



APPLICATION NOTE

DEBKS: A Tool to Detect Differentially Expressed Circular RNAs



Zelin Liu^{1,#}, Huiru Ding^{2,#}, Jianqi She¹, Chunhua Chen², Weiguang Zhang^{2,*},
Ence Yang^{1,3,4,*}

¹ Institute of Systems Biomedicine, Department of Medical Bioinformatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

² Department of Human Anatomy, Histology & Embryology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

³ Department of Microbiology & Infectious Disease Center, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

⁴ Chinese Institute for Brain Research, Beijing 102206, China

Received 31 January 2020; revised 22 October 2020; accepted 1 January 2021

Available online 23 February 2021

Handled by Fangqing Zhao

KEYWORDS

RNA-seq;
Gene expression;
Circular RNA;
Back-splicing;
Ribo-Zero

Abstract Circular RNAs (circRNAs) are involved in various biological processes and disease pathogenesis. However, only a small number of functional circRNAs have been identified among hundreds of thousands of circRNA species, partly because most current methods are based on circular junction counts and overlook the fact that a circRNA is formed from the host gene by **back-splicing** (BS). To distinguish the expression difference originating from BS or the host gene, we present differentially expressed back-splicing (DEBKS), a software program to streamline the discovery of differential BS events between two rRNA-depleted RNA sequencing (RNA-seq) sample groups. By applying to real and simulated data and employing RT-qPCR for validation, we demonstrate that DEBKS is efficient and accurate in detecting circRNAs with differential BS events between paired and unpaired sample groups. DEBKS is available at <https://github.com/yangence/DEBKS> as open-source software.

Introduction

Circular RNAs (circRNAs) are a class of long non-coding RNAs with a covalently closed continuous loop structure that is formed via back-splicing (BS). Although circRNAs are believed to play important regulatory roles in a number of biological processes [1–4], only a small number of functional circRNAs have been recognized among hundreds of thousands

* Corresponding authors.

E-mail: yangence@pku.edu.cn (Yang E), zhangwg@bjmu.edu.cn (Zhang W).

Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2021.01.003>

1672-0229 © 2022 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

of circRNA species [5–8]. A straightforward strategy to identify functional circRNAs is to detect differentially expressed circRNAs (DECs) that may contribute to certain traits or diseases [9–11].

Previously developed software, such as *find_circ* [12], *CIRI2* [13], and *CIRCexplorer2* [14], focused on the identification of circRNAs, as well as quantification by circular junction counts (CJCs), while recent tools, including *CIRI-full* [15], *CircSplice* [16], and *CircAST* [17], could detect circRNA alternative splicing (AS) and quantify CJCs at the isoform level. As these programs are not able to detect DECs, differentially expressed gene (DEG) detection software (mostly *DESeq2* [18] and *edgeR* [19]) are employed to analyze the data obtained with these CJC-based quantification tools. However, the circRNA expression level is usually associated with the host gene [20], and it is therefore difficult to differentiate whether the difference is due to BS or is only a byproduct of the host gene.

Some programs, such as *DCC* [21], *CIRI2*, *CIRIquant* [22], and *CLEAR* [23], could measure circRNA expression level by the ratio of BS events, instead of CJCs, to normalize the noise of host gene expression. Given that the BS ratio is not applicable for *DESeq2* and *edgeR*, here we report differentially expressed back-splicing (DEBKS), a tool that detects differential BS events between sample groups with paired or unpaired rRNA-depleted RNA-seq data. DEBKS includes four functional modules: *merge* for reformatting junction counts from circRNA detection software, *count* for quantifying expression level of circRNA linear counterpart, *anno* mainly for estimating circRNA length, and *dec* for detecting DECs. Using public, in-house, and simulated RNA-seq datasets, we demonstrate that DEBKS effectively improves the BS ratio calculation and has a better performance in DEC detection than current methods.

Method

Workflow of DEBKS for identifying DECs

DEBKS is designed to analyze DECs from rRNA-depleted RNA-seq data without erasing linear RNA. DEBKS includes four modules: *merge*, *count*, *anno*, and *dec* (Figure 1A). After detecting circRNAs by users, *merge* can collect and reformat CJCs from files generated by detection software. This module can also simultaneously collect linear junction counts (LJCs), if available, which is defined as junction counts across flanking introns of candidate back-spliced sites. Alternatively, *count* is employed for calculating LJCs from BAM files produced by RNA-seq aligners, such as *HISAT2* [24] and *STAR* [25]. Similar to the percentage spliced in [26], the ratio of BS events is measured by percentage back-spliced in (PBSI, ψ), which is estimated by:

$$\psi = (CJC/I_c)/(LJC/I_l + CJC/I_c) \quad (1)$$

where I_l represents the effective linear length and I_c represents the effective back-spliced length (Figure 1B). The values of I_l and I_c are estimated by:

$$I_l = \begin{cases} 4L & \text{if } e \geq 2L \\ 2L + e & \text{if } e < 2L \end{cases} \quad (2)$$

$$I_c = \min(2L, e) \quad (3)$$

where L represents the RNA-seq read length minus the minimum overhang length, and e represents the circRNA length, which can be optionally predicated with gene annotation by *anno* or full-length analysis software, such as *CIRI-full* and *CircAST*. For smaller circRNAs ($e < 2L$), ψ is adjusted with smaller effective lengths. For larger circRNAs ($e \geq 2L$), ψ is equivalent to the calculation of the ratio in other programs, such as *CIRI2* and *CIRIquant*. Finally, the statistical framework of *rMATS* [26], which was developed for detecting differential AS, is employed to compare the difference of ψ ($|\Delta\psi|$) with a user-defined threshold between the paired or unpaired sample groups.

Public RNA-seq datasets

Two public rRNA-depleted RNA-seq datasets from the NCBI Sequence Read Archive database (SRA: SRP050149 and SRP156355) were employed to verify the performance of DEBKS. SRP050149 includes four samples of HEK293 cells with the co-depletion of *ADAR1* and *ADAR2* by RNAi (*ADAR1/2* knockdown) and four untreated samples. SRP156355 includes five invasive ductal carcinoma (IDC) samples and five paired normal samples.

Mouse model of transient middle cerebral artery occlusion

Transient focal ischemia of three male mice (C57BL/6, 6-to-7 weeks old, 21 ± 1 g) was induced by middle cerebral artery occlusion (MCAO) using the endovascular suture method [27]. The operation was performed under isoflurane anesthesia. The right external carotid artery (ECA), internal carotid artery (ICA), and common carotid artery (CCA) were exposed under a stereoscopic anatomical microscope. The suture was gently pushed from the ECA to the ICA to block the origin of the middle cerebral artery (MCA; 9–10 mm). After 1 h, the suture was withdrawn to allow MCA reperfusion. After 6 h, mice were sacrificed. Then, the brain tissue from the ischemic core area on the right side of the brain was removed, and the normal brain tissue from the corresponding area on the left side was removed and utilized as a control. Next, RNA extraction was performed immediately.

RNA extraction

Total RNA of the ischemic core and normal areas was extracted by the FastPure Cell/Tissue Total RNA Isolation Kit (Catalog No. RC101, Vazyme, Nanjing, China) with 10–20 mg of brain tissue according to the manufacturer's instructions.

Library preparation for RNA-seq

The library construction and sequencing of rRNA-depleted RNA were performed by Annoroad Gene Technology (Beijing, China). A total of 3 μ g RNA per sample was used as the input material for rRNA removal with the Ribo-Zero rRNA Removal Kit (Catalog No. MRZG12324, Illumina, San Diego, CA). Libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (Catalog No. E7420, NEB, Ipswich, MA), followed by sequencing on an Illumina HiSeq X platform with 150-bp paired-end reads.

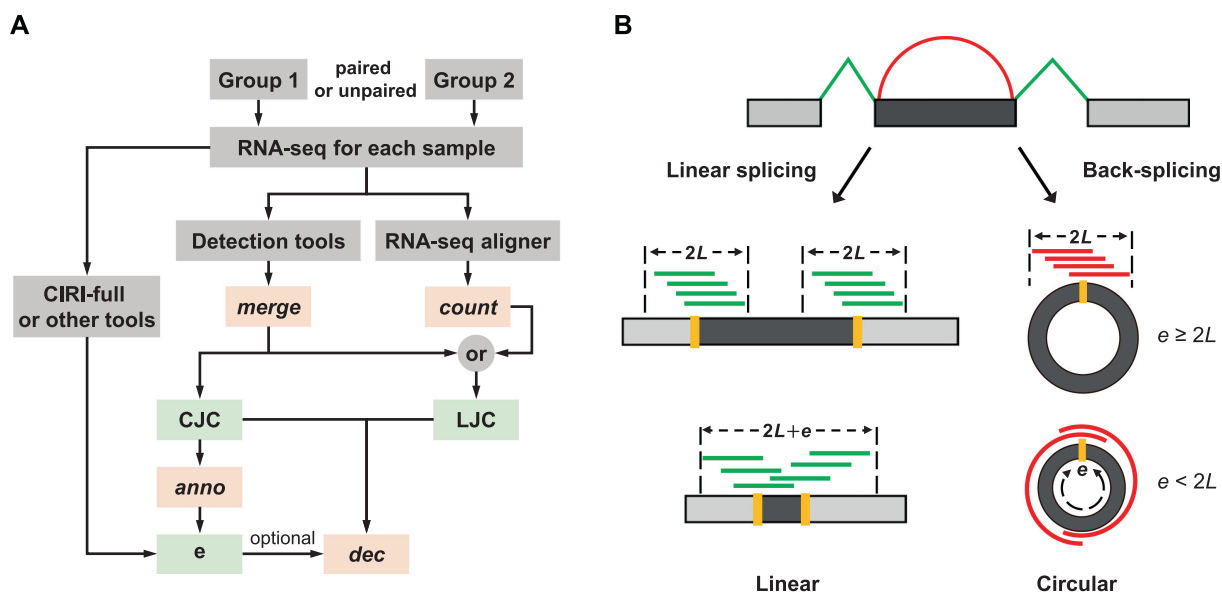


Figure 1 DEBKS workflow

A. The workflow of DEBKS consists of four modules: *merge*, *count*, *anno*, and *dec*. **B.** The schematic diagram of back-splicing. Red reads represent CJCs. Green reads represent LJCs. CJC, circular junction count; LJC, linear junction count; L , RNA-seq read length minus the minimum overhang length; e , circRNA length.

RT-qPCR analysis

Total RNA (500 ng) was reverse-transcribed into complementary DNA (cDNA) by using HiScript III RT Strand SuperMix for qPCR (+gDNA wiper) (Catalog No. R323, Vazyme). For each circRNA, two primers inside the back-spliced site and two primers outside the back-spliced site were designed to detect circular junctions and upstream and downstream linear junctions (Figure S1; Table S1). RT-qPCR was performed by using ChamQ Universal SYBR qPCR Master Mix (Catalog No. Q711, Vazyme) on an ABI7500 system (Applied Biosystems, Foster, CA) according to the manufacturer's procedures, and *Actb* was used as an internal reference gene. Then, ψ was calculated by determining the ratio of the relative expression level of the circular junctions to the sum of the circular junctions and the average of the upstream and downstream linear junctions.

Identification of DECs and DEGs in real RNA-seq datasets

For each RNA-seq dataset, raw reads were mapped to the reference genome (hg19 for SRP050149 and SRP156355; mm10 for mouse RNA-seq) by STAR (v2.5.3a). Next, the LJCs of circRNAs were calculated by DEBKS, while CJCs of circRNAs and counts of host genes were calculated by CIRCExplorer2 [14] (v2.3.8) and featureCounts (v1.6.0) [28], respectively. Only circRNAs with average CJCs ≥ 2 were kept in the following differential expression analysis. DECs and DEGs between the two groups were analyzed using DEBKS and DESeq2, respectively.

Simulation of rRNA-depleted RNA-seq datasets

For the simulated RNA-seq dataset1, we randomly selected 10,000 circRNA host genes based on human gene annotation

(GENCODE v19) for 3 vs. 3 sample groups. We assumed that each host gene contained only one linear product but that half genes contained one circular product, while the other half contained two circular isoforms (same back-spliced site with exon skipping in one isoform). For each gene, the total number of linear and circular RNA molecules was sampled from a normal distribution, with the mean and the standard deviation (SD) being derived from the sum of CJCs and LJCs of the in-house MCAO mouse RNA-seq data. For the two sample groups, the ratios of BS events were simulated with 9500 pairs from a beta distribution ($p = 1, q = 12$) and 500 pairs from a beta distribution ($p = 7, q = 10$). Half of the ratio pairs had differences greater than 5% as DECs, and the ratio of BS events for each sample was simulated with the mean equal to the group ratio and the SD at five different levels, that is, 0.01, 0.02, 0.05, 0.10, and 0.20. Next, the number of circular and linear molecules was calculated with the total number of molecules and the ratio of BS events. For circRNAs with AS, the percentage of one circRNA isoform against total circular molecules was simulated with uniform distribution. Next, the number of each circular isoform molecule was calculated with the percentage and the total number of circular molecules. Finally, simulated rRNA-depleted RNA-seq dataset1 was generated with simulated numbers of circular and linear molecules.

For the simulated RNA-seq dataset2, again, 10,000 circRNA host genes based on human gene annotation (GENCODE v19) were randomly selected for 3 vs. 3 sample groups. Each gene contained one linear product and two circular isoforms from the same back-spliced site. The total number of linear and circular RNA molecules was sampled as described in dataset1. The ratio pairs between groups of each circular isoform against the circular and the linear isoforms were simulated with beta distribution ($p = 1, q = 12$ for 9500 ratio pairs; $p = 7, q = 10$ for 500 ratio pairs). For the two circular isoforms from the same gene, there was one and

only one isoform that had a difference of more than 5% from a DEC at the isoform level. Next, the number of circular and linear molecules was calculated with the total number of molecules and the ratio of BS events. Finally, simulated rRNA-depleted RNA-seq dataset2 was generated with a number of circular and linear molecules.

Identification of DECs in simulated RNA-seq datasets

For each RNA-seq dataset, raw reads were mapped to the reference genome (hg19) by BWA mem [29] (v0.7.17-r1188). circRNAs were detected by CIRI2 (v2.0.6) with gene annotation (GENCODE v19). In the simulated RNA-seq dataset1, DEBKS, CircTest [21], and Fisher's exact test (FET) were employed to identify DECs based on the CJC and LJC from the output file of CIRI2. In the simulated RNA-seq dataset2, the CJC of each isoform of circRNAs were detected by CIRI-full, and the corresponding LJC were obtained from CIRI2.

Results

Identification of DECs with DEBKS in two public RNA-seq datasets

We first evaluated the performance of DEBKS using two public rRNA-depleted RNA-seq datasets. The first dataset includes samples of HEK293 cells with or without the knock-down of both *ADAR1* and *ADAR2*. Given that RNA editing is exclusive to circRNA formation [9], we expected to observe more up-regulated circRNAs in the *ADAR1/2* knockdown treatment. With $|\Delta\psi| > 0.01$ at false discovery rate (FDR) < 0.05 , DEBKS detected 14 DECs, all of which were up-regulated in *ADAR1/2* knockdown cells as expected (Figure 2A).

The second dataset includes RNA-seq data of five IDC samples and five paired normal samples. With or without the parameter '-p', DEBKS detected 377 DECs with the paired model and detected 62 DECs with the unpaired model (Figure 2B). Most (372 of 377) DECs detected by the paired

model were down-regulated in cancer samples, which is in keeping with the results of a previous study [30]. The $|\Delta\psi|$ of 315 DECs only detected in the paired model was significantly lower ($P = 2.6 \times 10^{-5}$, as determined by a two-tailed Student's *t*-test, Figure 2C) than that of the 62 DECs detected by both models, indicating that the statistical framework with paired information can result in considerably improved sensitivity in paired samples.

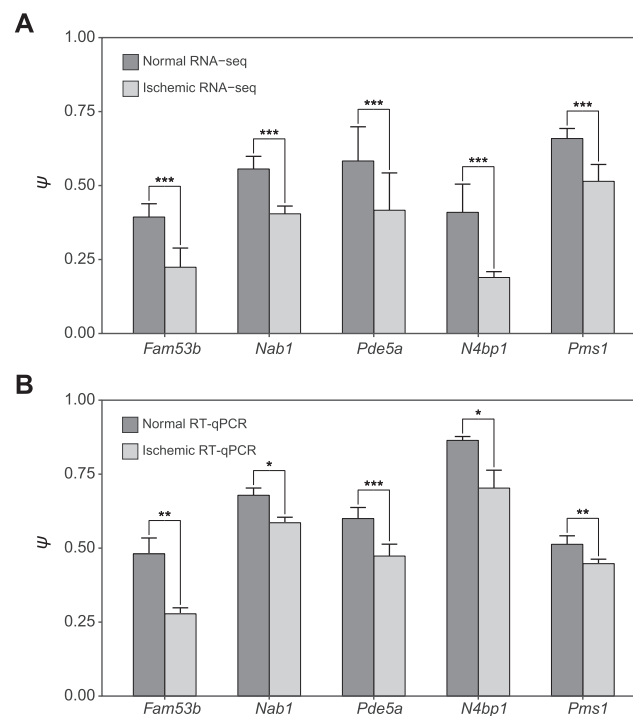


Figure 3 Five circRNAs validated by RT-qPCR in the mouse brain. Comparison of ψ values of five circRNAs in ischemic core and control areas of the mouse brain obtained by RNA-seq (A) and RT-qPCR (B). *P* values were evaluated by DEBKS for RNA-seq, and by single-tailed paired Student's *t*-test for RT-qPCR. *, $P < 0.1$; **, $P < 0.05$; ***, $P < 0.01$.

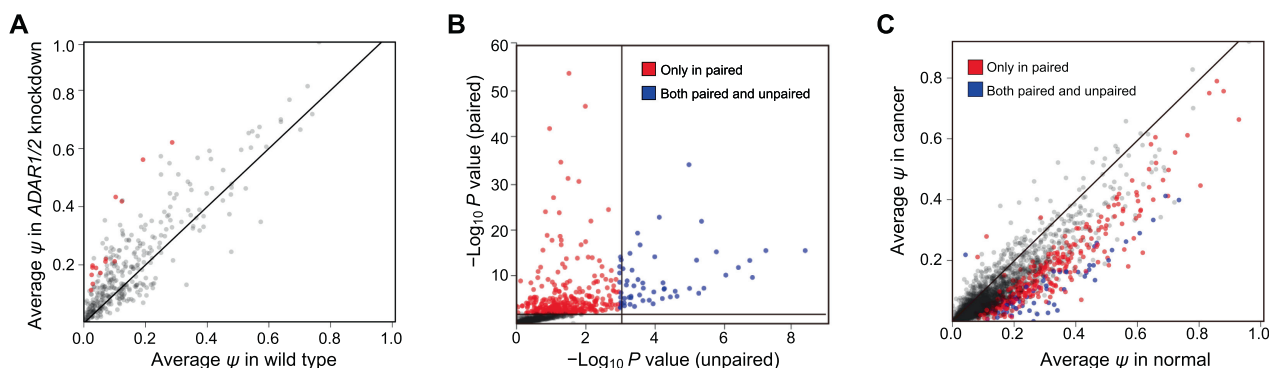


Figure 2 Distribution of DECs in public RNA-seq data

A. Average ψ in *ADAR1/2* knockdown and normal HEK293 groups. Red points represent DECs detected by DEBKS. B. and C. *P* values (B) and average ψ (C) of circRNAs detected by DEBKS with paired or unpaired model between IDC and paired normal samples. Red points represent DECs only detected in paired model, and blue points represent DECs detected in both paired and unpaired models. DEC, differentially expressed circRNA; IDC, invasive ductal carcinoma.

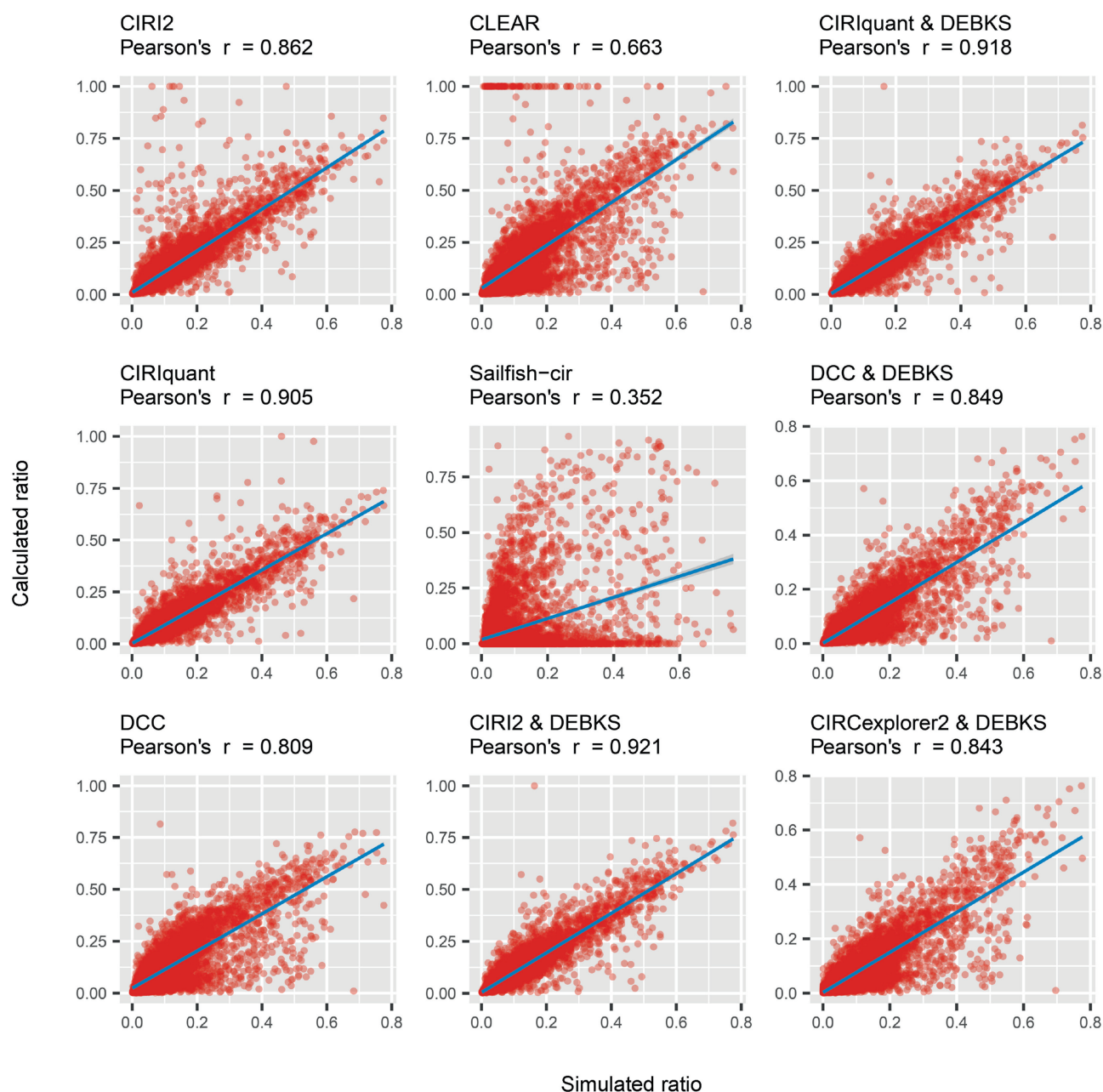


Figure 4 BS ratio calculation with simulated dataset1 is improved in current tools when combining with DEBKS

For CIRI2, CIRIquant, and DCC, BS ratios were calculated by $CJC/(LJC/2 + CJC)$. Because CLEAR and Sailfish-cir used normalized circular (C) and linear (L) expression, we first calculated the ratio by $C/(L/2 + C)$ and $C/(L + C)$ separately and selected $C/(L/2 + C)$ for CLEAR and $C/(L + C)$ for Sailfish-cir based on better Pearson's r . For the ψ calculated by DEBKS, CJCs were from CIRI2, CIRIquant, DCC, and CIRCexplorer2, while LJCs were calculated by DEBKS *count* module with STAR alignment. Blue lines indicate the fitting line. BS, back-splicing.

DEBKS is consistent with RT-qPCR validation in mouse brain

Next, we detected DECs between the ischemic core and control area of the mouse brain with MCAO, which can induce differential expression of circRNAs in the ischemic core [31]. At the threshold of $|\Delta\psi| > 5\%$ and $FDR < 0.05$, a total of 74 DECs (10 up-regulated and 64 down-regulated in the ischemic area)

were identified using rRNA-depleted RNA-seq data by DEBKS with a paired statistical model. Among highly expressed DECs (average $CJC \geq 10$), five were randomly selected for validation by RT-qPCR, and they were all consistent with the RNA-seq data (Figure 3A and B). Notably, we did not detect any DECs with either DEBKS in the unpaired model or CircTest of DCC, suggesting that paired samples

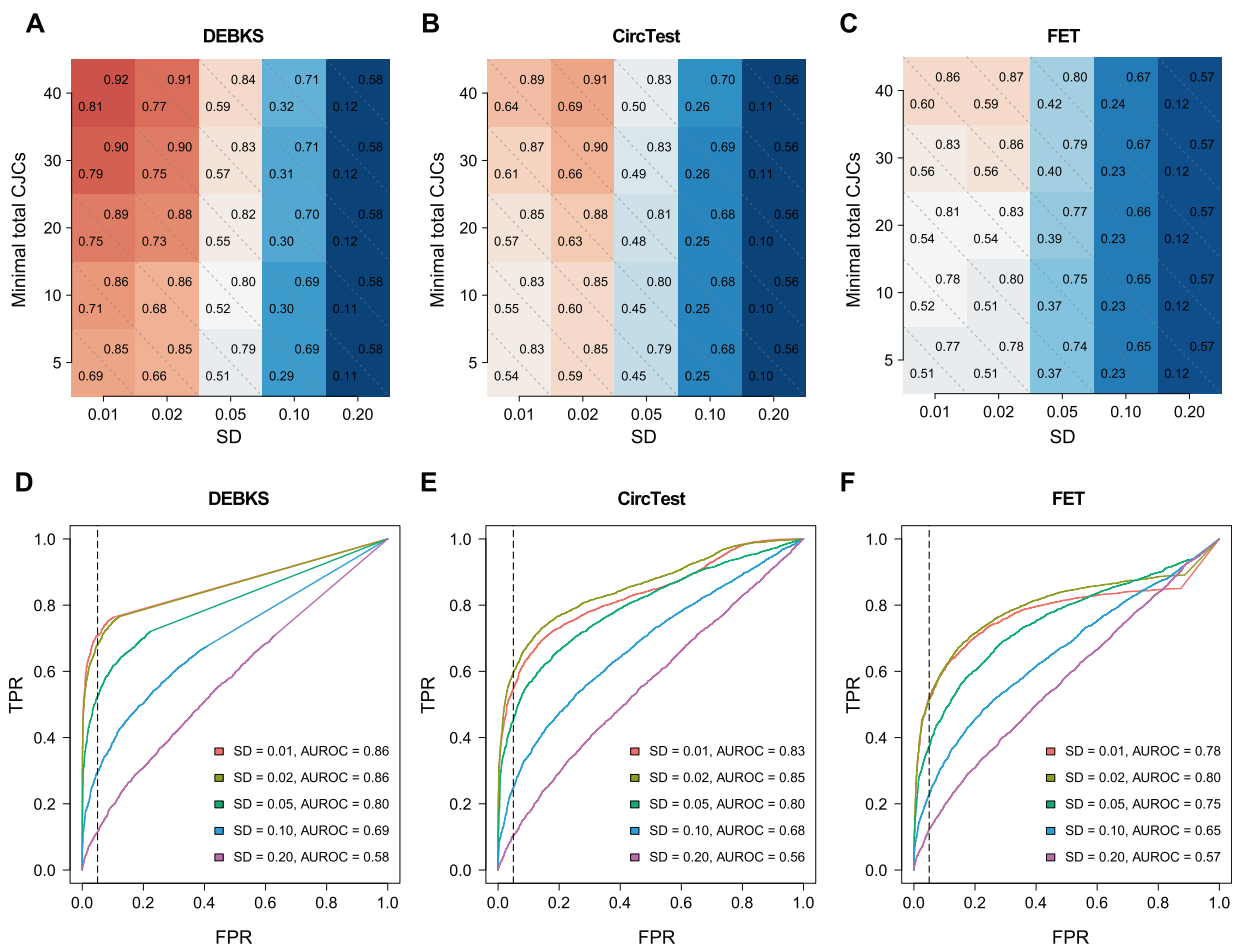


Figure 5 DEBKS shows higher TPR and AUROC than CircTest and FET under various group variances and junction counts with the simulated dataset1

A.–C. The values of TPR at 5% FPR (lower triangle) and AUROC (upper triangle) of DECs identified by DEBKS (A), CircTest (B), and FET (C) for circRNAs with minimal total CJCs of all samples ranging from 5 to 40 counts and SDs from 0.01 to 0.20. **D.–F.** ROC curves of five simulation results at SDs from 0.01 to 0.20 for circRNAs with minimal total CJCs of 10 counts for DEBKS (D), CircTest (E), and FET (F). The vertical dashed line indicates FPR = 5%. TPR, true positive rate; FPR, false positive rate; FET, Fisher's exact test; ROC, receiver operating characteristic; AUROC, area under receiver operating characteristic; SD, standard deviation.

are more powerful for detecting DECs, especially with DEBKS, the only software demonstrated to detect differential BS for paired groups with replicates to date.

DEBKS performs better than current methods in simulation data

To further evaluate the performance of DEBKS, we simulated rRNA-depleted RNA-seq data based on the real junction count distribution with five intragroup SDs, that is, 0.01, 0.02, 0.05, 0.10, and 0.20, which represented the variability of the ratio of BS events among replicates. In the simulated data, the ratio of BS events followed the beta distribution, which could better reflect the ratio distribution in real RNA-seq data (Figure S2A and B). We first compared the performance of BS ratio quantification with other programs, including CIRI2, CIRIquant, DCC, CIRCexplorer2, CLEAR, and sailfish-cir [32] (Figure 4). We found that ψ calculated by CJC from CIRI2 and LJC from DEBKS was most consistent with the simulated ratio. In addition, when combined with LJC

from DEBKS, ψ was more consistent with the simulated ratio than the ratio derived from CIRI2, CIRIquant, or DCC alone, supporting that DEBKS possesses an advantage for ratio calculation.

With minimal total CJCs of all samples ranging from 5 to 40 counts and SD from 0.01 to 0.20 in the simulated dataset1, we evaluated the performance of *dec* of DEBKS, as well as the performance of CircTest and FET for comparison (Figure 5). In terms of the values of area under receiver operating characteristic (AUROC) and true positive rate (TPR) at 5% false positive rate (FPR), DEBKS was determined to perform better than any combination of minimal total CJCs and SD, indicating that DEBKS is powerful in DEC detection for both weakly and strongly expressed circRNAs. However, by applying CIRI-full to calculate CJCs of each circular isoform in the simulated dataset2, we observed that DEBKS possessed low accuracy for detecting DECs at the isoform level (Figure S3), which was primarily observed because of the inaccurate quantification of CJCs (Figure S4).

DEBKS is less heavily influenced by host gene expression than DESeq2

Given the low abundance of circRNAs in rRNA-depleted RNA-seq data, ratio-based methods were considered to be more likely to detect DECs than CJC-based methods, such as DESeq2. By using DESeq2 at $FDR < 0.05$, we failed to detect any DECs in the *ADARI/2* knockdown and MCAO RNA-seq data, while detected 34 DECs in the IDC RNA-seq data, among which 47% (16/34) were also detected by DEBKS. In addition, the host genes of DECs detected by DESeq2 were considerably more likely to be DEGs (odds ratio = 7.33, $P = 4.02 \times 10^{-8}$, two-tailed FET) compared to those detected by DEBKS with the paired model (odds ratio = 1.08, $P = 0.67$, two-tailed FET) and the unpaired model (odds ratio = 1.63, $P = 0.15$, two-tailed FET), suggesting that DEBKS identifies DECs with less influence from host gene expression.

Discussion

Given that circRNAs are BS isoforms of host genes, we developed DEBKS to identify DECs at the splicing level by BS ratio, which normalizes CJs with local LJs. In addition to CJC-based methods, ratio-based methods may help to identify potential functional circRNAs based on the biogenesis mechanism of circRNAs [4]. In our evaluation, DEBKS efficiently identified DECs with less influence from host gene expression. For example, among the 3080 circRNAs detected in IDC RNA-seq data, a strong correlation of \log_2 fold change of expression level (\log_2 FC) was observed between circRNA and the host gene (Pearson's $r = 0.53$, $P < 2.2 \times 10^{-16}$) (Figure S5), indicating that circRNA expression level was significantly affected by host gene expression. In contrast, only a weak correlation was observed between circRNA $\Delta\psi$ and \log_2 FC of the host gene (Pearson's $r = 0.06$, $P = 4.7 \times 10^{-4}$), suggesting that BS events measured by $|\Delta\psi|$ were less strongly influenced by the expression level of the host gene.

At present, several software, including CircTest and CIRI-quant, can detect DECs with ratio-based methods. Compared with simulated data, DEBKS shows better performance for length adjustment of smaller circRNAs and employment of the statistical framework of rMATS. In addition, DEBKS could be applied in paired or unpaired groups with replicates, which is distinct from CIRIquant, which utilizes a ratio-based test to detect DECs between groups without replicates.

Because DEBKS is based on the ratio of CJs and LJs, it is only applicable in RNA-seq analyses of both circRNAs and linear RNAs, such as rRNA-depleted or exome-capture RNA-seq [10]. In addition, intergenic circRNAs without linear counterparts, which are estimated to account for less than 10% of total circRNAs [33], are not suitable for DEBKS.

Ethical statement

All experimental animals were purchased from the Experimental Animal Center of Peking University, China. All protocols involving animals, behavioral testing, surgery, and animal care

were approved by the laboratory animal welfare ethics committee.

Code availability

The source code of DEBKS and the script to simulate the RNA-seq dataset are available at <https://github.com/yan-gence/DEBKS>.

Data availability

RNA-seq data of the mouse MCAO model have been deposited in the Genome Sequence Archive [34] at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformatics (GSA: CRA002696), which are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>.

CRedit author statement

Zelin Liu: Conceptualization, Software, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Huiru Ding:** Methodology, Validation, Writing - review & editing. **Jianqi She:** Investigation, Data curation, Writing - original draft. **Chunhua Chen:** Resources, Project administration, Funding acquisition. **Weiguang Zhang:** Supervision, Resources, Writing - review & editing. **Ence Yang:** Conceptualization, Supervision, Resources, Funding acquisition, Writing - original draft, Writing - review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared that they have no competing interests.

Acknowledgments

This study was supported by the Beijing Municipal Science and Technology Commission, China (Grant No. Z181100001518005) and the Chinese Institute for Brain Research, Beijing, China to EY, the National Natural Science Foundation of China (Grant No. 81873769) to CC, and the Innovation Fund for Outstanding Doctoral Candidates of Peking University Health Science Center, China (Grant No. BMU2020BSS001) to ZL.

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2021.01.003>.

ORCID

ORCID 0000-0002-3516-3999 (Zelin Liu)
ORCID 0000-0003-1798-5806 (Huiru Ding)

ORCID 0000-0001-7210-6638 (Jianqi She)
 ORCID 0000-0002-8502-8189 (Chunhua Chen)
 ORCID 0000-0001-5055-4907 (Weiguang Zhang)
 ORCID 0000-0002-9526-2737 (Ence Yang)

References

- [1] Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet* 2013;9:e1003777.
- [2] You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* 2015;18:603–10.
- [3] Chen LL. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol* 2020;21:475–90.
- [4] Kristensen LS, Andersen MS, Stagsted LVW, Ebbesen KK, Hansen TB, Kjems J. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 2019;20:675–91.
- [5] Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;20:1666–70.
- [6] Dong R, Ma XK, Li GW, Yang L. CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinformatics* 2018;16:226–33.
- [7] Wu W, Ji P, Zhao F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol* 2020;21:101.
- [8] Liu M, Wang Q, Shen J, Yang BB, Ding X. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol* 2019;16:899–905.
- [9] Ivanov A, Memczak S, Wylter E, Torti F, Porath HT, Orejuela MR, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* 2015;10:170–7.
- [10] Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, et al. The landscape of circular RNA in cancer. *Cell* 2019;176:869–81.
- [11] Chen S, Huang V, Xu X, Livingstone J, Soares F, Jeon J, et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell* 2019;176:831–43.
- [12] Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;495:333–8.
- [13] Gao Y, Zhang J, Zhao F. Circular RNA identification based on multiple seed matching. *Brief Bioinform* 2018;19:803–10.
- [14] Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* 2016;26:1277–87.
- [15] Zheng Y, Ji P, Chen S, Hou L, Zhao F. Reconstruction of full-length circular RNAs enables isoform-level quantification. *Genome Med* 2019;11:2.
- [16] Feng J, Chen K, Dong X, Xu X, Jin Y, Zhang X, et al. Genome-wide identification of cancer-specific alternative splicing in circRNA. *Mol Cancer* 2019;18:35.
- [17] Wu J, Li Y, Wang C, Cui Y, Xu T, Wang C, et al. CircAST: full-length assembly and quantification of alternatively spliced isoforms in circular RNAs. *Genomics Proteomics Bioinformatics* 2019;17:522–34.
- [18] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
- [19] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- [20] Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* 2017;66:22–37.
- [21] Cheng J, Metge F, Dieterich C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* 2016;32:1094–6.
- [22] Zhang J, Chen S, Yang J, Zhao F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun* 2020;11:90.
- [23] Ma XK, Wang MR, Liu CX, Dong R, Carmichael GG, Chen LL, et al. CIRCexplorer3: a CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics Proteomics Bioinformatics* 2019;17:511–21.
- [24] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015;12:357–60.
- [25] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- [26] Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* 2014;111:E5593–601.
- [27] Moisse K, Welch I, Hill T, Volkening K, Strong MJ. Transient middle cerebral artery occlusion induces microglial priming in the lumbar spinal cord: a novel model of neuroinflammation. *J Neuroinflammation* 2008;5:29.
- [28] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- [29] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- [30] Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, et al. Correlation of circular RNA abundance with proliferation-exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep* 2015;5:8057.
- [31] Duan X, Li L, Gan J, Peng C, Wang X, Chen W, et al. Identification and functional analysis of circular RNAs induced in rats by middle cerebral artery occlusion. *Gene* 2019;701:139–45.
- [32] Li M, Xie X, Zhou J, Sheng M, Yin X, Ko EA, et al. Quantifying circular RNA expression from RNA-seq data using model-based framework. *Bioinformatics* 2017;33:2131–9.
- [33] Liu Z, Ran Y, Tao C, Li S, Chen J, Yang E. Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex. *Genome Biol* 2019;20:99.
- [34] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* 2021;19:578–83.