Published online 11 February 2016

Nucleic Acids Research, 2016, Vol. 44, No. 9 ***e87***
doi: 10.1093/nar/gkw075

Circular RNA profile in gliomas revealed by identification tool UROBORUS

Xiaofeng Song[1], Naibo Zhang[2], Ping Han[3], Byoung-San Moon[2], Rose K. Lai[4,*], Kai Wang[5,*] and Wange Lu[2,*]

[1]Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, [2]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Department of Stem Cell Biology and Regenerative Medicine, University of Southern California, Los Angeles, CA 90089, USA, [3]Department of Gynecology and Obstetrics, The First Affiliated Hospital with Nanjing Medical University, Nanjing 210029, China, [4]USC Norris Comprehensive Cancer Center, Department of Neurology and Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA and [5]Zilkha Neurogenetic Institute, Department of Psychiatry and Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA

Received September 05, 2015; Revised December 15, 2015; Accepted January 31, 2016

ABSTRACT

Recent evidence suggests that many endogenous circular RNAs (circRNAs) may play roles in biological processes. However, the expression patterns and functions of circRNAs in human diseases are not well understood. Computationally identifying circRNAs from total RNA-seq data is a primary step in studying their expression pattern and biological roles. In this work, we have developed a computational pipeline named UROBORUS to detect circRNAs in total RNA-seq data. By applying UROBORUS to RNA-seq data from 46 gliomas and normal brain samples, we detected thousands of circRNAs supported by at least two read counts, followed by successful experimental validation on 24 circRNAs from the randomly selected 27 circRNAs. UROBORUS is an efficient tool that can detect circRNAs with low expression levels in total RNA-seq without RNase R treatment. The circRNAs expression profiling revealed more than 476 circular RNAs differentially expressed in control brain tissues and gliomas. Together with parental gene expression, we found that circRNA and its parental gene have diversified expression patterns in gliomas and control brain tissues. This study establishes an efficient and sensitive approach for predicting circRNAs using total RNA-seq data. The UROBORUS pipeline can be accessed freely for non-commercial purposes at http://uroborus.openbioinformatics.org/.

INTRODUCTION

In the diverse world of RNA, most are non-coding RNA, including lincRNA, rRNA, tRNA, snRNA and microRNA (miRNA), which play important regulatory roles in numerous biological processes. These non-coding RNAs are known to be linear molecules with 5′ and 3′ termini. However, some non-coding RNA also exists in a covalently closed continuous loop, called circular RNA (circRNA). In the decades since the initial discovery of circRNA in a viroid in 1976 (1), some endogenous circRNAs from back-spliced exons have been found in mammals, including *Mus musculus* (2) and *Homo sapiens* (3–9).

With the development of next generation sequencing, abundant circRNAs have been identified in human cell lines based on poly(A)-minus or RNase R treated RNA-seq data (10–12). Sorek *et al*. first used RNase R treated RNA-seq data to reveal genome-wide circRNAs in *Archaea* (10). By using a non-conical exon ordering strategy based on MapSplice, Sharpless and colleagues performed high-throughput sequencing on RNase R treated RNA (rRNA-depleted) from human fibroblasts, and identified more than 25 000 circRNAs (11,13). After collecting rRNA-depleted poly(A)-minus RNA-seq in multiple human cell lines downloaded from the ENCODE project and aligning them with the customized scrambled exon database, Brown *et al*. identified widespread cell-type specific circRNA expression in humans (12,14). However, the above three research groups did not publish their tools. Stadler *et al*. proposed an unbiased circRNA detection algorithm integrated into the mapping tool segemehl, which can detect splicing, trans-splicing and gene fusion events from single-end read data based on an enhanced suffix array, chaining and dynamic programming algorithms (15). Zhao *et al*. proposed a *de novo* cir-

*To whom correspondence should be addressed. Tel: +1 323 442 1618; Fax: +1 323 442 4040; Email: wangelu@usc.edu
Correspondence may also be addressed to Rose K. Lai. Tel: +1 323 442 1618; Fax: +1 323 442 4040; Email: roselai@usc.edu
Correspondence may also be addressed to Kai Wang. Tel: +1 323 442 1618; Fax: +1 323 442 4040; Email: kaiwang@usc.edu

cRNA identification tool named CIRI using multiple filtration strategies (16). Rajewsky *et al*. developed a pipeline to identify thousands of circRNAs in rRNA-depleted total RNA sequencing data from multiple human cell lines such as CD19+ and HEK293, and they also revealed that circRNA functions as sponge of miRNA (17–19). One year after publishing these results, they released their linux script pipeline named find_circ. Shortly afterward, based on Tophat_fusion, Yang *et al*. developed a circRNA detection pipeline named *CIRCexplorer*, and revealed that circRNA formation is mediated by complementary sequences flanking the exons (20,21). Due to the detection strategy bias used in Tophat_fusion, which is specifically designed for identifying fusion genes, the same bias exists in *CIRCexplorer*. In addition, all of the above circRNA detection pipelines have been primarily applied to certain RNA-seq data sets, such as poly(A) minus RNA-seq data or RNase R treated total RNA-seq data. Those poly(A) minus RNA-seq data exclude coding transcripts, and can be used for detecting circRNA (without a polyA tail) with high sensitivity. After rRNA depleted total RNA is treated with RNase R, which can digest the linear mRNA transcripts, an abundance of circRNA can be enriched and easily detected by high-sensitivity tools. However, there is a large amount of total RNA-seq data without poly(A) depletion or RNase R treatment readily at hand for scientists. These data can be used for many research purposes, including non-coding RNA and mRNA differential expression analysis, with the goal of saving sequencing costs. Therefore, there is an urgent and widespread need for a computational tool for the unbiased identification of circRNAs from total RNA-seq data (rRNA depleted).

In this project, we developed a tool named UROBORUS, a Greek word referring to the ancient symbol of the snake or dragon eating its own tail. UROBORUS is based on TopHat and Bowtie tools using PERL language, which can identify circRNA from total RNA-seq data (rRNA depleted) (22,23). We implemented UROBORUS on RNA-seq data from glioma samples to reveal the circRNA profiling of gliomas, which represent 30% of all brain and central nervous system tumors and 80% of all malignant brain tumors (24). This study included two types of glioma with distinct molecular pathogeneses: *de novo* GBM (WHO glioma grade IV) and oligodendroglioma (glioma WHO grade II and III). Therefore, we sequenced the transcriptomics of 7 oligodendroglioma, 20 glioblastoma (GBM) and 19 normal tissue samples, and tried to reveal the circular RNA profile in gliomas using UROBORUS.

MATERIALS AND METHODS

Patient sample collection and RNA isolation

We retrieved 19 fresh-frozen *de novo* GBM and 7 oligodendroglioma (3 WHO grade III and 4 WHO grade II) from the University of Southern California Brain Tumor Bank. All tumor tissues passed the neuropathological criteria of having >80% tumor nuclei and <50% necrosis (GBM). For controls, we obtained 20 fresh-frozen autopsy brain tissues from the University of Miami Brain Bank. These non-tumor controls were neuropathologically verified to be free of any neurological diseases. Tumors and control subject characteristics were summarized in Supplementary Table S1.

Total RNA was extracted using RNeasy Kit (Qiagen) with DNase I digestion according to the manufacturer's instructions. RNA integrity was verified on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). cDNA was synthesized from 2 μg of total RNA using Superscript III (Life Technologies Inc.) and random primers (Life Technologies Inc.).

Library preparation and sequencing

Ribosomal depleted libraries were constructed for all samples using Illumina's TruSeq Total RNA Library Prep Kit. End-repair and adaptor ligation were performed according to manufacturer's instructions. Libraries were size-selected for 250–300 bp cDNA fragments on a 3.5% agarose gel and PCR-amplified for 15–20 cycles. PCR products were then purified on a 2% agarose gel and gel-extracted. Library quality was credentialed by assaying each library on an Agilent 2100 Bioanalyzer of product size and concentration. RNA seq was then generated using Illumina Hi-seq 2000, with 50 paired-end (PE) reads. Each sample yielded at least 60 million reads.

circRNA detection pipeline

CircRNA can be produced from either an exon or an intron locus. However, most human circRNAs are exonic circRNAs, which is derived from a canonical splice donor site. We developed a computational pipeline named UROBORUS combined with TopHat and Bowtie to detect junction reads from these back-spliced exons in order to reveal the circRNA profile in gliomas. In order to filter those canonical splicing supporting reads, the RNA-seq reads were initially mapped to the human reference genome (hg19) using TopHat, which is capable of detecting the canonical splicing event. However, TopHat cannot map junction reads supporting back-spliced exons to the reference genome, and circRNA should be detected in those unmapped reads. Therefore, UROBORUS used the unmapped.sam file from the TopHat results as input data.

The main steps of UROBORUS are briefly described in Figure 1. The first step is to extract 20 bp from two ends (head and tail) of reads in an unmapped.sam file to form an artificial paired-end seed in fastq file format (R20_1.fastq and R20_2.fastq). Then, we aligned this short 20 bp paired-end seed to the human reference genome (hg19) with maximum of 2 bp mismatches using TopHat with default parameters.

In the above mapping results, we should identify two cases of junction reads from the mapping results in the SAM file format: balanced mapped junction (BMJ) reads and unbalanced mapped junction (UMJ) reads. BMJ reads are represented as reads aligned to the joining region of two back-spliced exons with minimum 20 bp of overhang at any end of the reads; UMJ reads are represented as reads aligned to the joining region of two back-spliced exons with less than 20 bp of overhang at one end of the reads. Intuitively, this is a balanced or unbalanced mapping event, so we call it a BMJ or UMJ read.

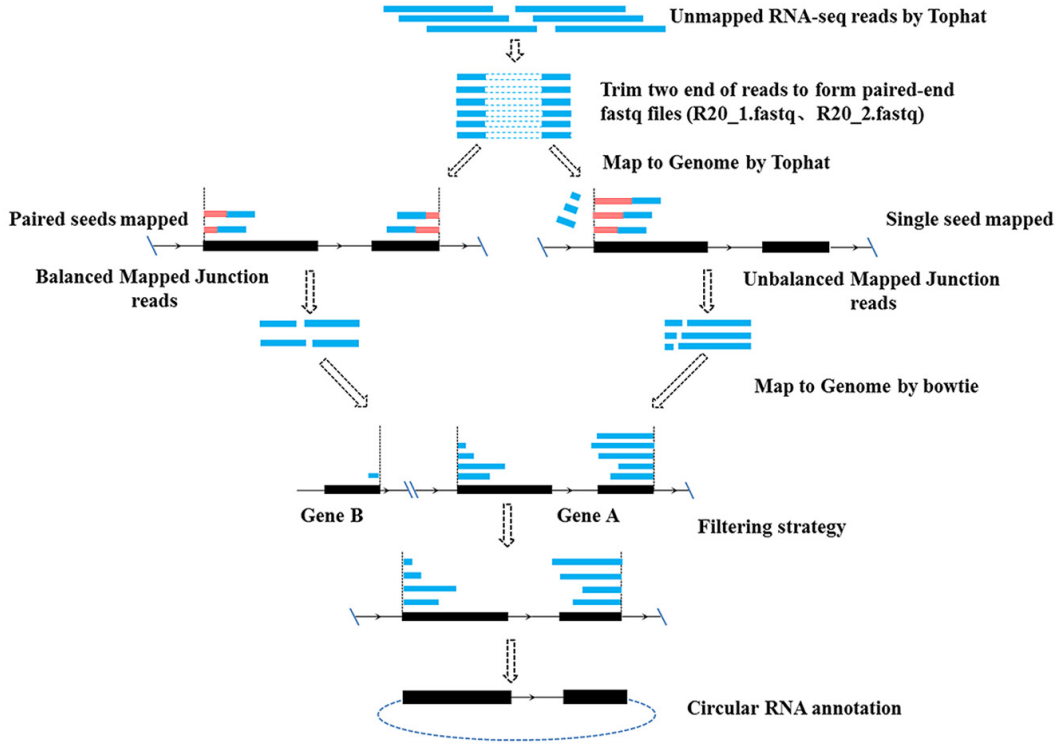


Figure 1. The UROBORUS pipeline for identifying circRNA based on total RNA-seq. The artificial paired-end seed (20 bp) was first extracted from two ends of reads in an unmapped.sam file, and then aligned to the reference genome. The results would have two cases of reads spanning the spliced site: balanced mapped junction (BMJ) reads and unbalanced mapped junction (UMJ) reads. The UROBORUS pipeline designed algorithm to deal with BMJ and UMJ reads, and detects more circRNA supported reads.

Most circRNAs are involved in non-linear exon ordering in the same chromosome and same gene in the second mapped SAM format file (converted from .bam file). It should be noted that some of these short 20 bp paired seeds have multiple mapping locations in the genome. First, we collected a short 20 bp seed and its mate in the primary alignment in the opposite orientation within 3 million bp distance (estimated longest gene length) along the same chromosome. Thus, some spurious alignments not supporting circRNA would be filtered. For instance, the head seed should be aligned at an exon downstream of the tail seed mapped exon for those plus strand genes. The head seed should be aligned at an exon region upstream of the tail seed mapped exon for those minus strand genes. Next these mapped paired-end seeds were separately extended outward to a splice site to form paired-end segments. If these paired-end segments could be mapped to the genome in the next step, they would join together to produce the candidate BMJ read supporting circRNA.

For detecting UMJ reads, the single mapped seed was selected, because its mate was too short to be mapped to the genome. Next, the seed was extended and mapped to the genome within the distance to the splice sites (acceptor site or donor site). We filtered out those single mapped seeds with a distance to the spliced site of larger than the read length minus 3 bps. Then the extended mapped seed (less than read length minus 3 bps) was kept as one segment of UMJ read, while the rest of sequence was kept as the other segment of UMJ read, which still remained not to do alignment as showed in Figure 1.

To obtain the integrated mapping *.bam file, we collected the above BMJ and UMJ reads to do full read alignments to the human reference genome (hg19) using Bowtie again. The paired-end reads not aligned to the same chromosome or the same gene were filtered by filtering strategy in the UROBORUS. Those BMJ or UMJ paired-end reads that aligned to the same chromosome but in opposite orientations (detected by UROBORUS algorithm) were annotated as candidate back-spliced junction reads.

In the end, those candidate back-spliced junction reads were collected and statistically analyzed. Those supporting junction reads above 2 reads were annotated to be candidate circRNA.

RT-PCR and Sanger sequencing

One microgram of total RNA for each sample was treated by RNase R (Epicentre, San Diego, CA, USA) at 37°C for 3 h. Two 1 μg of Total RNA for each sample were reverse transcribed to cDNA using an iScript RT SuperMix kit (Bio-Rad, Hercules, CA, USA), according to the manufacturer's protocols for subsequent amplification. Divergent and convergent primers were designed by an NCBI primer design tool and synthesized from Integrated DNA Technology (see Supplementary Table S2). We performed 35 cycles of PCR under standard conditions with 30 s 55°C extension. The PCR products were then subjected to elec-

trophoresis in 2% ethidium bromide-stained agarose gel. To conform to the PCR results, the PCR products were purified using a Zymoclean Gel DNA Recovery Kit. The Sanger sequencing for PCR product was performed by GENEWIZ.

RESULTS

UROBORUS was tested on total RNA-seq data (rRNA depleted) from 7 oligodendroglioma, 20 glioblastoma and 19 normal brain tissues. First, we mapped the RNA-seq reads to the human reference genome (hg19) using TopHat. Then we employed UROBORUS to detect circRNA in the unmapped reads, and identified thousands of novel circRNAs in both the normal and the cancer tissues (see Table 1). Some known circRNAs (circular HIPK3, CAMSAP1, MAN1A2, FBXW4, REXO4, ZKSCAN1, ZBTB44, FAM120A, MAP3K1, ZBTB46, NUP54, RARS, CRKL and XPO1) reported in previous research were also detected in our data (11,13,20).

One of the biggest computational challenges for finding circRNA using short 20 bp seeds is the multiple locations of seeds in the reference genome. To solve this problem, we developed UROBORUS's filtering algorithm to eliminate those spurious alignments (see Materials and Methods). Thus, we can reduce UROBORUS's false positive rate of circRNA, and obtain more circRNA with strong support from multiple reads.

One of the defects in previous pipelines is that they only use alignments with BMJ reads, in which the mapping region across junction point is more than the seed length, and neglect UMJ reads, in which the overhang region across break point is less than the seed length (usually ∼20 bp). In these cases, many circRNAs supported by short spanning reads are missed or underestimated in their expression level. However, the circular RNAs supported by UMJ reads are not rare events. UROBORUS was designed to address this problem.

Figure 2 illustrates two typical cases of the circRNAs identified by UROBORUS. Figure 2A shows the BMJ and UMJ reads spanning a break point of circular ERC1 between exon14 and exon15. As illustrated in the figure, only 30 BMJ read counts clearly span the break point of circular ERC1, but the 43 UMJ read counts spanning the boundary of circular ERC1 should not be neglected. Figure 2B shows a circular CORO1C not supported by any BMJ reads, and only supported by 7 UMJ read counts, indicating that the circular CORO1C would be missed by those pipelines that only employ BMJ reads. By employing both BMJ and UMJ reads, UROBORUS improves the sensitivity of circRNA detection.

Estimation of the false positive rate

In order to estimate false positive rate, we should run the pipeline on the RNA-seq data from RNA sample that does not contain circRNA. We expect that no circular RNA should be identified. If circRNAs were detected, that would be false positive circRNAs. A similar strategy was widely employed by previous researchers in detecting the fusion gene (25). Therefore, in order to estimate UROBORUS's false positive rate, we ran UROBORUS on poly(A)+ RNA-seq data in which circRNAs should be absent. We tested UROBORUS on ploy(A)+ RNA-seq data, and detected only one circRNA with 2 reads counts supported. Considering that this sample comprised ∼1.27 million reads, and assuming that all reported circRNA are false positives, the false positive rate would be ∼0.79 per million reads. Thus, we can estimated that UROBORUS will have FDR < 0.013 (62.4 million reads, 3875 circRNAs).

Comparison with CIRCexplorer and find_circ

Two available computational pipelines—*find_circ* and *CIRCexplorer*—were selected to compare with UROBORUS. The number of circRNAs detected by three methods was listed in Table 1, showing that UROBORUS can detect more circRNAs than find_circ, but fewer circRNAs than *CIRCexplorer*. Due to the detection strategy bias, *CIRCexplorer* inevitably misses those UMJ reads supporting circRNAs in the first step and also overestimates the first step detected circRNA expression level. In this case it seems that CIRCexplorer detected more circular RNA due to its detection bias. The facts that *find_circ* missed many circRNAs supported by short spanning reads or underestimated in their expression level lead to the less circRNAs detected.

Among all the detected circRNAs, there are about 156 circRNAs expressed at a level above 0.1 RPM (reads per million mapped reads) in all samples identified by *find_circ*, 1417 circRNAs expressed at a level above 0.1 RPM in all samples identified by *CIRCexplorer*, and about 572 circRNAs expressed at a level above 0.1 RPM in all samples identified by UROBORUS. It seems that UROBORUS can identify more circRNAs than *find_circ,* but fewer circRNAs than *CIRCexplorer*. This might be due to the reason that overestimation of expression level by *CIRCexplorer.* When all the circRNAs (supported by 2 reads) identified by the three methods were compared using a Venn Diagram, only 307 common circRNAs were identified by the three pipelines, as shown in Figure 3A. However, *find_circ* identified 196 circRNAs that are different from *CIRCexplorer* and UROBORUS; *CIRCexplorer* identified 524 circRNAs that are different from *find_circ* and UROBORUS; and UROBORUS identified 726 circRNAs that are different from *find_circ* and *CIRCexplorer*. These results indicate that UROBORUS can detect more novel circRNAs, and these three methods can be complementary tools, based on their different mapping and filtering strategies. It is notable that all three pipelines can identify about 90% of those highly expressed circRNAs above 0.3 RPM, and detect more common circRNAs as the expression level of circRNAs increases, as shown in Figure 3B. This indicates that three pipelines are equally useful in predicting highly expressed circRNA.

As shown in Supplementary Table S3, 24 randomly selected circRNAs validated by experiments were all predicted by UROBORUS. However, *find_circ* predict 14 validated circRNAs. *CIRCexplorer* only predict 9 validated circRNAs. This suggests that UROBORUS may have better performance than *find_circ* or *CIRCexplorer* in predicting circRNAs in total RNA-seq data. We successfully validated 16 circRNAs at low expression level among the 19 circRNAs randomly selected from the discrepancy group,

Table 1. Number of circRNAs identified by UROBORUS, find_circ and CIRCexplorer

Cancer tissue*		OLIGO2 (4 subjects)				OLIGO3 (3 subjects)		
ID		RL8	RL9	RL11	RL12	RL10	RL13	RL14
Number of fragments(10^6)		89.5	57.6	63.9	65.9	56.9	57.7	63.7
Number of circRNAs	UROBORUS	**3074**	**1232**	**2190**	**2311**	**1294**	**1025**	**1530**
	Find_circ	2484	907	1601	1766	772	753	1097
	CIRCexplorer	3565	1372	2434	2592	1233	1076	1626

Cancer tissue*		GBM (20 subjects)																		
ID		RL1	RL2	RL3	RL4	RL5	RL6	RL7	RL27	RL28	RL29	RL30	RL31	RL32	RL33	RL34	RL35	RL36	RL37	RL38	RL39
Number of fragments(10^6)		62.5	59.3	59.6	60.5	62.2	55.6	65.3	57.4	65.4	69.3	67.3	63.9	68.9	66.7	60.0	65.6	59.7	74.3	60.9	66.1
Number of circRNAs	UROBORUS	**1832**	**934**	**70**	**76**	**115**	**764**	**814**	**527**	**940**	**975**	**786**	**501**	**541**	**1899**	**463**	**569**	**383**	**904**	**853**	**1194**
	Find_circ	1258	634	38	89	79	666	593	349	670	753	598	392	352	887	284	385	248	664	560	885
	CIRCexplorer	1876	973	17	107	71	862	852	512	986	1045	824	532	597	2138	426	584	293	997	880	1235

Normal Tissue*		CEREBELLUM (6 subjects)						CORTEX (13 subjects)												
ID		RL15	RL16	RL17	RL18	RL19	RL20	RL21	RL22	RL23	RL24	RL25	RL26	RL40	RL41	RL42	RL43	RL44	RL45	RL46
Number of fragments(10^6)		62.4	91.5	59.8	65.4	59.7	58.4	64.7	66.9	64.3	58.6	64.2	59.0	68.1	67.7	66.0	63.1	67.2	57.5	69.6
Number of circRNAs	UROBORUS	**3875**	**4965**	**2116**	**3079**	**2606**	**2407**	**3292**	**2425**	**2176**	**1689**	**2159**	**1998**	**2005**	**1213**	**2003**	**1703**	**2288**	**1479**	**2257**
	Find_circ	2530	4314	2148	2644	2184	2076	2705	1997	1744	1665	1664	1580	1703	961	1611	1442	1842	1083	1785
	CIRCexplorer	4613	5849	2942	3744	3064	2878	3857	2794	2450	2278	2455	2350	2413	1332	2369	2037	2671	1654	2534

*GBM = glioblastoma multiforme; OLIGO2 = oligodendroglioma WHO grade II; OLIGO3 = oligodendroglioma WHO grade III; cortex = cortical brain controls; cerebellum: cerebellar controls.

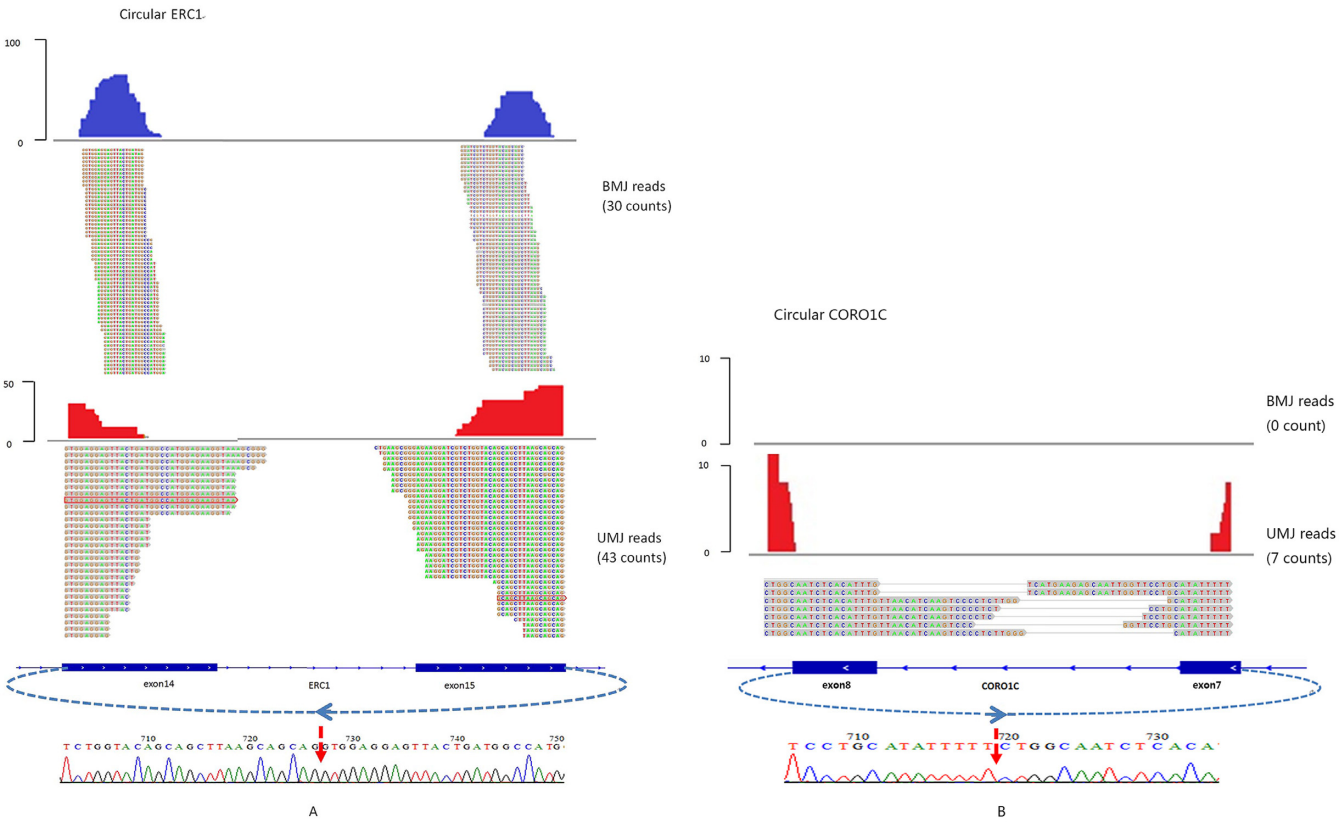
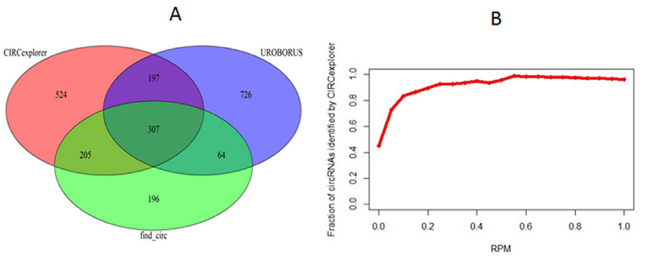


Figure 2. CircRNAs supported by BMJ and UMJ reads. UMJ reads are neglected by other approaches, leading to underestimate of the circRNA expression level or complete missing of those only UMJ supported circRNAs. (**A**) UROBORUS can identify both BMJ and UMJ reads supported circular ERC1, while find_circ and CIRCexplorer can only detect fraction of BMJ reads of circular ERC1, leading to expression level underestimation of circular ERC1. Sanger sequencing confirmed ERC1 circRNA. (**B**) Unlike find_circ and CIRCexplorer, one of the most powerful features in UROBORUS is its ability to identify those only UMJ reads supported circular RNA such as CORC1C, which were further confirmed by Sanger sequencing.



	Language	Depended tools	Memory	Time
UROBORUS	perl	Tophat、bowtie	6.3 G	4.8 h
CIRCexplorer	python	Tophat_fusion	5.2 G	2.3 h
Find-circ	python	bowtie	5.8 G	2.8 h

Figure 3. Comparison of three pipelines. (**A**) CircRNAs (supported by 2 reads) identified by the three methods were compared using a Venn Diagram; (**B**) UROBORUS and CIRCexplorer identify more common circRNAs as shown by the factions when the expression level of circRNAs increases; (**C**) The comparison of the three methods in language used, depending tools, memory and running time.

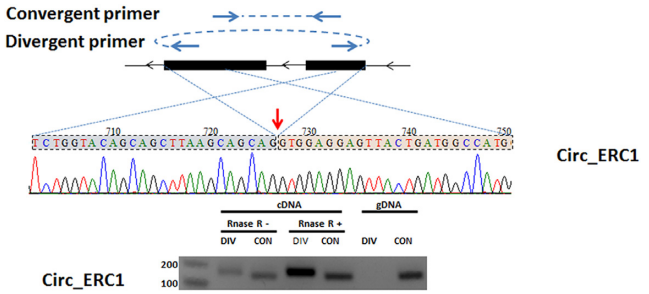


Figure 4. Primer design, PCR and Sanger sequencing validation for circ_ERC1. Two pairs of primers (convergent or divergent primers) were designed for each candidate circRNA (primer information in Supplementary Table S2). The junction region of circular RNA sequenced by Sanger sequencing validated the circular ERC1 formation; Gel electrophoresis validated the existence of circular ERC1.

in which UROBORUS identified, while *find_circ* or *CIRCexplorer* did not hit. This result indicated that UROBORUS can detect more novel lower expressed circRNAs and has lower false positive rate than other two methods.

As shown in Figure 3C, UROBORUS has maximum memory of 6.3 G. This is slightly more than the other two pipelines, which use maximum memory of 5.2 G and 5.8 G, suggesting that these three pipelines have similar space complexity. UROBORUS can detect circRNAs on 10 G RNA-seq data with a running time of 4.8 h, while the other two pipelines run about 2.3 or 2.8 h on the same data in the same biocluster (compute node: Intel Xeon CPU X5650@2.67GHz, 12 core, 80 G memory).

Experimental validation

To verify circRNAs identified by UROBORUS, we randomly selected 27 circRNAs from normal brain samples, GBM and oligodendrogliomas for implementing RT-PCR and Sanger sequencing for validation. The expression level of the selected 27 circRNAs ranges from 0.017 to 3.104 RPM. Two pairs of primers (convergent or divergent primers) were designed for each candidate circRNA. Convergent primers are expected to amplify both the linear and circular transcripts, and divergent primers are expected to amplify only circular transcripts (Figure 4). We performed RT-PCR separately with total RNA or RNase R-treated RNA from normal, oligodendroglioma or GBM tissue, and a genomic DNA sample was used as a control. RNase R can digest linear RNA and enrich circRNA, so we can avoid detecting the false positive circRNA derived from genomic re-arrangement using an RNase R treated RNA sample. GAPDH was employed as a control for linear mRNA digestion in RNase R treatment.

For the 27 randomly selected circRNAs, we successfully amplified 24 circRNAs from total RNA samples or RNase R treated RNA samples, in which circular CORO1C (0.061 RPM, supported by 7 read counts), expressed at lower levels, was also validated, but not amplified any circRNAs in genomic DNA samples. The experimental results were shown in Supplementary Figure S1. Only 3 circRNAs (circ_RPPH1, circ_MORC3 and circ_TTC3) were not validated from total RNA samples or RNase R treated RNA samples by RT-PCR. The Sanger sequencing on the RT-PCR product amplified by divergent primers further confirmed 24 circRNAs existence.

The findings—that *CIRCexplorer* did not identify 15 circRNAs, and *find_circ* did not identify 10 circRNA among 24 validated circRNAs by experiment in normal, GBM and oligodendroglioma tissue—imply that UROBORUS performs better than the *CIRCexplorer* and *find_circ* in predicting circRNA in total RNA-seq data.

Genomic features of circRNAs in gliomas

Sequencing depths of glioma and control brain samples ranged from 55.6 to 91.5 million reads at the data size of about 10 G (details provided in Supplementary Table S4). About 85% of these RNA-seq reads were mapped to the human reference genome (hg19 version). When applying UROBORUS on the remaining 15% unmapped RNA-seq reads, we identified thousands of circRNA with at least two back-spliced junction reads in each sample of glioma and normal tissues.

Genomic features of highly expressed circRNAs (RPM > 0.1) were investigated and compared between normal and glioma tissue. First, we found no significant difference in the exon number of highly expressed circRNAs between normal and glioma tissues, as shown in Figure 5A, indicating that circRNA is formed in a regular way, and may not be affected by biological conditions. We also found that the genomic distance of the back-splicing site in most circRNA is within 50 kb both in normal and glioma tissue, with only a few circRNAs spanning 100–300 kb, as shown in Figure 5B, suggesting that circRNA formation is most likely within one gene region, may be coupled to RNA splicing and cannot be altered by the biogenesis of gliomas. These two findings indicate that the circRNA is formed in a specific pathway, rather than as random byproduct of a canonical splicing event or as transcriptomic noise.

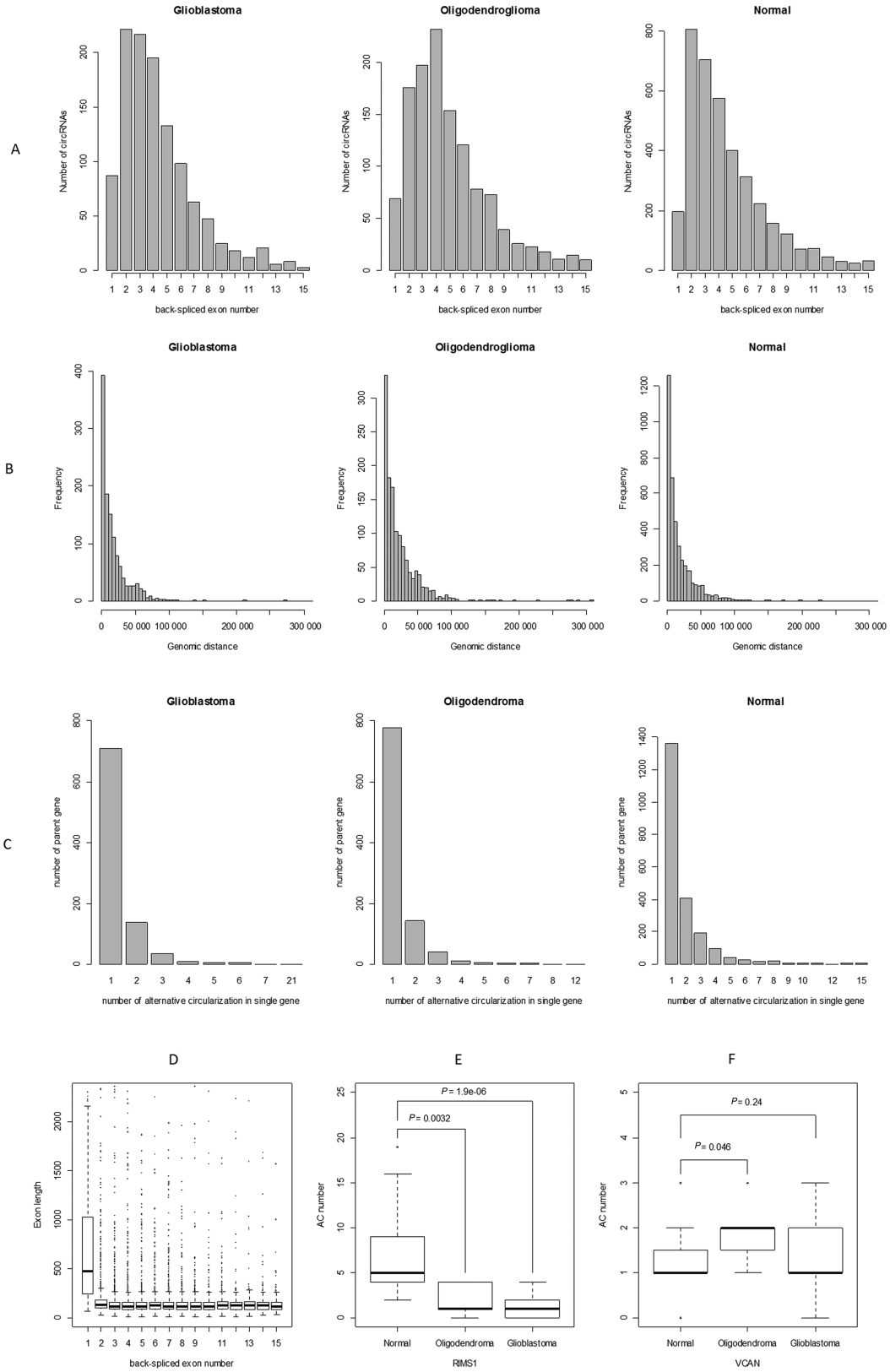


Figure 5. Genomic features of circRNA in gliomas and normal tissue. (**A**) There is no significant difference in the exon number of highly expressed circRNAs between normal and glioma tissues; (**B**) The genomic distance of the back-splicing site in most circRNAs is within 50 kb both in normal and glioma tissue, with only a few circRNAs spanning 100–300 kb; (**C**) The number of alternative circularization numbers in normal tissue is significantly higher than in glioma tissue; (**D**) Exon length of circRNA in all samples; (**E**) RIMS circRNA has more alternative circularization in normal tissue than in glioma tissue; (**F**) VCAN circRNA isoforms are similar in normal and glioma tissue.

Next in all samples including normal and glioma tissue, about 10% of circRNAs contain only one exon, as shown in Figure 5D. Most circRNAs, either in normal or glioma tissue, contain two to seven exons. In circRNAs with only one exon, the exon length was much longer than those from circRNAs with multiple exons. This is the case in both normal and glioma tissue, suggesting that back-splicing processing may prefer a certain length to maximize exon(s) circularization, and not be related to the biogenesis of gliomas. These findings are consistent with previous results from other researchers (12,20). The only difference is that most circRNAs in our samples contain more exons than those circRNAs detected in previous research (10,20), suggesting that different circRNAs may exist in different species or tissues. These similar genomic features were obtained from highly expressed circRNAs identified by our pipeline UROBORUS and others (*find_circ*, *CIRCexplorer*).

In addition, we discovered that multiple circRNA isoforms can be produced from single parental gene, consistent with previous reports (12,20). However, in this study, most of the highly-expressed circRNAs only have one isoform in normal and glioma tissue, with only a small fraction of circRNAs generated in multiple isoforms. Furthermore, the majority of these multiple isoforms were expressed at low levels, with only one or two isoforms expressed at a higher level. These findings suggest that each parental gene can produce one or two primary highly expressed circRNA after competing with other circRNA isoforms.

Statistical analysis was also performed for the alternative circularization numbers (AC numbers) of each parental gene. We found that some parental genes have different AC numbers in normal and glioma tissue, as shown in Figure 5C, indicating that alternative RNA circularization may be involved in gliogenesis. For example, the RIMS gene produces about 5 circRNA isoforms in normal tissue samples, but only 2 circRNA isoforms in oligodendroglioma or GBM samples, as shown in Figure 5E, suggesting that tumor biogenesis reduced RIMS gene alternative circularization. However, the same number of circRNA isoforms derived from the VCAN gene in both normal and glioma tissue, as shown in Figure 5F, indicating that VCAN gene alternative circularization may not be involved in tumor biogenesis.

CircRNA expression profile in gliomas

Thousands of detected circRNAs are distributed in all chromosomes in normal and gliomas tissue, except for the X and Y chromosomes, as illustrated in Figure 6. This distribution is not uniform in the chromosomes, either in normal tissue or gliomas. We also found that the detected circRNA number in GBM was significantly lower compared to brain control tissues or oligodendroblastoma (Wilcoxon rank-sum test, P-value = 1.944e-09, 0.0002117 separately), and there is no significant difference in the circRNA number between normal tissue and oligodendroblastoma (Wilcoxon rank-sum test, P-value = 0.1516).

Approximately 83%(476) of total circRNAs is differentially expressed in normal and GBM tissue with q value < 0.05 (Wilcoxon rank-sum test and Benjamini–Hochberg correction) among 572 highly expressed circRNAs (RPM > 0.1) identified in all 46 samples (details in Supplementary Table S5). There are about 468 circRNAs, such as circ_CDR1, circ_ATRNL1, circ_AKT3, circ_SPTAN1, circ_ZNF483, circ_RIMS1, circ_FKBP8, circ_UNC13C, that are more highly expressed in normal tissue than in GBM with a q value < 0.05 (Wilcoxon rank-sum test and Benjamini–Hochberg correction), indicating that most of the highly expressed circRNAs are important factor in maintaining normal physiological function. Anna *et al*. and colleagues also found that most circRNAs are highly expressed in human normal tissue comparing with colorectal and ovarian cancer, idiopathic lung fibrosis (26).

There are eight circRNAs, such as circ_COL1A2, circ_PTN, circ_VCAN, circ_SMO, circ_PLOD2, circ_GLIS3, circ_EPHB4, circ_CLIP2, that are more highly expressed in GBM than in normal tissues with a q value < 0.05 (Wilcoxon rank-sum test and Benjamini–Hochberg correction), indicating that these eight circRNAs could be potential biomarkers for GBM. The heatmap of highly expressed circRNAs (in RPM) and their parental mRNA expression level (Fragments per kilo base of exon per million mapped reads (FPKM)) is displayed in Figure 7, showing that the circRNAs were differentially expressed among oligodendroglioma, GBM and control brain tissues, and even between different normal tissues (cerebellum and cortex). Moreover, Figure 7 also illustrated that the corresponding parental mRNAs of these circRNAs were also differentially expressed among oligodendroglioma, GBM and control brain tissues, and even between different normal tissues (cerebellum and cortex). These results indicate that both of circRNA and its parental gene are expressed in tissue or cancer specificity.

Previous researchers found that the expression of circular RNA is positively correlated with that of linear parental mRNA (20), and even promote or regulate transcription of their parental genes (27,28). In order to further investigate the correlation of circRNAs expressed with their parental gene, we have calculated the Pearson correlation coefficient (PCC) between the expression levels of circRNAs and their parental mRNAs. The results indicated that about 50% of highly expressed circRNA (>0.1 RPM) have a strong positively correlation (PCC > 0.5) with their parental mRNAs in normal and gliomas. Therefore, some circRNAs in the remaining half of 572 highly expressed circRNAs may have potential effects on gliomas. We also found from Figure 8A that about 185 highly expressed circRNAs and their parental mRNAs has the similar correlation pattern in normal or gliomas (Root mean square error (RMSE) < 0.15). The correlation of the remaining about 387 highly expressed circRNAs with their parental mRNAs have different pattern in normal and gliomas. It is found on the PCC Frequency distribution in Figure 8B that there are more circRNAs strongly correlated (PCC > 0.7) with their parental gene in normal than in gliomas. In addition to differential expression, we also studied the patterns of circRNA expression and their parental gene mRNA expression (Supplementary Table S6). We compared differential expression patterns between GBM and controls, and between oligodendroglioma and controls. For both GBM and oligodendroglioma, there were 3 patterns of circRNA differential expressions: (i) circRNAs more highly expressed in normal tis-

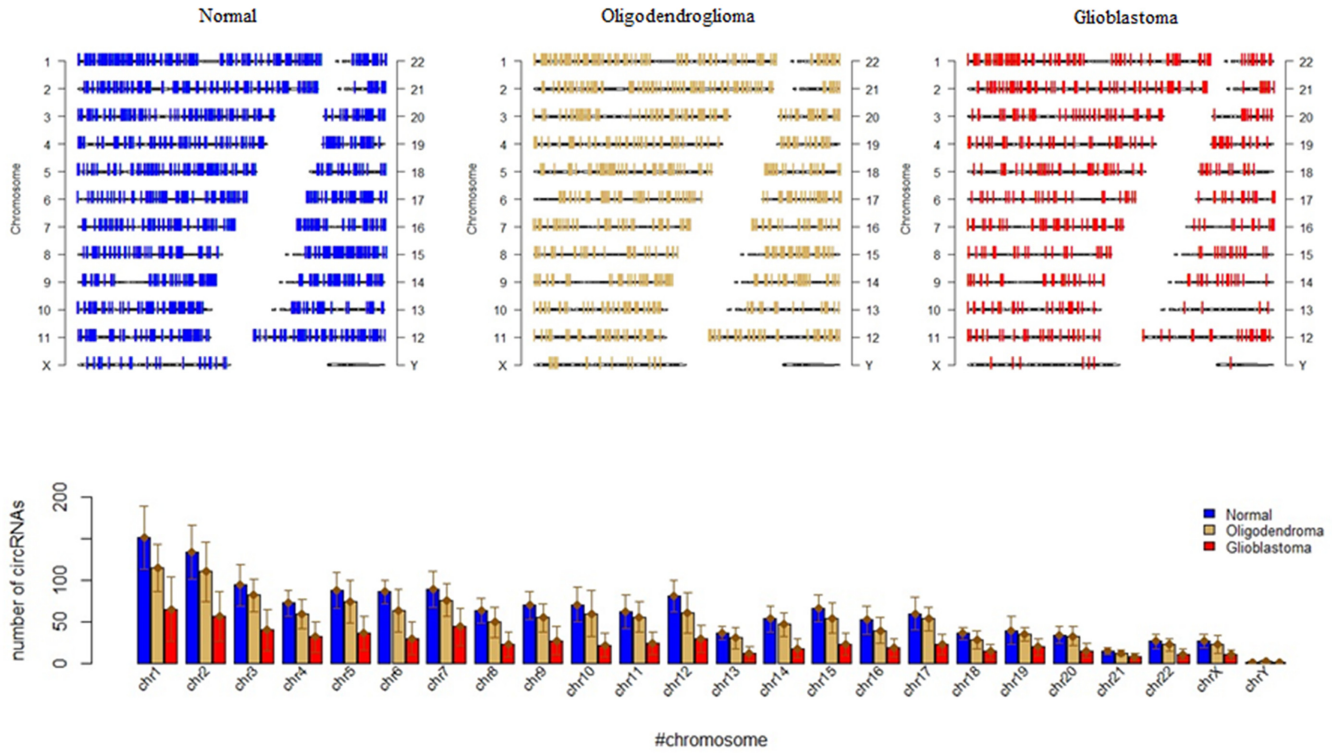


Figure 6. CircRNA distribution in normal tissue and gliomas. The upper panel shows the circRNA distribution in different chromosomes in normal tissue and gliomas. The circRNA number in GBM was significantly lower compared to normal tissues or oligodendroblastoma (Wilcoxon rank-sum test, *P*-value = 1.944e-09, 0.0002117 separately); no significant difference in the circRNA number between normal tissue and oligodendroblastoma (Wilcoxon rank-sum test, *P*-value = 0.1516). The lower panel shows the circRNA number in each chromosomes in normal tissue and gliomas.

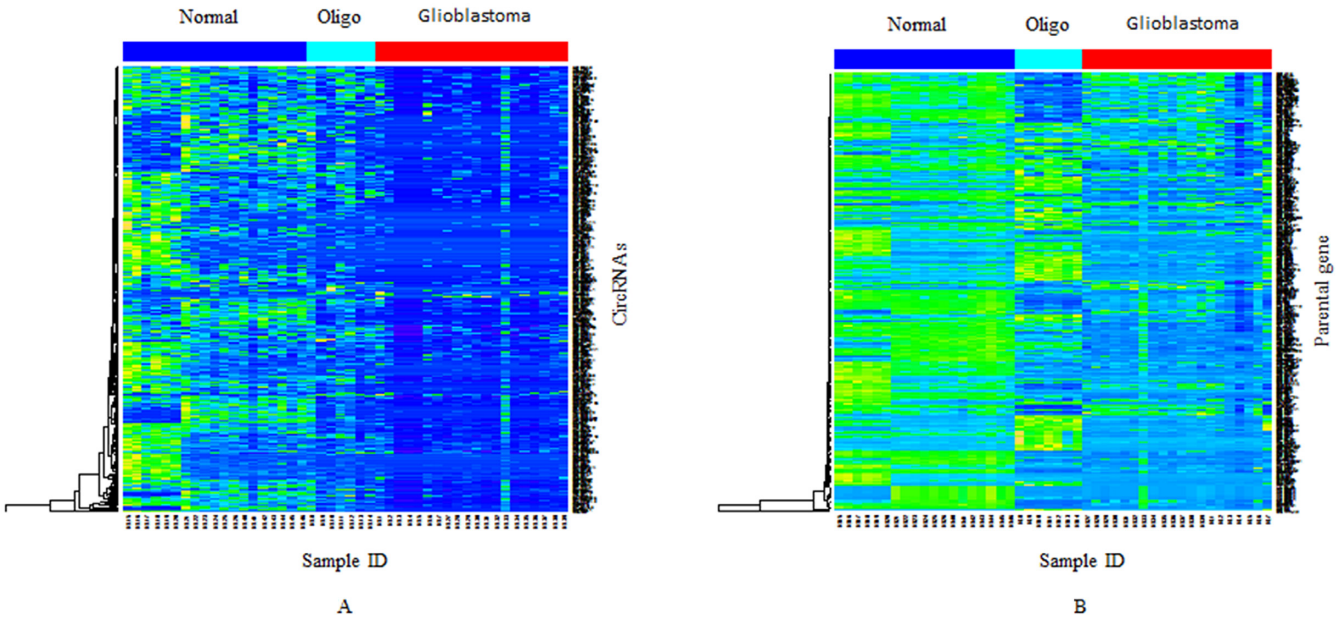


Figure 7. Differential circRNA expression and their parental mRNA expression in normal tissue and gliomas. (**A**) The heat map of expressed circRNAs in normal, oligodendroblastoma, and GBM (RPM > 0.1); (**B**) The heat map of parental gene related to circRNAs in normal, oligodendroblastoma and GBM.

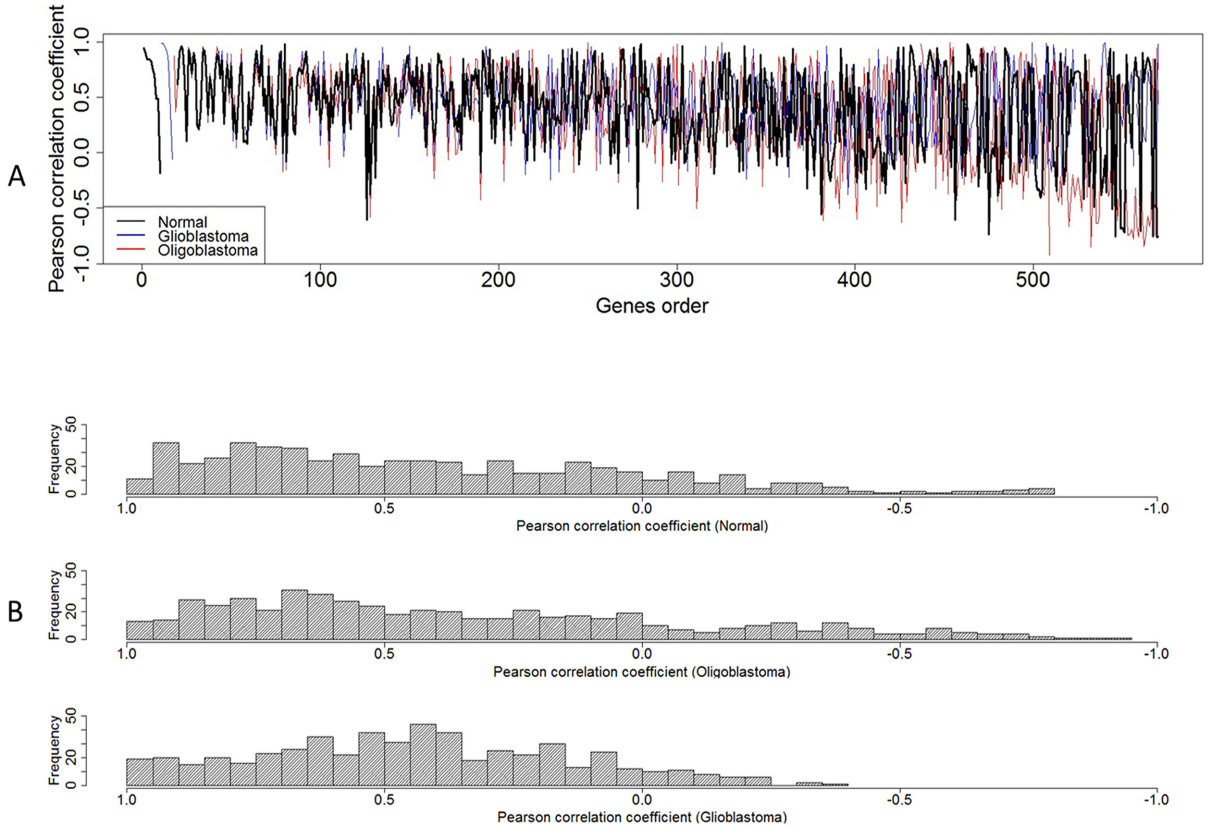


Figure 8. Pearson correlation coefficient between circRNAs and their parental mRNA. (**A**) The genes were arranged from left to right according to the RSME value from lower to higher of PCC among normal, GBM, and oligodendroblastoma. Thus, the left 185 circRNAs and their parental mRNAs have the similar correlation pattern in normal or gliomas (RMSE < 0.15). The correlations of the remaining about 387 circRNAs with their parental mRNAs have different patterns in normal and gliomas; (**B**) The PCC Frequency distribution in normal and gliomas.

sue than in tumor tissues (Expression Pattern A (CPA)); (ii) circRNAs more highly expressed higher in tumors than in controls (Expression Pattern B (CPB)); and (iii) circRNAs with no difference in expression between normal and tumor tissue (Expression Pattern C (CPC)). For each pattern of circRNA expression, their parental gene expression also exists in three patterns: (i) parental gene expression is higher in normal than in tumor tissue, which is labeled by mRNA expression pattern A (MPA); (ii) parental gene expression is lower in normal than in tumor tissue, which is labeled by mRNA expression pattern B (MPB); and (iii) there is no difference in parental gene expression between normal and tumor tissue, which is labeled by mRNA expression pattern C (MPC). We listed these expression patterns of the circRNAs and their parental genes in Table 2. We estimated that circRNAs in these different expression patterns may play different roles in normal or abnormal biological processes.

First, those 353 CPA circRNAs and their MPA parental genes (such as CDR1, ATRNL1, AKT3, SPTAN1, ZNF483 and RIMS1) have a similar expression patterns in normal and GBM. Therefore, we hypothesized that they are byproducts from alternative splicing variants (29), or that they may function as miRNA sponges to elevate target parental gene expression in normal biological processes (18). For the 115 CPA circRNAs and their MPB or MPC parental genes (such as QKI, PSMA7, RHOBTB3, TMEM165, LIFR, CSNK1A1, VAMP3 and CAPZA1), we speculated that these circRNAs and their parental genes may function as tumor suppressors. The seven CPB circRNAs have a similar expression pattern to their MPB parental genes (such as COL1A2, PTN, VCAN, PLOD2, SMO, GLIS3 and EPHB4.). Therefore, we hypothesized that they are byproducts from alternative splicing variants (30), or that they may function as miRNA sponges to elevate target parental gene expression in abnormal biological processes (18). We did not detect any CPB circRNAs and their MPA parental genes in normal tissues and GBM. For the one CPB circRNA, circ_CLIP2, and its MPC parental gene, it is speculated that it may function as an oncogene in GBM. However, circ_CLIP2 is only expressed in a few of subjects. In the future, we need more subjects to investigate its regulatory roles in GBM biogenesis.

There are more than 476 differentially expressed circRNAs in normal tissue and GBM, but only 113 circRNAs are expressed differentially between normal tissue and oligodendroglioma. In addition, the number of CPA circRNAs and their MPA parental genes in normal and oligodendroglioma is far less than that in normal tissue and GBM. Although the number of highly expressed circRNAs in oligodentroglioma is relatively more than that in GBM, circ_VCAN is highly expressed in both of oligodentroglioma and GBM. It is worth mentioning

Table 2. circRNA and its parental gene expression pattern in normal tissue and gliomas

Normal versus Glioblastoma circRNA Expression pattern	mRNA Expression pattern	Gene name	Gene number
CPA (N > G) 468	MPA (N > G)	CDR1;ATRNL1;AKT3;SPTAN1;ZNF483;RIMS1;FKBP8;UNC13C; NTRK2;GABRB2;CADPS2; NRXN1;GLS;DYNC1H1;PDE4DIP;CAMK4; PJA2;TIAM1;ATP9A;CLASP2;SYNE1;	353
	MPB (N < G)	QKI;PSMA7;RHOBTB3;TMEM165;LIFR;CSNK1A1;VAMP3;CAPZA1; CORO1C;WBSCR22;PDE 4B;NUP205;ZFAND6;BACH1;SLC30A7;ASXL1; TTC28;TNPO3;DHX34;NCOA3;DCUN1D4	22
	MPC (N = G)	FAT1;MAN2A1;FAM208A;LZIC;PICALM;RBM39;CCDC9; QSER1;LPXN;PSMD5;DHDDS;PAN 3;.....	93
CPB (N < G) 8	MPA (N > G)	/	0
	MPB (N < G)	COL1A2;PTN;VCAN;PLOD2;SMO;GLIS3;EPHB4;.	7
	MPC (N = G)	CLIP2.	1
CPC (N = G)	OPHN1;NCOR2;KANSL1L;MDM2;QKI;CHD7;RELL1;LPHN3;MAML2;NUP54; FKBP8;PSD3;SAMD4A;TBC1D1;LPHN3;;.....		96
Normal versus Oligodendroglioma			
CPA (N > O) 93	MPA (N > O)	SPTAN1;RIMS1;AKT3;UNC13C;ZNF483;CADPS2;GLS;FKBP8;SYNE1; AKAP12;STXBP5L;DN AJC6;ANK3;SOGA2;PDE4DIP;ATP8A2; TTLL7;RPH3A.....	65
	MPB (N < O)	PDE4B;SMAD5;LRP6;TDRD3.	4
	MPC (N = O)	PICALM;ANKRD36B;PHIP;CRKL;PPA2;USP24;ZBTB44;USP45;NTRK2; PSMB1;MGA;PTK2;K LHL24;MDN1;KLHL24;FGGY;CNOT2;RBM33; MIB1;PHF21A;RPAP2;SLC12A2;	24
CPB (N < O) 20	MPA (N > O)	TBCD	1
	MPB (N < O)	SCARNA2;VCAN;ZBTB20;OPHN;AGXT2L1;MKLN1;AASS;NFIB;SOX6; LPHN3;LPHN3;TEX9 ;MTHFD2L;SMO;MAP3K1.	15
	MPC (N = O)	KDM4B;ZNF292;ZNF362;LDLRAD3.	4
CPC (N = O)	MDM2;GLIS3;GLI1;CSMD2;COL1A2;ANKIB1;MAN2A1;CAMSAP1;CDK8;TBC1D1; UBR5;CUX1;PICALM;UBA2;DDX3Y;ARH GAP12;FAM208A;UBE3D;LZIC;CDK13;.....		460

CPA: circRNA expression pattern A (normal > tumor); CPB: circRNA expression pattern B (normal < tumor); CPC: circRNA expression pattern C (normal = tumor); MPA : parental mRNA expression pattern A (normal > tumor); MPB : parental mRNA expression pattern A (normal < tumor); MPC : parental mRNA expression pattern A (normal = tumor); normal > tumor means that circRNA or mRNA expression level is higher in normal than that in tumor tissue, normal < tumor means that circRNA or mRNA expression level is lower in normal than that in tumor tissue, normal < tumor means that there is no difference between normal and tumor tissue, the difference significance is tested by Wilcoxon rank-sum test and Benjamini–Hochberg correction ($q < 0.05$).

that circ_VCAN parental genes have been known to be expressed in brain tissue (31,32), and that the VCAN gene has been experimentally validated to be the causal gene of gliomas (31,32). Thus, the highly expressed VCAN-derived circRNAs in gliomas are also supposed to be related to glioma biogenesis. However, the circRNA shares the same exons with one isoform of its parental gene, so the canonical knock down or over-expressed experimental methods cannot discriminate the circular or linear isoform of the gene (33). It is imperative that an experimental method specifically designed for studying circRNA function should be developed in the circRNA research field in the near future. This would allow us to determine whether circRNA or its parental gene is the main causal gene of glioma.

DISCUSSION

We have developed a computational pipeline named UROBORUS with the goal of identifying genome-wide circRNA based on total RNA-seq data. Employing this pipeline in gliomas and normal tissue, we first revealed the circRNA profile in gliomas, detected thousands of circRNAs in each sample and found a set of highly expressed circRNAs not detected by two other methods. We observed a notable decrease in circRNA amounts from normal tissue to oligodendroglioma and GBM, suggesting that the abnormal conditions in tumors leads to the failure of the back-spliced events required for forming circRNA. The alternative circularization numbers (AC numbers) for some parental genes in glioma tissue were decreased compared to AC numbers in normal tissue, indicating that alternative circRNA formation was also altered in tumor conditions.

Compared with *CIRCexplorer* and *find_circ* methods, the UROBORUS pipeline is a user-friendly and efficient tool for detecting circRNAs expressed at low levels in total RNA-seq data. The *find_circ* is a linux script tool, which cannot be easily used by biologists. The amount of circRNA identified by *find_circ* in our RNA-seq data was much less than by the *CIRCexplorer and* UROBORUS pipelines. *CIRCexplorer,* a modification from TopHat_fusion that employs a two-step mapping strategy, creates a Bowtie index from spliced contigs formed by the first step of mapping results. *CIRCexplorer* then aligns all of the un-mapping reads against the spliced contigs within one gene region. Thus, *CIRCexplorer* inevitably misses those UMJ reads supporting circRNAs and overestimates the circRNA expression level. The UROBORUS pipeline can avoid these detecting biases by taking into consideration short spanning reads, and re-aligning both BMJ and UMJ reads to the whole genome as opposed to a Bowtie index from spliced contigs. Although each of the three tools use different strategies to find different circRNAs, UROBORUS can find more common, highly-expressed circRNA than the other two tools as the expression level increases. This suggests that the detection of circRNA expressed at low levels is key to these tools. Unlike the other two approaches based on 20 bp spanning two sites of the junction, the UROBORUS pipeline can find those circRNA supported by short spanning reads. This capability increases its sensitivity and allows it to find circRNAs expressed at lower levels with 0.013 FDR. The six circRNAs expressed at relatively low levels in some samples (circ_STXBP5L, circ_FKBP8, circ_CORO1C, circ_ZNF148, circ_ RNF10, and circ_ SAFB2), successfully validated in the RT-PCR and Sanger sequencing experiments, and also indicated that the UROBORUS pipeline can detect circRNA expressed at low levels, when compared with the other two available tools.

The current version of the UROBORUS pipeline can only detect those circRNAs supported by exon–exon junctions, and may miss those circRNAs derived from an intron region or intergenic region. Although these intron- or intergenic region-derived circRNAs only represent a small frac-

tion of total circRNAs, we also plan to develop a second-generation UROBORUS pipeline to detect more circRNAs derived from intron regions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the anonymous reviewer's comments and suggestions.

FUNDING

National Natural Science Foundation of China [61171191, 61571223 to X.S, 81270700 to P.H]; Specialized Research Fund for the Doctoral Program of Higher Education [20133218110016 to X.S]; 'Six Talent Peak' Project of Jiangsu Province [SWYY-021 to X.S]; China Scholar Council [201203070162 to X.S]. Funding for open access charge: Natural Science Foundation of China [61171191, 61571223, 81270700]; Specialized Research Fund for the Doctoral Program of Higher Education in China [20133218110016]; 'Six Talent Peak' Project of Jiangsu Province [SWYY-021]; China Scholar Council [201203070162].
Conflict of interest statement. None declared.

REFERENCES

1. Sanger,H.L., Klotz,G., Riesner,D., Gross,H.J. and Kleinschmidt,A.K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 3852–3856.
2. Capel,B., Swain,A., Nicolis,S., Hacker,A., Walter,M., Koopman,P., Goodfellow,P. and Lovell-Badge,R. (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, **73**, 1019–1030.
3. Cocquerelle,C., Daubersies,P., Majérus,M.A., Kerckaert,J.P. and Bailleul,B. (1992) Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.*, **11**, 1095–1098.
4. Cocquerelle,C., Mascrez,B., Hétuin,D. and Bailleul,B. (1993) Mis-splicing yields circular RNA molecules. *FASEB J.*, **7**, 155–160.
5. Burd,C.E., Jeck,W.R., Liu,Y., Sanoff,H.K., Wang,Z. and Sharpless,N.E. (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.*, **6**, e1001233.
6. Braun,S., Domdey,H. and Wiebauer,K. (1996). Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. *Nucleic Acids Res.*, **24**, 4152–4157.
7. Pasman,Z., Been,M.D. and Garcia-Blanco,M.A. (1996) Exon circularization in mammalian nuclear extracts. *RNA*, **2**, 603–610.
8. Guo,J.U., Agarwal,V., Guo,H. and Bartel,D.P. (2014) Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.*, **15**, 409–422.
9. Lasda,E. and Parker,R. (2014) Circular RNAs: diversity of form and function. *RNA*, **20**, 1829–1842.
10. Danan,M., Schwartz,S., Edelheit,S. and Sorek,R. (2012). Transcriptomewide discovery of circular RNAs in Archaea. *Nucleic Acids Res.*, **40**, 3131–3142.
11. Jeck,W.R., Sorrentino,J.A., Wang,K., Slevin,M.K., Burd,C.E., Liu,J., Marzluff,W.F. and Sharpless,N.E. (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, **19**, 141–157.
12. Salzman,J., Gawad,C., Wang,P.L., Lacayo,N. and Brown,P.O. (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One*, **7**, e30733.
13. Jeck,W.R. and Sharpless,N.E. (2014) Detecting and characterizing circular RNAs. *Nat. Biotechnol.*, **32**, 453–461.
14. Salzman,J., Chen,R.E., Olsen,M.N., Wang,P.L. and Brown,P.O. (2013) Cell-type specific features of circular RNA expression. *PLOS Genet.*, **9**, e1003777.
15. Hoffmann,S., Otto,C., Doose,G., Tanzer,A., Langenberger,D., Christ,S., Kunz,M., Holdt,L.M., Teupser,D., Hackermüller,J. *et al.* (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.*, **15**, R34.
16. Gao,Y., Wang,J. and Zhao,F. (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.*, **16**, 4–19.
17. Memczak,S., Jens,M., Elefsinioti,A., Torti,F., Krueger,J., Rybak,A., Maier,L., Mackowiak,S.D., Gregersen,L.H., Munschauer,M. *et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333–338.
18. Hansen,T.B., Jensen,T.I., Clausen,B.H., Bramsen,J.B., Finsen,B., Damgaard,C.K. and Kjems,J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384–388.
19. Ivanov,A., Memczak,S., Wyler,E., Torti,F., Porath,H.T., Orejuela,M.R., Piechotta,M., Levanon,E.Y., Landthaler,M., Dieterich,C. *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.*, **10**, 170–177.
20. Zhang,X.O., Wang,H.B., Zhang,Y., Lu,X., Chen,L.L. and Yang,L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134–147.
21. Zhang,Y., Zhang,X.O., Chen,T., Xiang,J.F., Yin,Q.F., Xing,Y.H., Zhu,S., Yang,L. and Chen,L.L. (2013) Circular intronic long noncoding RNAs. *Mol. Cell*, **51**, 792–806.
22. Trapnell,C., Pachter,L. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
23. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
24. Huse,J.T., Wallace,M., Aldape,K.D., Berger,M.S., Bettegowda,C., Brat,D.J., Cahill,D.P., Cloughesy,T., Haas-Kogan,D.A., Marra,M. *et al.* (2014) Where are we now? And where are we going? A report from the Accelerate Brain Cancer Cure (ABC2) Low-grade Glioma research workshop. *Neuro Oncol.*, **16**, 173–178.
25. Kim,D. and Salzberg,S.L. (2011) TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.*, **12**, R72.
26. Bachmayr-Heyda,A., Reiner,A.T., Auer,K., Sukhbaatar,N., Aust,S., Bachleitner-Hofmann,T., Mesteri,I., Grunt,T.W., Zeillinger,R. and Pils,D. (2015) Correlation of circular RNA abundance with proliferation–exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.*, **27**, 8057–8066.
27. Li,Z., Huang,C., Bao,C., Chen,L., Lin,M., Wang,X., Zhong,G., Yu,B., Hu,W., Dai,L. *et al.* (2015) Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.*, **22**, 256–264.
28. Conn,S.J., Pillman,K.A., Toubia,J., Conn,V.M., Salmanidis,M., Phillips,C.A., Roslan,S., Schreiber,A.W., Gregory,P.A. and Goodall,G.J. (2015) The RNA binding protein quaking regulates formation of circRNAs. *Cell*, **160**, 1125–1134.
29. Cocquerelle,C., Mascrez,B., Hetuin,D. and Bailleul,B. (1993) Mis-splicing yields circular RNA molecules. *FASEB J.*, **7**, 155–160.
30. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
31. Yang,W. and Yee,A.J. (2013) Versican V2 isoform enhances angiogenesis by regulating endothelial cell activities and fibronectin expression. *FEBS Lett.*, **587**, 185–192.
32. Uhlén,M., Fagerberg,L., Hallström,B.M., Lindskog,C., Oksvold,P., Mardinoglu,A., Sivertsson,Å., Kampf,C., Sjöstedt,E., Asplund,A. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
33. Glazar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.