in the downstream research on circRNAs.

Keywords: Circular RNA · Assembly · RNA-seq.

Introduction

Splicing is a ubiquitous and essential post-transcriptional modification of precursor mRNAs. In this process, introns are excised and the two flanking exons are stitched together. Canonical splicing covalently connects

[†]contributed equally to this work

[*]to whom correspondence should be addressed; email: mxs2589@psu.edu

the 3' end of an upstream exon to the 5' end of the immediate downstream exon. A class of noncanonical splicing, known as back-splicing, stitches the 3' end of a downstream exon to the 5' end of an upstream exon via a back-splicing junction (BSJ), forming a closed circular structure called circular RNA (circRNA). Both canonical and back-splicing junctions may experience alternative splicing; consequently, a gene can express multiple circRNAs Ji et al. (2019); Li et al. (2018b). CircRNAs are more prevalent and conserved than previously thought. More than 60% of human genes express at least one circRNAs Ji et al. (2019). Notably, *BIRC6*, a gene that has only 12 linear transcripts, was found to express 243 circRNAs Ji et al. (2019). CircRNAs usually have a lower expression level than their linear forms, but certain circRNAs are significantly more abundant and constitute the major isoform of their genes Vromman et al. (2023); Ji et al. (2019); Szabo and Salzman (2016); Kristensen et al. (2022). Expression of circRNAs is also often tissue and developmental stage-specific in a spatiotemporal pattern Venø et al. (2015); Rybak-Wolf et al. (2015); Memczak et al. (2013).

An increasing volume of research has been evidencing the regulatory functionality of circRNAs and their use in disease diagnosis Kristensen et al. (2022); Memczak et al. (2013). To name a few, *circBIRC6* has been found to play a role in cell pluripotency Yu et al. (2017), similarly, circHIPK3 in cell growth Zheng et al. (2016) and circular *Foxo3* in suppression of cancer cell proliferation and survival Yang et al. (2016). Due to its circular structure, circRNAs are resistant to most RNA degradation mechanisms Li et al. (2018b). In particular, its lack of free 5' and 3' ends protects the molecule itself from exonucleases Suzuki and Tsukahara (2014), making them more stable and have a longer half-life than their linear counterparts Jeck et al. (2013); Jeck and Sharpless (2014). CircRNAs have been reported to serve as novel biomarkers in carcinogenesis and pathogenesis Rybak-Wolf et al. (2015); Kristensen et al. (2022); Wang et al. (2016) and have found their potential use in non-invasive diagnosis Vo et al. (2019).

It is therefore of great interest to detect expressed circRNAs in cells. The RNA sequencing (RNA-seq) protocols that target on linear RNAs or mRNAs with poly(A) tails will neglect circRNAs Kristensen et al. (2022). Tailored RNA-seq experiments have been designed for circRNAs, for example, using RNase R to digest linear RNAs and therefore to enrich circRNAs followed by sequencing Ji et al. (2019); Vo et al. (2019). These approaches are efficient in raising the sensitivity of circRNA detection, but often low-throughput and costly. The total RNA-seq technology, on the other hand, can capture and sequence the entire population of RNA molecules, including circRNAs, in a biological sample. Total RNA-seq is high-throughput, and has been widely used in many studies about RNAs. Large-scale total RNA-seq datasets are available in public repositories such as GEO Clough and Barrett (2016), SRA Leinonen et al. (2010), and ENCODE Consortium et al. (2012), providing a rich resource to further study circRNAs.

Numerous computational methods to detect circRNAs from total RNA-seq were published lately (see Vromman et al. (2023) for a recent review). However, many of them require a fully annotated transcriptome, including CYCLeR Stefanov and Meyer (2023), psirc Yu et al. (2021), CircAST Wu et al. (2019), CIRCexplorer2 Zhang et al. (2016), and CIRCexplorer3 Ma et al. (2019). Those methods' dependency on an existing annotation significantly limits their capability to detect novel circRNAs and their applicability to non-model species without a well-annotated transcriptome. Other tools, including CIRI-full Zheng et al. (2019), Circall Nguyen et al. (2021), CircMiner Asghari et al. (2020), CIRI2 Gao et al. (2018), CIRI-AS Gao et al. (2016), and CircMarker Li et al. (2018a), can be operated annotation-free, are constrained to identifying BSJ only, in a deficiency of assembling full-length circRNAs. Additionally, some methods have to be combined with experimental enrichment of circRNA Stefanov and Meyer (2023). Due to the complexity of alternative splicing and low circRNA abundance, current computational methods unfortunately fail to accurately detect BSJs while also producing exceedingly unsatisfying full-length assemblies. Therefore, the problem of *in silico* assembly of circRNAs, from highly sensitive BSJ identification to full-length circRNA reconstruction, remains largely unresolved.

Here we present TERRACE (accura**T**e ass**E**mbly of circ**R**NAs using b**R**idging and m**AC**hine l**E**arning), a new tool for assembling full-length circRNAs from paired-end total RNA-seq data. TERRACE stands out by assembling circRNAs accurately without relying on annotations, a feature absent in most existing tools. TERRACE starts with a fast, light-weight algorithm to identify back-spliced reads (i.e., reads containing BSJs). We realize that the key to assembling circRNAs is to correctly "bridge" the three fragments in a back-spliced read. We formulate this task as seeking paths in a weighted splice graph that can connect the three fragments and that its "bottleneck" weight is maximized, and solve this formulation using an efficient dynamic programming algorithm. TERRACE also features a new machine-learning model that is trained to assign confidence scores to assembled circRNAs which we show outperforms abundance-based ranking.

Results

Experimental setup

TERRACE. We implemented the algorithm explained in Methods as a new circRNA assembler named TERRACE. TERRACE takes the alignment (in BAM format; can be produced by any RNA-seq aligner such as STAR Dobin et al. (2013) or HISAT Kim et al. (2015)) of total RNA-seq reads as input and produces a list of full-length, annotated circRNAs in the GTF format. TERRACE is designed to assemble reliable circRNAs without the need for an annotation but comes with the flexibility of an optional reference annotation file.

Methods for Comparison. Although there are an abundance of methods that predict BSJs, very few of them produce the full-length annotation (please refer to CYCLeR paper Stefanov and Meyer (2023) for a classification). Full-length circRNA assemblers that do not require a reference annotation are even rarer. We only identify CIRI-full as such a tool and it serves as the state-of-the-art in the field. We include CIRCexplorer2, CIRI-full and CircAST for comparison when the reference annotation is provided. CircAST needs to be provided with a list of BSJs from CIRI2 or CIRCexplorer2; we choose to use CIRI2 as the results are worse when CIRCexplorer2 is used. CYCLeR is not suitable for comparison since it necessitates both control total RNA-seq samples and circRNA-enriched samples, whereas our study exclusively focuses on total RNA libraries. Sailfish-cir and CIRCexplorer3 are tools that primarily target circRNA quantification and hence are not included.

Datasets. We use both real dataset and simulated dataset for evaluation. The real dataset is chosen from the isoCirc Xin et al. (2021) paper (BIGD accession number: PRJCA000751) which consists of 8 human tissue samples (lung, brain, skeletal muscle, heart, testis, liver, kidney, and prostate). Full-length circRNAs for these samples are cataloged by isoCirc using a combination of a reference gene annotation and long reads, which we use as ground-truth. The simulated dataset is generated using CIRI-simulator Gao et al. (2015), previously used by several methods Gao et al. (2016); Zhang et al. (2020); Jia et al. (2019); Nguyen et al. (2021). The expressed circRNAs are available in the simulation which we use as ground-truth. We simulate 10 samples and report the average performance.

Evaluation. We define an assembled circRNA to be *correct* if the coordinates of its BSJ and its intron-chain all exactly match a circRNA in the ground-truth. We then calculate *recall*, defined as the proportion of circRNAs in the ground truth that are correctly assembled, and *precision* which is the percentage of assembled circRNAs that are correct. It is common that a method exhibits high precision but low sensitivity, or vice versa. To still conclude in this case, we report *Fscore*, calculated as $(2\times recall \times precision)/(recall + precision)$. We also draw the precision-recall curve for TERRACE. With the curve we can calculate an *adjusted precision* w.r.t. another method, defined as the precision of TERRACE at a point on the curve when its recall matches the recall of the compared method. We draw and compare two precision-recall curves for TERRACE, by using either the "abundance" or the "score" inferred by the random forest.

Comparison of assembly accuracy

Fig. 1(A-H) shows the accuracy of TERRACE and CIRI-full on the real dataset when the reference annotation is not provided. TERRACE yields a significantly higher number of correct circRNAs than CIRI-full across all samples (228% more on average) at a comparable precision (43% vs 40% on average). The Fscore of

TERRACE is consistently higher than that of CIRI-full (22% vs 10% on average). Both the precision-recall curves of TERRACE are above CIRI-full, proving TERRACE's much enhanced accuracy. Specifically, the

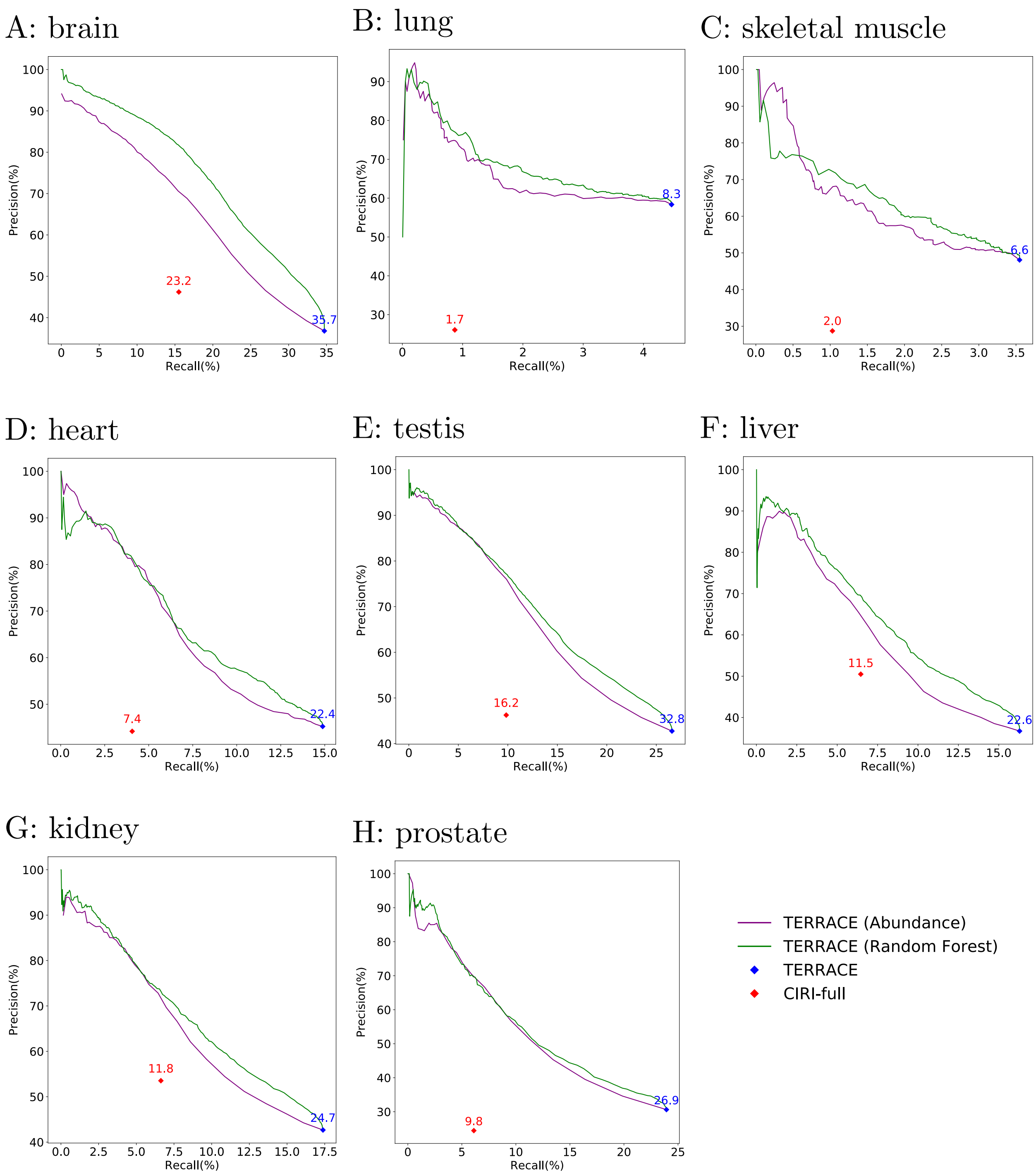

Fig. 1: Comparison of assembly accuracy without annotation. Fscores (%) are indicated on top of data points.

average adjusted precision of TERRACE is 75% (using the random-forest curve; it is 72% using the coverage curve) comparing with CIRI-full at 40%.

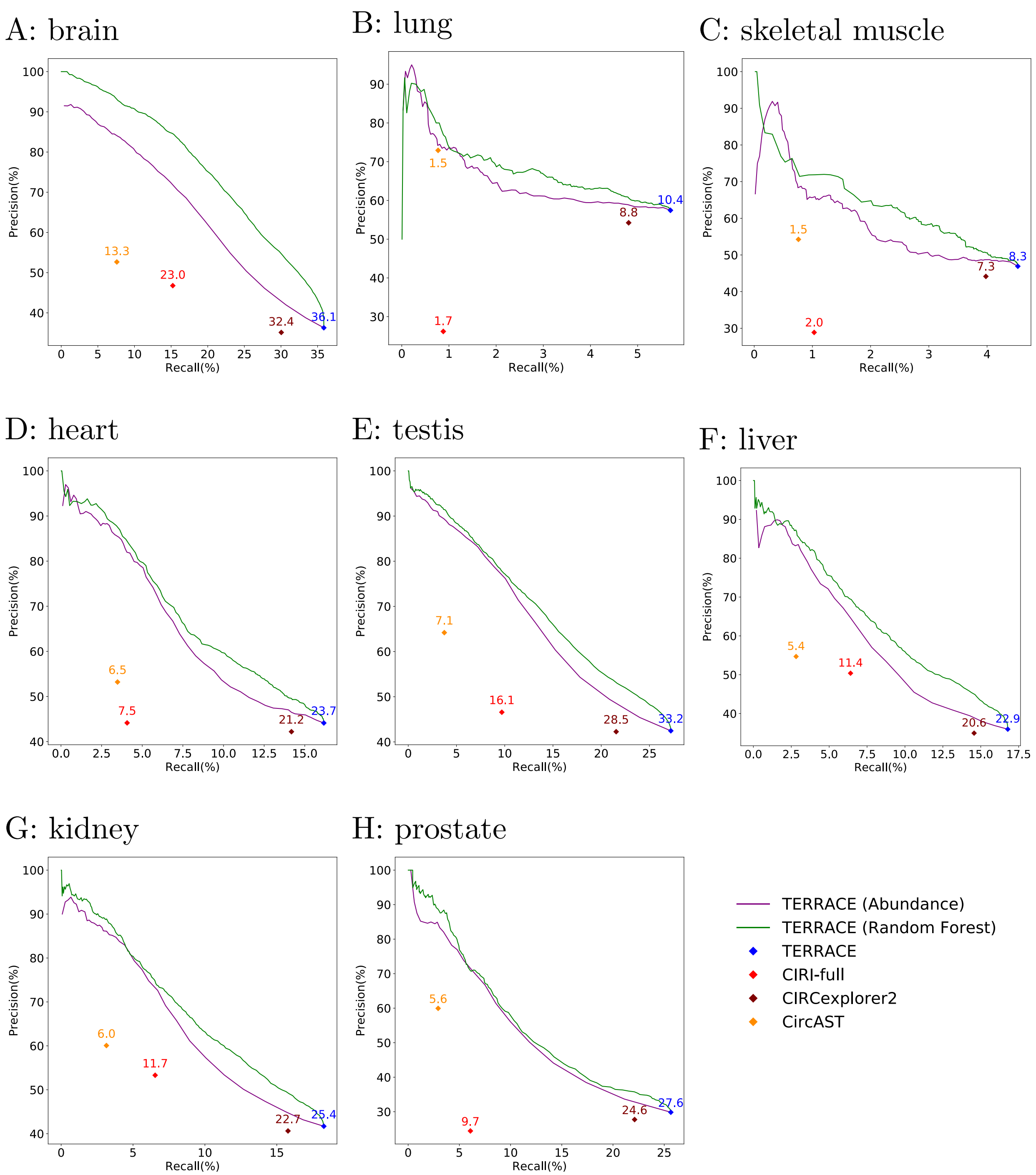

Fig. 2: Comparison of assembly accuracy with annotation. Fscores (%) are indicated on top of data points.

The confidence scores produced by random forest result in a superior curve than that using abundance as the varying parameter. This improvement can be attributed to the machine learning model's ability to learn a more effective and accurate scoring function from a broad class of features (including abundance). Of note, when the abundance is high, simply using it can identify correct circRNAs accurately in a performance similar to random forest (the low-recall regions in the figures). However, when the abundance becomes low, correct and incorrect circRNAs become indistinguishable by just using abundance. The trained scoring function effectively tackles this problem, ensuring a bigger improvement in precision in the high-recall regions.

Fig. 2(A-H) compares the accuracy of different methods running with an annotation. TERRACE reconstructs 17% more correct circRNAs than CIRCexplorer2 while obtaining higher precision on all samples. TERRACE assembles 271% more correct ones than CIRI-full at a comparable precision (42% vs 40% on average). CircAST exhibits higher precision, but at a cost of much reduced recall. Again, TERRACE obtains a higher Fscore than all other methods on all samples. Also, the precision-recall curves of TERRACE consistently lie well above the all other data points, demonstrating its improved accuracy. In particular, the average adjusted precision of TERRACE (using the random-forest curve) is 76% compared to CIRI-full at 40%, 50% compared to CIRCexplorer2 at 40%, and 86% compared to CircAST at 59%.

We also compare the precision recall curves of TERRACE with the curves generated by varying the abundance thresholds of other methods. Supplemental Figs. S4 and S5 confirm the superiority of TERRACE over the curves of other methods. To better quantify the improvement, we measure the area under the precision-recall curve. Since the precision or recall ranges of different methods may vary substantially, we compare TERRACE with each alternative method by locating a shared range (constrained by recall or by precision), and calculate the partial area under the precision-recall curve (pAUC). See Supplementary Tables S3, S4, and S5 for details.

Fig. 3(A-B) shows the accuracy of TERRACE compared to other tools on simulated data. When annotation is not provided, TERRACE identifies an average of 37% more correct circular transcripts and achieves better precision than CIRI-full. In the presence of annotations, TERRACE consistently outperforms CIRI-full and CIRCexplorer2 on all measures. TERRACE exhibits an average precision comparable to CircAST while maintaining much better recall, resulting in a much higher overall Fscore. We conducted additional experiments to evaluate TERRACE's performance across a range of simulation parameters, including read length, circular transcript coverage, and linear transcript coverage. Across all variations, TERRACE demonstrated improved overall performance compared to other methods (see Supplementary Tables S6 and S7). We observe that recall rates for TERRACE is below 80% when the circular coverage is reduced to x5. This because we require the BSJ of a back spliced read to match a splice junction inferred from read alignments. When the sequencing coverage is low, there exist sparse regions where the aligner is unable to infer a splice junction. Due to the absence of these junctions, many correct back-spliced reads get discarded resulting in low recall. One possible approach to address this would be to adjust the splice junction requirement based on coverage depths.

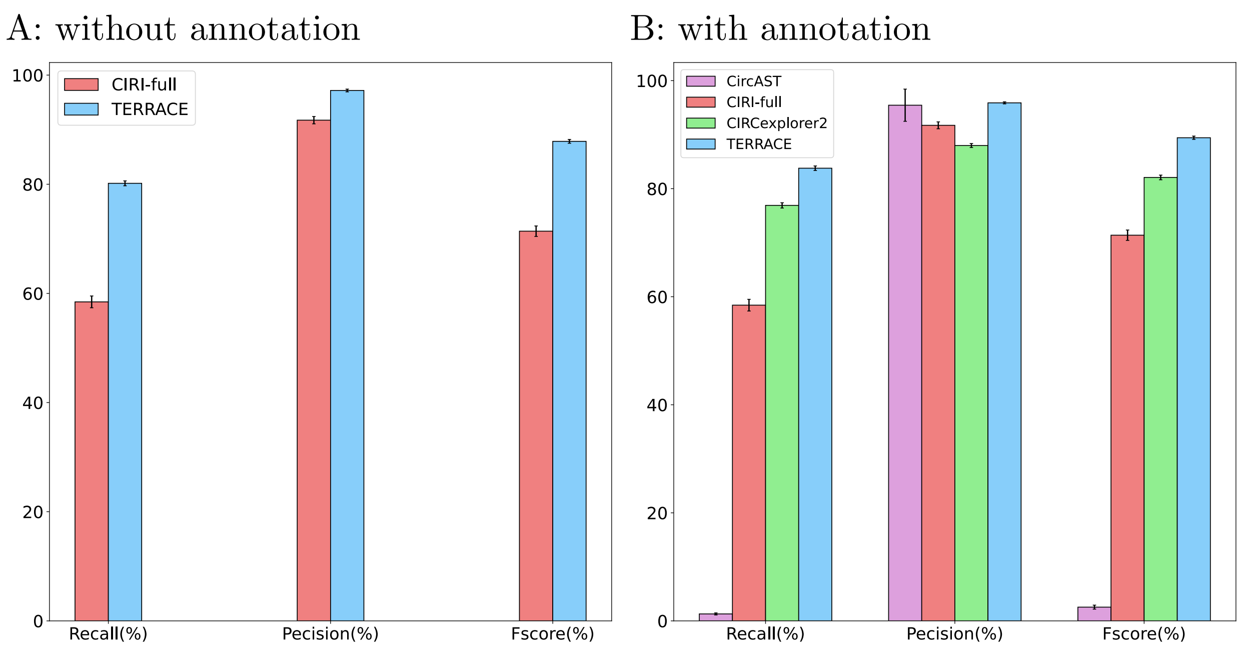

Fig. 3: Average accuracy of different tools on simulated data. The error bars show the standard deviation over 10 simulated samples.

We realize that the recall rates of various methods including TERRACE are quite low on real data (Figs. 1 and 2), particularly in lung and muscle tissues. This trend is also evident in the precision values. One plausible explanation for this discrepancy could be linked to the use of annotated circRNAs from long reads data as the ground truth for evaluation. Upon analyzing the count of annotated circRNAs across the samples and comparing them to the limited number detected by various methods (refer to Supplementary Table S8), we hypothesize that long reads and short-reads total RNA-seq data (which is used here) capture divergent sets of expressed circRNAs. This may result in accurate predictions from short-reads being misclassified as incorrect, leading to an underestimation of both precision and sensitivity. As an evidence, we notice that the read coverage across the gene loci in muscle and lung samples is highly non-uniform: many reads cluster densely within a few number of gene loci, leaving other genes with sparse coverage. This causes all methods, including TERRACE, to construct much less number of correct circRNAs compared to the ground truth. The number of circRNAs annotated using long reads is in the similar range with other samples, so the coverage non-uniformity seems not to exist in the long reads dataset. Nevertheless, we emphasize that although the ground truth used for this study may not fully capture the absolute accuracy, it still serves as a fair benchmark for comparing the relative accuracy of different methods. Given the underestimated recall rates of TERRACE on the biological samples, we resort to simulations to illustrate that achieving high recall is possible in an unbiased setting. The performance of TERRACE on simulated samples strongly reinforces this assertion.

Comparison with long reads assembler

We also conducted experiments to compare the results of TERRACE with a long read assembler, CIRI-long Zhang et al. (2021). CIRI-long analyses various sequencing protocols to measure their effect on circRNA detection and concludes that the optimal procedure for circRNA detection using nanopore technology should include RNaseR, A-tailing, reverse transcription with SMARTer RT under RNase H- conditions and 1kb (long) fragment size selection. We use nanopore sequencing reads from the optimal protocol to run CIRI-long, and illumina short reads (RNaseR and Total) from the same study to run TERRACE. The datasets from two biological replicates of mouse brains are available at BIGD (accession number : CRA003317). Since an established ground truth is not available, we resort to finding number of overlapping circRNAs between CIRI-long and TERRACE as illustrated in Supplementary Figs. S7 and S8. The low number of overlapping circRNAs compared to the number detected uniquely by each tool reconfirms the hypothesis that long-read and short-reads express divergent sets of circRNAs. Additionally, the large number of unique detection by TERRACE is an indication of many *de novo* circRNAs that are likely to be true given the relatively higher precision of TERRACE on the benchmarking datasets.

Discussion

The substantial growth in research dedicated to circRNAs in recent years underscores their significance in biology and medicine. Despite the abundance of experimental and computational techniques designed for circRNA detection, inherent limitations persist within these methods. Current experimental protocols often require special enrichment of libraries for accurate detection while computational methods are hindered by their dependence on annotation and inability to reconstruct full-length molecules. TERRACE made a significant advancement towards closing this critical gap. TERRACE utilized a fast algorithm that effectively detects previously overlooked back-splicing junctions. A bridging system, which we originally proposed for improving (linear) transcript assembly, was re-designed to reconstruct the full-length circRNAs. Instead of using abundance for ranking, TERRACE learned a better score function with a broader class of informative features. TERRACE outperforms existing tools drastically, especially in scenarios where annotations are unavailable. We anticipate widespread adoption of TERRACE, particularly in studies involving species lacking well-annotated transcriptomes.

Further improvements can be made for TERRACE. The precision-recall curves (in Figs. 1 and 2) are not satisfactory. We would investigate the extraction of more features and advanced learning approaches. For instance, we can incorporate the count of splicing positions or partial exons supporting a BSJ as extra features. Additionally, we could explore training a model with complete sequences of *bona fide* circRNAs, possibly sourced from established circular RNA databases, to better differentiate between accurate and erroneous ones. An LSTM model may be used for such sequence-based training. We also consider extending TERRACE to incorporate additional types of data. Leveraging long reads and/or circRNA-enriched libraries for detection may reveal less obvious circRNAs, enhancing the assembly accuracy.

Methods

TERRACE takes the alignment of paired-end total RNA-seq reads and, optionally, a reference annotation as input. TERRACE first identifies back-spliced reads, each of which will be assembled into a set of candidate, full-length circular paths in the underlying splice graph. The candidate paths, optionally augmented by the annotated transcripts, are subjected to a selection process followed by a merging procedure to produce the resultant circRNAs (refer to Supplemental Figures S1-S3 for illustration). A score function is learned that can assign a confidence score to each assembled circRNA (refer to Supplemental Note for a list of features utilized to learn the score function). TERRACE is outlined in Fig. 4. An extended version of the Methods with full descriptions of all the steps is provided as a Supplemental Methods section. Analysis and comparison of runtime and memory usage is available in Supplemental Note and Tables S1-S2 respectively.

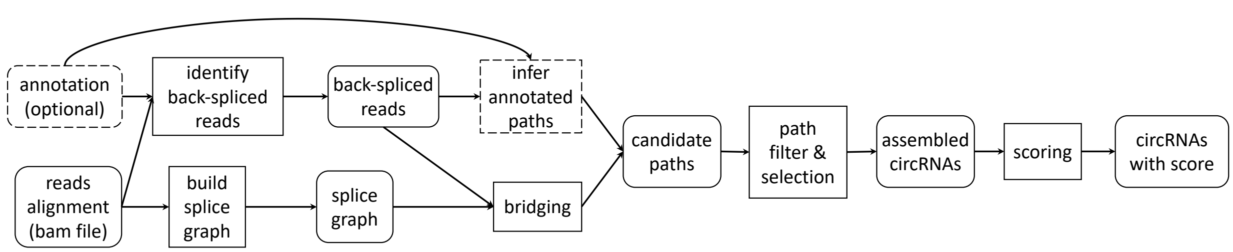

Fig. 4: Outline of TERRACE. Rounded boxes represent data and data structures. Rectangles represent procedures. Dashed boxes indicate optional.

Software Availability

The source code of TERRACE is freely available at `https://github.com/Shao-Group/TERRACE` and also uploaded as Supplemental Code. TERRACE is also available as a conda package at `https://anaconda.org/bioconda/terrace`. The scripts, evaluation pipelines, and instructions that can be followed to reproduce the experimental results of this work is available at `https://github.com/Shao-Group/TERRACE-test`. The alignment files (by STAR) of these samples and the raw sequences and alignment files of the simulated data have been hosted at Penn State Data Commons (DOI: `https://doi.org/10.26208/AZ99-RQ38`). Please refer to Supplemental Material for additional information.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work is supported by the US National Science Foundation (2019797 and 2145171 to M.S.) and by the US National Institutes of Health (R01HG011065 to M.S.). We thank Qimin Zhang for helpful input. We thank Fangqing Zhao and Jinyang Zhang for their assistance with CIRI-full.

Author contributions: T.Z. and M.S. designed the project. All authors designed and implemented the methods. T.Z. conducted the experiments. All authors discussed the results and approved the final manuscript.

References

Asghari H, Lin YY, Xu Y, Haghshenas E, Collins CC, and Hach F. 2020. CircMiner: accurate and rapid detection of circular RNA through splice-aware pseudo-alignment scheme. *Bioinformatics* **36**: 3703–3711.

Clough E and Barrett T. 2016. The gene expression omnibus database. *Statistical Genomics: Methods and Protocols* pp. 93–110.

Consortium EP et al.. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Dobin A, Davis C, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras T. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Gao Y, Wang J, and Zhao F. 2015. Ciri: an efficient and unbiased algorithm for de novo circular rna identification. *Genome biology* **16**: 1–16.

Gao Y, Wang J, Zheng Y, Zhang J, Chen S, and Zhao F. 2016. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nature Communications* **7**: 12060.

Gao Y, Zhang J, and Zhao F. 2018. Circular RNA identification based on multiple seed matching. *Briefings in Bioinformatics* **19**: 803–810.

Jeck WR and Sharpless NE. 2014. Detecting and characterizing circular RNAs. *Nature Biotechnology* **32**: 453–461.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, and Sharpless NE. 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**: 141–157.

Ji P, Wu W, Chen S, Zheng Y, Zhou L, Zhang J, Cheng H, Yan J, Zhang S, Yang P, et al.. 2019. Expanded expression landscape and prioritization of circular rnas in mammals. *Cell reports* **26**: 3444–3460.

Jia Gy, Wang Dl, Xue Mz, Liu Yw, Pei Yc, Yang Yq, Xu Jm, Liang Yc, and Wang P. 2019. CircRNAFisher: a systematic computational approach for de novo circular RNA identification. *Acta Pharmacologica Sinica* **40**: 55–63.

Kim D, Langmead B, and Salzberg S. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**: 357–360.

Kristensen LS, Jakobsen T, Hager H, and Kjems J. 2022. The emerging roles of circRNAs in cancer and oncology. *Nature Reviews Clinical Oncology* **19**: 188–206.

Leinonen R, Sugawara H, Shumway M, and Collaboration INSD. 2010. The sequence read archive. *Nucleic Acids Research* **39**: D19–D21.

Li X, Chu C, Pei J, Măndoiu I, and Wu Y. 2018a. CircMarker: a fast and accurate algorithm for circular RNA detection. *BMC Genomics* **19**: 572.

Li X, Yang L, and Chen LL. 2018b. The biogenesis, functions, and challenges of circular RNAs. *Molecular Cell* **71**: 428–442.

Ma XK, Wang MR, Liu CX, Dong R, Carmichael GG, Chen LL, and Yang L. 2019. CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics, Proteomics & Bioinformatics* **17**: 511–521.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al.. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333–338.

Nguyen DT, Trac QT, Nguyen TH, Nguyen HN, Ohad N, Pawitan Y, and Vu TN. 2021. Circall: fast and accurate methodology for discovery of circular rnas from paired-end rna-sequencing data. *BMC bioinformatics* **22**: 1–18.

Rybak-Wolf A, Stottmeister C, Glažar P, Jens M, Pino N, Giusti S, Hanan M, Behm M, Bartok O, Ashwal-Fluss R, et al.. 2015. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Molecular Cell* **58**: 870–885.

Stefanov SR and Meyer IM. 2023. CYCLeR–a novel tool for the full isoform assembly and quantification of circRNAs. *Nucleic Acids Research* **51**: e10–e10.

Suzuki H and Tsukahara T. 2014. A View of Pre-mRNA Splicing from RNase R Resistant RNAs. *International Journal of Molecular Sciences* **15**: 9331–9342.

Szabo L and Salzman J. 2016. Detecting circular RNAs: bioinformatic and experimental challenges. *Nature Reviews Genetics* **17**: 679–692.

Venø MT, Hansen TB, Venø ST, Clausen BH, Grebing M, Finsen B, Holm IE, and Kjems J. 2015. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biology* **16**: 245.

Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, Wu YM, Dhanasekaran SM, Engelke CG, Cao X, et al.. 2019. The landscape of circular RNA in cancer. *Cell* **176**: 869–881.e13.

Vromman M, Anckaert J, Bortoluzzi S, Buratin A, Chen CY, Chu Q, Chuang TJ, Dehghannasiri R, Dieterich C, Dong X, et al.. 2023. Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nature Methods* **20**: 1159–1169.

Wang F, Nazarali AJ, and Ji S. 2016. Circular RNAs as potential biomarkers for cancer diagnosis and therapy. *American Journal of Cancer Research* **6**: 1167–1176.

Wu J, Li Y, Wang C, Cui Y, Xu T, Wang C, Wang X, Sha J, Jiang B, Wang K, et al.. 2019. CircAST: full-length assembly and quantification of alternatively spliced isoforms in circular RNAs. *Genomics, Proteomics & Bioinformatics* **17**: 522–534.

Xin R, Gao Y, Gao Y, Wang R, Kadash-Edmondson KE, Liu B, Wang Y, Lin L, and Xing Y. 2021. isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nature communications* **12**: 266.

Yang W, Du WW, Li X, Yee AJ, and Yang BB. 2016. Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis. *Oncogene* **35**: 3919–3931.

Yu CY, Li TC, Wu YY, Yeh CH, Chiang W, Chuang CY, and Kuo HC. 2017. The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nature Communications* **8**: 1149.

Yu KHO, Shi CH, Wang B, Chow SHC, Chung GTY, Lung RWM, Tan KE, Lim YY, Tsang ACM, Lo KW, et al.. 2021. Quantifying full-length circular rnas in cancer. *Genome Research* **31**: 2340–2353.

Zhang J, Chen S, Yang J, and Zhao F. 2020. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nature communications* **11**: 90.

Zhang J, Hou L, Zuo Z, Ji P, Zhang X, Xue Y, and Zhao F. 2021. Comprehensive profiling of circular rnas with nanopore sequencing and ciri-long. *Nature biotechnology* **39**: 836–845.

Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Chen LL, and Yang L. 2016. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Research* **26**: 1277–1287.

Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, Luo Y, Lyu D, Li Y, Shi G, et al.. 2016. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature Communications* **7**: 11215.

Zheng Y, Ji P, Chen S, Hou L, and Zhao F. 2019. Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Medicine* **11**: 1–20.