

ExRNA Sequencing Results of Different Volumes of Serum

Liujing Wang, Xiangwei Zhao*

State Key Laboratory of Bioelectronics, School of Biological Science & Medical Engineering, Southeast University, Nanjing, China
Email: m15926479650@163.com, *xwzhao@seu.edu.cn

How to cite this paper: Wang, L.J. and Zhao, X.W. (2020) ExRNA Sequencing Results of Different Volumes of Serum. *Journal of Biosciences and Medicines*, 8, 201-208.

<https://doi.org/10.4236/jbm.2020.86019>

Received: May 24, 2020

Accepted: June 27, 2020

Published: June 30, 2020

Abstract

Circulating extracellular RNAs (exRNAs) are very important potential biomarkers of disease. But few have been reported the extracellular RNA profiles in serum. In this study, we tried to use RNA-seq technology to analyze the extracellular RNA by using low volume peripheral blood serum and compared the sequencing results between different volumes of serum (500 μ l, 1.5 ml, 2.5 ml, 3.5 ml). According to the results, we found that most of the exRNA in the serum is lncRNA, protein_coding and MT-rRNA. And when the volume of serum is between 0.5 - 3.5 ml, the numbers and reads of exRNA, including the reads of circRNA, have not changed significantly. It indicates that the sequencing result of exRNA in serum has no significance with the initial serum volume.

Keywords

Extracellular RNAs, Serum, RNA-seq Technology, Serum Volume

1. Introduction

Liquid biopsies have gradually become a rapidly developing technique of in vitro diagnostic (IVD) due to its low invasiveness and availability [1]. Most body fluids including serum, urine, saliva all can be the carriers of disease biomarkers [2] [3]. In the 1940s, extracellular nucleic acids (exDNA and exRNA) were found in the blood [4]. The nucleic acids were passively or actively released into the blood from various tissues and cells [5]. Therefore, the nucleic acid in the serum provides a potential window into the health status of the individual [5] [6]. For example, exDNA testing found in serum has greatly promoted the development of non-invasive testing and diagnosis [7], including prenatal testing and fetal genome sequencing for fetal aneuploidy, as well as cancer diagnosis and monitor-

ing [8].

Since the discovery of extracellular RNAs (exRNAs) in the circulatory system and other body fluids, many studies have been working hard to classify and evaluate whether exRNAs can be used as biomarkers of diseases [5] [6] [7] [8]. There have many exRNAs been identified, such as mRNA, long non-coding RNA, PiwiRNA (piRNA), small nuclear RNA (snRNA), small nuclear RNA (snoRNA), ribosomal RNA (rRNA), transport RNA (tRNA), Y-RNA, microRNA(miRNA) and circular RNA (circRNA), etc. [9] [10]. Because exRNAs in serum are exposed to the environment of RNase R, the exRNAs, especially mRNA, are very unstable in serum, and RNA are usually in a fragmented state, it is a challenge to characterize RNA expression profiles in serum [10] [11].

In this study, we perform the RNA-seq library preparation in different volumes of serum, making the global measurements of RNA species, including linear RNA and circRNA, and evaluate the sequencing results of exRNA in different volumes of serum.

2. Materials and Methods

2.1. Study Population and Serum Sample Collection

We collected peripheral blood (about 2 ml) from 16 individuals, then the blood samples were centrifugated at 3000 rpm for 10 min at 4°C. The serum samples were collected and 500 µl serum of every individual were mixed, then the mixed serum samples were divided into 500 µl, 1.5 ml, 2.5 ml, 3.5 ml, stored at -80°C until use [12]. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Jiangsu Province Hospital.

2.2. Serum exRNA Extraction

The total RNA from serum (extracellular RNA) was extracted by using miR-Neasy Serum/Plasma Kit according to the manufacturer's instructions with modifications. Simply, every 500 µl serum add 1 ml TRIzol reagent, mixed by vortex and incubated at room temperature for 5 min, followed 250 µl chloroform were added to each tube, vortexed and incubated at room temperature for 15 min [13]. After centrifuge the sample at 12,000 g for 15 minutes at 4°C, the aqueous phase was placed into a new tube. Then we added equal volume of 100% isopropanol to the aqueous phase and incubated at room temperature for 40 min. After that, the aqueous phase transferred to one spin column according to the instructions. Finally, exRNA was dissolved 25 µl RNase-free water. The quality of total exRNA was determined by NanoDrop ND-1000.

2.3. Total exRNA Sequencing Library Preparation

DNA contaminations from extracted total RNA were removed by Recombination DNase I according to the manufacturer's instructions. After incubated at

37°C for 25 min, RNA was purified by VAHTS RNA clean beads following the protocol. Because the concentration of extracted RNA is too low, we adopted the exRNA library preparation without rRNA depletion. The exRNA-seq library constructed following manufacturer instructions using the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina®. And exRNA libraries were amplified by PCR cycles using Illumina compatible index primers. The amplified libraries were resolved on a 2% agarose gel and recollect the 250 - 500 bp cDNA fragments by using TIANgel Midi Purification Kit. Finally, 4 sequencing libraries were pooled into a single sequencing lane and sequenced by Hi-seq-PE150 (Illumina).

2.4. Analysis of Sequencing Data

Raw data of fastq format were firstly processed through in-house perl scripts. In this step, reads containing ploy-N, adapter and low-quality reads were filtered. Q20, Q30 and GC content the clean data were calculated. All the downstream analyses were based on the clean data with high quality. HISAT2 (version 2.1.0) was used to align exRNA-seq paired-end clean reads to the reference genome (hg38). The number RNA transcripts were counted by feature Counts [14] (version 1.6.3) using the aligned reads and the gene annotation file (genome.v30.annotation.gtf) as the input file. Finally, using CIRI (CircRNA Identifier) software [15] to predict the “junction read” of the sequence, and identify circRNA from these “junction read”, and use circbase to annotate and count circRNA.

3. Results

3.1. The Results of Serum exRNA Library Preparation

We prepared 4 different volumes (500 µl, 1.5 ml, 2.5 ml, 3.5 ml) of serum samples, then the exRNAs were extracted and the RNA-seq library were prepared. Using Qubit2.0 to measure the concentration of the extracted RNA, but we found that the RNA concentration was too low to be detected. And using agarose gel electrophoresis (**Figure 1(A)**) and Agilent 2100 (**Figure 1(B)**) to show the quality of cDNA library, and we found the library fragment size was 250 - 500 bp, the average concentration of the cDNA library is 1.84 nmol/l.

3.2. The Quality of Sequencing Data

We use fastQC software to analyze the quality of sequencing data. In the exRNA sequencing data, most of the reads reached the Q30 standard (the error rate of each base is 0.1%) (**Figure 2(A)**). The average GC content of each read also tends to the standard level (**Figure 2(B)**).

3.3. Summary of the Sequencing Alignments

We evaluated the results of serum exRNA sequencing of different volumes (500 µl, 1.5 ml, 2.5 ml, 3.5 ml) and the circRNA prediction. We found that the

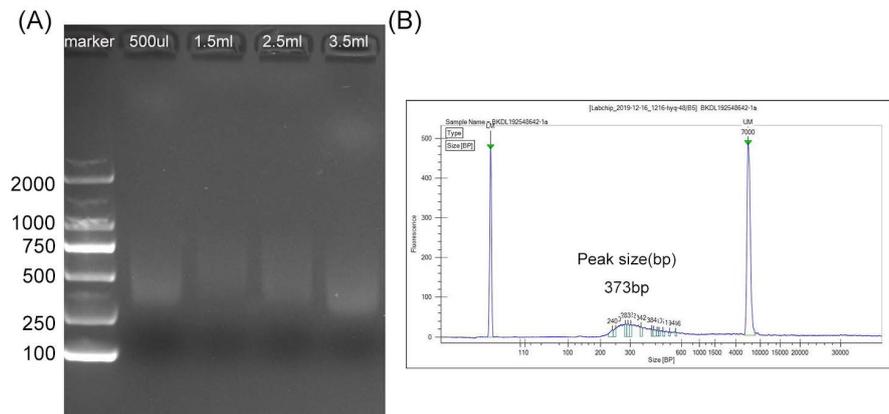


Figure 1. Preparation of four sample libraries. (A) Agarose gel electrophoresis was used to detect 4 libraries to prepare amplification products; (B) 4 samples mixed samples were prepared from Agilent 2100 detection library.

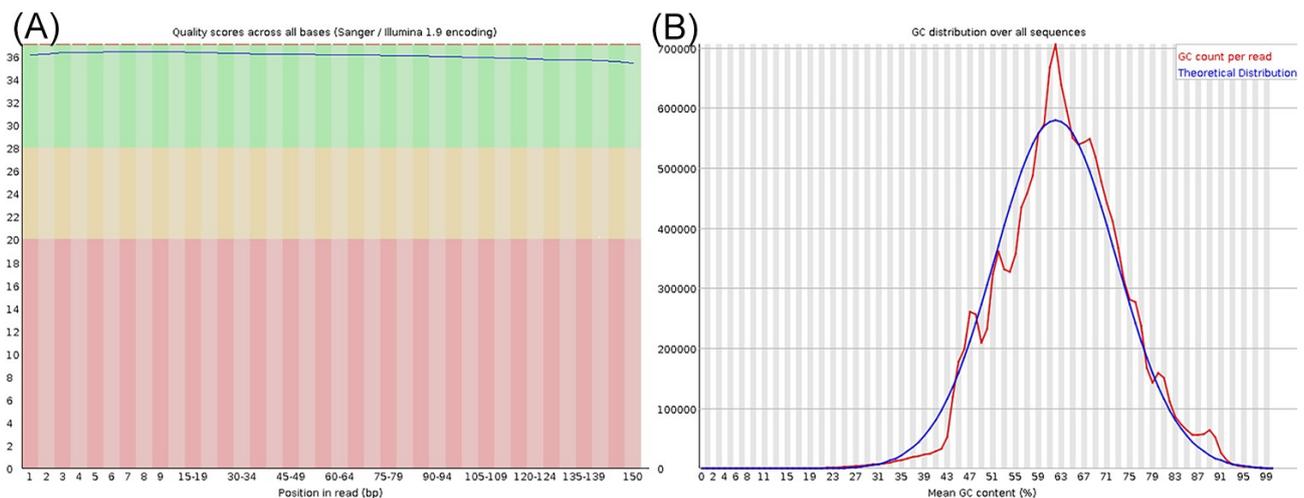


Figure 2. FastQC report of sequencing data. (A) Per base sequence quality; (B) Per sequence GC content.

number of genes, the reads of gene, circRNA number, and circRNA counts obtained from the analysis of different volumes of serum exRNA sequencing did not change with the volume (Table 1).

3.4. ExRNA Biotypes Distribution Analysis

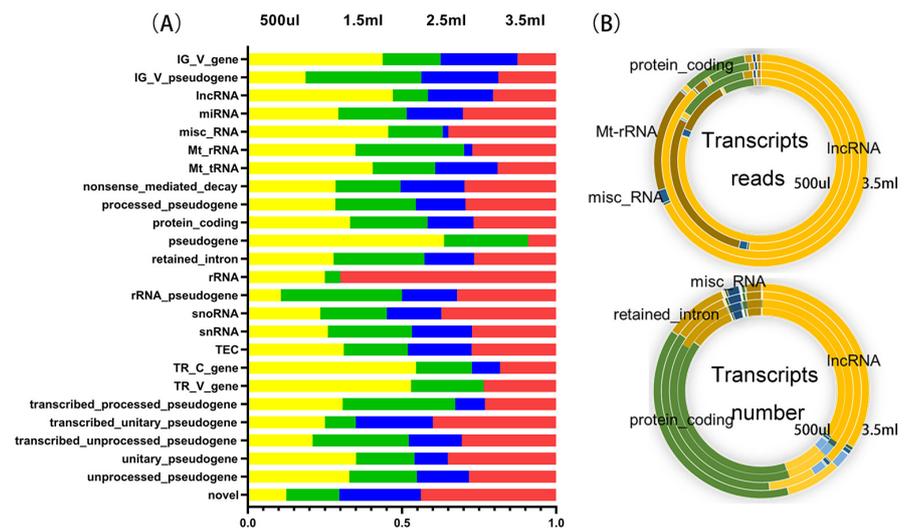
To get a better understanding about different volumes of serum's exRNA biotypes distribution, we analyzed the total aligned reads of each biotype between different volumes of serum (Figure 3(A)) and visualized this using a donut graph where each layer represents the reads and numbers of RNA biotype for each volume (Figure 3(B)). The results show the reads and numbers of the main biotypes of different volumes serum did not change with the volume.

3.5. Distribution of Highly Expressed Transcripts

We select the transcripts that have over 20 reads in each sample and analyze their expression level in different volumes of serum. The results show that there

Table 1. Overview of exRNA sequencing of different volumes of serum.

	the volumes of serum			
	500 μ l	1.5 ml	2.5 ml	3.5 ml
clean reads	15406446	17083788	15178400	21202145
mapping rate	57.04%	54.44%	79.89%	69.99%
transcripts number	5541	5274	3772	6009
transcripts reads	202745	147639	206618	321286
circRNA number	40	27	19	21
circRNA counts	200	110	95	103

**Figure 3.** The changes of transcript biotype in different volumes of serum. (A) Percentage of total aligned reads for RNA from different volumes of serum by ENSEMBL biotype; (B) The ratio of transcripts reads and numbers of RNA biotype for each volume.

is no obvious connection between gene expression and serum volume, and most genes are averagely distributed among the 4 samples (**Figure 4**).

3.6. Distribution of circRNA in Serum

We used CIRI software to predict the circRNA of RNA sequencing data from 4 samples. In total, we found that the total reading of 48 circRNA in serum was greater than 3 (**Figure 5(A)**). In general, most circRNAs have very low abundance, and only a few circRNAs are detected with higher counts (**Figure 5(B)**). However, the highly expressed circRNA expressed the similar counts in the 4 samples.

4. Discussion

RNA expression profiles can reveal individual health status. Peripheral blood serum is an ideal material for studying gene expression because of its easy to access and low invasiveness. But due to the small amount and fragmentation of

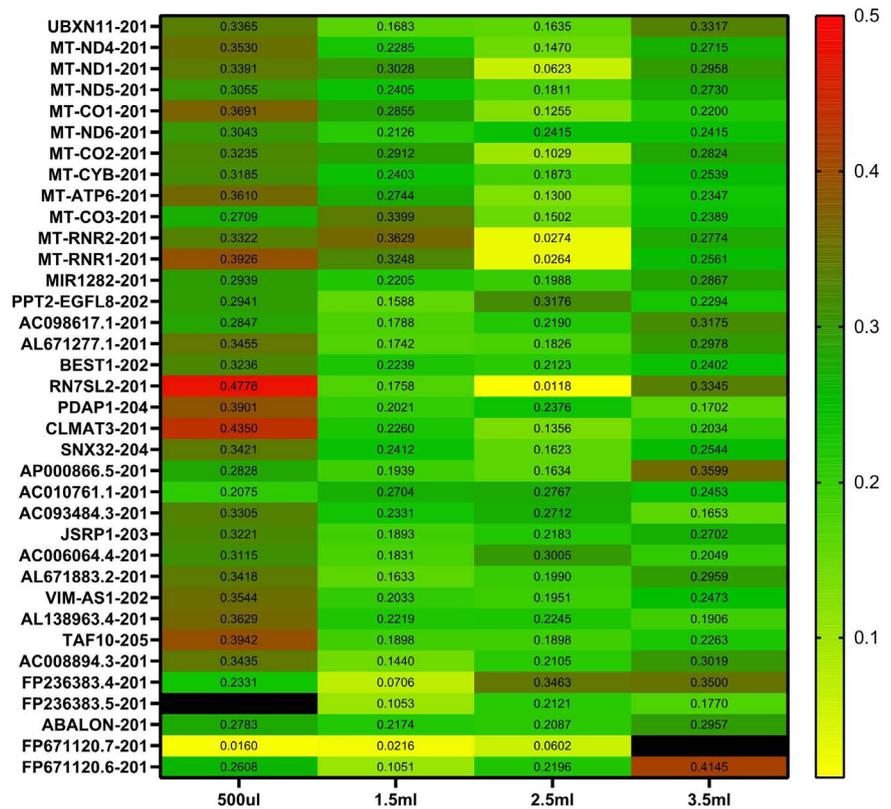


Figure 4. The heat map of distribution of highly expressed transcripts. The green part indicates the transcript is expressed averagely in the 4 groups.

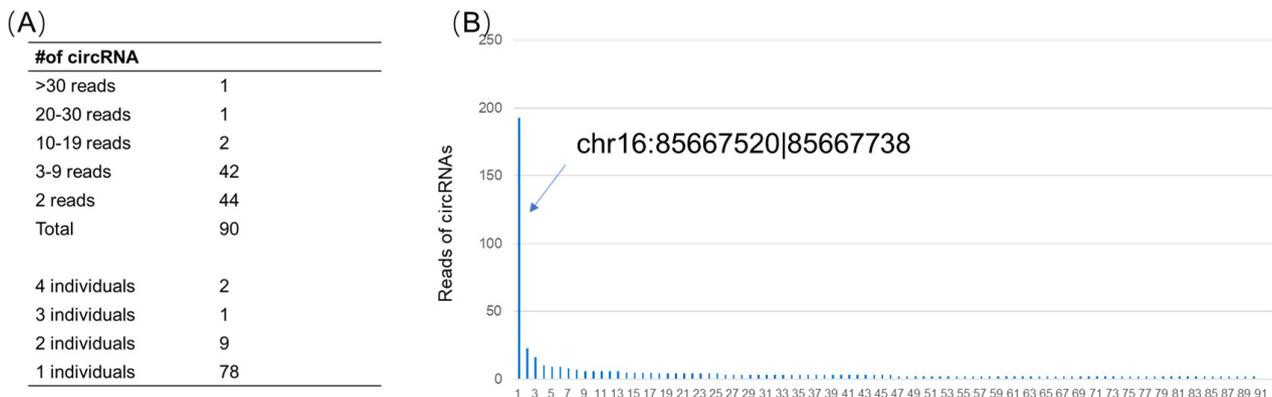


Figure 5. CircRNA in serum. (A) Analysis of the reads of circRNA in serum by sequencing; (B) Distribution of circRNA in serum.

exRNA in serum, the exRNA only 10 pg in 5 - 7 μ L serum, comparable to the amount of RNA in a single cell, so the amount of exRNA is under 1 ng in 500 μ L serum [6]. It is difficult to build the library of exRNA.

In this study, we explored the method of total exRNA sequencing in low volume serum and we discuss the types and distribution of exRNA in serum, as well as the effect of different serum volumes on the results of exRNA sequencing. About 20 million clean reads were acquired by RNA-seq technology per sample and 50% - 80% were mapped to the hg38. In short, we acquired 4000 -

6000 transcripts and about hundreds of thousands counts in each individual sample from serum RNA-seq.

Besides, according to the sequencing results, we found that most of the exRNA in the serum is lncRNA, protein_coding and MT-rRNA, of which lncRNA and protein_coding account for a relatively high proportion in terms of expression level and type, while MT-rRNA only has 2 transcripts, but the expression is very high. In general, when the volume of serum is between 0.5 - 3.5 ml, the exRNA sequencing results, including the amount of circRNA, have not changed significantly, maybe the key is the depth of sequencing, and it requires further experimentation.

It cannot be denied that there still have limits in our study. The main weakness is the small number of samples. However, serum exRNA sequencing technology is still an important technology for us to explore disease biomarkers. The experiment learned that at the same depth of sequencing, the initial volume of serum does not affect the sequencing results, and the volume of serum can be reduced in subsequent experiments.

Acknowledgements

Thanks to the grant from the National Natural Science Foundation of China [No.81827901].

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Heitzer, E., Haque, I.S., Roberts, C.E.S. and Speicher, M.R. (2019) Current and Future Perspectives of Liquid Biopsies in Genomics-Driven Oncology. *Nature Reviews Genetics*, **20**, 71-88. <https://doi.org/10.1038/s41576-018-0071-5>
- [2] Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. and Vogelstein, B. (2003) Transforming Single DNA Molecules into Fluorescent Magnetic Particles for Detection and Enumeration of Genetic Variations. *Proc Natl Acad Sci USA*, **100**, 8817-8822. <https://doi.org/10.1073/pnas.1133470100>
- [3] Leary, R.J., Sausen, M., Kinde, I., Papadopoulos, N., Carpten, J.D., Craig, D., *et al.* (2012) Detection of Chromosomal Alterations in the Circulation of Cancer Patients with Whole-Genome Sequencing. *Science Translational Medicine*, **4**, 162ra154. <https://doi.org/10.1126/scitranslmed.3004742>
- [4] Mandel, P. and Metais, P. (1948) Les acidesnucléiques du plasma sanguin chez l'Homme. *CR Acad Sci Paris*, **142**, 241-243.
- [5] Pan, W.Y., Ngo, T.T.M., Camunas-Soler, J., Song, C.X., Kowarsky, M., Blumenfeld, Y.J., *et al.* (2017) Simultaneously Monitoring Immune Response and Microbial Infections during Pregnancy through Plasma cfRNA Sequencing. *Clinical Chemistry*, **63**, 1695-1704. <https://doi.org/10.1373/clinchem.2017.273888>
- [6] Koh, W., Pan, W.Y., Gawad, C., Fan, H.C., Kerchner, G.A., Wyss-Coray, K., *et al.* (2014) Noninvasive *In Vivo* Monitoring of Tissue-Specific Global Gene Expression

- in Humans. *Proceedings of the National Academy of Sciences*, **111**, 7361-7366. <https://doi.org/10.1073/pnas.1405528111>
- [7] Bianchi, D.W., Mieghem, T.V., Shaffer, L.G., Faas, B.H.W., Chitty, L.S., Ghidini, A. and Deprest, J. (2014) In Case You Missed It: The Prenatal Diagnosis Section Editors Bring You the Most Significant Advances of 2013. *Prenatal Diagnosis*, **34**, 1-5. <https://doi.org/10.1002/pd.5007>
- [8] Zhou, Z.X., Wu, Q.Y., Yan, Z.M., Zheng, H.Z., Chen, C.J., Liu, Y., *et al.* (2019) Extracellular RNA in a Single Droplet of Human Serum Reflects Physiologic and Disease States. *Proc Natl Acad Sci USA*, **116**, 19200-19208. <https://doi.org/10.1073/pnas.1908252116>
- [9] Li, Y.C., Zhao, J.J., Yu, S.L., Wang, Z., He, X.G., Su, Y.H., *et al.* (2019) Extracellular Vesicles Long RNA Sequencing Reveals Abundant mRNA, circRNA, and lncRNA in Human Blood as Potential Biomarkers for Cancer Diagnosis. *Clinical Chemistry*, **65**, 798-808. <https://doi.org/10.1373/clinchem.2018.301291>
- [10] Amorim, M.G., Valieris, R., Drummond, R.D., Pizzi, M.P., Freitas, V.M., Nunes, D.N., *et al.* (2017) A Total Transcriptome Profiling Method for Plasma-Derived Extracellular Vesicles: Applications for Liquid Biopsies. *Scientific Reports*, **7**, 14395. <https://doi.org/10.1038/s41598-017-14264-5>
- [11] Dluzen, D.F., Hooten, N.N., De, S., Wood, W.H., Zhang, Y.Q., Evans, M.K., *et al.* (2018) Extracellular RNA Profiles with Human Age. *Ageing Cell*, **17**, e12785. <https://doi.org/10.1111/accel.12785>
- [12] Hu, X.D., Bao, J.T., Wang, Z., Zhang, Z.G., Gu, P.J., Tao, F., Cui, D. and Jiang, W.L. (2016) The Plasma lncRNA Acting as Fingerprint in Non-Small-Cell Lung Cancer. *Tumor Biology*, **37**, 3497-3504. <https://doi.org/10.1007/s13277-015-4023-9>
- [13] Hooten, N.N., Fitzpatrick, M., Wood, W.H., De, S., Ejiogu, N., Evans, M.K., *et al.* (2013) Age-Related Changes in MicroRNA Levels in Serum. *Ageing*, **5**, 725-740. <https://doi.org/10.18632/aging.100603>
- [14] Liao, Y., Smyth, G.K. and Shi, W. (2014) Feature Counts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics*, **30**, 923-930. <https://doi.org/10.1093/bioinformatics/btt656>
- [15] Gao, Y., Wang, J.F. and Zhao, F.Q. (2015) CIRI: An Efficient and Unbiased Algorithm for De Novo Circular RNA Identification. *Genome Biology*, **16**, 4. <https://doi.org/10.1186/s13059-014-0571-3>