The functional circular RNA, ciRS-7 (CDR1as), is biosynthesized using back-splicing promoted by inverted mammalian-wide MIRs but not primate-specific *Alu*s

Rei Yoshimoto[1], Karim Rahimi[2,3], Karoline K. Ebbesen[2,3], Thomas Hansen[2], Jørgen Kjems[2,3] & Akila Mayeda[1]*

[1]Division of Gene Expression Mechanism, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, 470-1192, Japan
[2]Department of Molecular Biology and Genetics, Aarhus University, C.F. Møllers Allé 3, 8000 Aarhus C, Denmark
[3]Interdisciplinary Nanoscience Center, Aarhus University, Gustav Wieds Vej 14, 8000 Aarhus C, Denmark

*Correspondence and requests for materials should be addressed to A.M. (e-mail: mayeda@fujita-hu.ac.jp)

SUMMARY

CircRNAs are abundant stable noncoding RNAs with covalently closed circular structure. The first characterized circRNA, ciRS-7 (CDR1as), acts as a 'sponge', or decoy, that regulates the stability and activity of miR-7 micro RNA. Here we demonstrate the definitive biosynthesis pathway of this important ciRS-7. By anti-sense oligoribonucleotide targeting of the relevant splice sites, we found that ciRS-7 is directly generated by back-splicing between the flanking splice sites rather than by subsequence splicing of the excised lariat RNA containing skipped exon and introns. We further validated this observation by CRISPR/Cas9-mediated genomic deletion of the relevant splice sites. Back-splicing circularization events are promoted by the flanking intronic complementary sequences, and only inverted repeats of *Alu*, primate-specific SINE, was previously identified. ciRS-7 lacks flanking available *Alu* elements, however, we identified conserved flanking inverted elements of MIR (mammalian-wide interspersed repeat). Splicing reporter assays in stably transfected HEK293 cells with mini-genes, with and without the inverted MIR elements, demonstrated they are required to generate ciRS-7. Using bioinformatics searches followed by mini-gene reporter tests, we identified several other MIR-dependent circRNAs. The MIR-dependent circRNAs are thus a new category of widely conserved mammalian circRNAs.

INTRODUCTION

Endogenous circular RNAs (circRNAs) were first observed as scrambled exon transcripts (Nigro et al., 1991) and these kinds of transcripts were found to have circular structures with covalently closed ends (Capel et al., 1993; Cocquerelle et al., 1993). For two decades, however, they were disregarded as rare oddities, or regarded as poorly expressed mis-spliced products.

High-throughput sequencing of full transcriptomes (RNA-Seq) has identified thousands of circRNAs in eukaryotes, which are now considered as common byproducts of many protein-coding genes (Salzman et al., 2012; Jeck et al., 2013; Memczak et al., 2013). The functions of the vast majority of these circRNAs remain unknown. However, some of these circRNAs have important biological functions. For instance, certain circRNAs control the stability and activity of micro RNAs (miRNAs), regulate transcription or alternative splicing, affect translation of host genes, or they can even be translated and produce proteins themselves (reviewed in Wilusz, 2018). ciRS-7 (also known as CDR1as) was one of the first functionally annotated circular RNAs. It is conserved among mammals and is mainly expressed in brain. ciRS-7 has many binding sites for a particular miRNA, miR-7, and a single binding site for miR-671 that triggers Argonaute2-catalized slicing of ciRS-7 (Hansen et al., 2011). *In cellulo* experiments suggested that ciRS-7 may function as sponges, or decoys, that reduce the available free miR-7 and thus prevent repression of miR-7 targeted mRNAs (Hansen et al., 2013; Memczak et al., 2013). Knockout mice lacking the ciRS-7 genomic locus down-regulated miR-7 in brain, suggesting ciRS-7 has a role in stabilizing or transporting miR-7 (Piwecka et al., 2017). As a result, the ciRS-7 knockout mice had impaired sensorimotor gating, which is associated with neuropsychiatric disorders.

Recent gene editing experiments revealed a comprehensive regulatory network in mouse brain, between a long non-coding RNA (ncRNA), circRNA, and two miRNAs (Kleaveland et al., 2018). This network involves Cyrano ncRNA that promotes the degradation of miR-7, which in turn enhances the miR-671-directed degradation of ciRS-7, and thus Cyrano causes an accumulation of ciRS-7. ciRS-7 is apparently a key component in a gene-regulatory network in brain, however, understanding the biosynthesis and transport of this particular circRNA remains an important challenge.

Two major circRNA biosynthesis pathways have been proposed (Figure 1A; reviewed in Jeck and Sharpless, 2014; Wilusz, 2018). Splicing reaction between downstream 5′ and upstream 3′ splice sites of circularized exon(s) are involved in both pathways. However, this

specific splicing occurs either directly on the loop-structure formed *via* the flanking intronic complementary sequences (Figure 1A; 'Back-splicing' pathway), or on the lariat-structure generated by exon-skipping splicing (Figure 1A; 'Intra-larat splicing pathway). In human, the pair of inverted repeats of *Alu*, primate-specific repetitive element, stimulates direct back-splicing (Jeck et al., 2013; Liang and Wilusz, 2014; Zhang et al., 2014). However, it is not known if similar intronic inverted sequences promotes back-splicing in non-primate circRNAs. In an artificial ciRS-7 expression vector, an insertion of 800 nucleotides (nt) of perfectly complementary sequences in an intron is technically sufficient to circularize ciRS-7 (Hansen et al., 2013). There is, of course, no such extensive complementarity around the endogenous ciRS-7 gene, and thus the mechanism of its production remains to be defined.

Here we elucidate definitive generation pathway of the endogenous ciRS-7. Remarkably, we identified highly conserved mammalian-wide interspersed repeat (MIR) sequences in the flanking introns of the ciRS-7, and we demonstrated that these flanking inverted MIR elements are required to generate ciRS-7. Bioinformatics analyses followed by reporter assays identified other distinct circRNAs generated by inverted MIR sequences. We conclude that a certain set of mammalian circRNAs that lack primate-specific *Alu* sequences are generated by mammalian-wide MIR-mediated back-splicing.

RESULTS

Genomic structure around ciRS-7 and pathways for ciRS-7 biosynthesis

To determine whether ciRS-7 is biosynthesized through the 'Intra-lariat splicing pathway' or 'Back-splicing pathway' (Figure 1A), it is crucial to examine the genomic structure around the corresponding sequence of ciRS-7. We therefore investigated the assembled transcripts from the RNA-Seq dataset of human brain (Gill et al., 2014). The ciRS-7 precursor extends over ~80 kb and this non-coding transcript has six exons and five introns (Supplemental Figure S1).

Our analysis of 3′ rapid amplification of cDNA ends (RACE) showed that the 3′ terminal of ciRS-7 precursor contained canonical poly (A) tail (from downstream of chrX: 139873517). The last exon 6 of the locus is long (~10 kb) and contains the entire ciRS-7 sequence (Supplemental Figure S1). We found the splice junction reads within the last two exons 5, 6 across the ciRS-7 locus (Supplemental Figure S2), which might be outer alternative splice sites to generate lariat RNA including ciRS-7 locus in the 'Intra-lariat splicing pathway' (Figure 1A; left panel).

The ciRS-7 precursors and multiple products were identified

To identify the intermediates and final products that could be generated by splicing events around the ciRS-7 locus, we performed RT–PCR analysis with human brain total RNA (Figure 1B). We took advantage of the RNase R (*E. coli* 3′ to 5′ exonuclease) digestion prior to RT–PCR that can distinguish opened linear RNA from closed RNA, such as circRNA and lariat RNA (Suzuki et al., 2006). The PCR-amplified products were analyzed by sequencing.

We successfully detected final closed circular form of ciRS-7 (C) together with linear form of several final spliced products (M) due to multiple alternative 5′ and 3′ splice sites (Figure 1B, C). The detection of precursor transcript (P) using primers across the 3′ end of ciRS-7 exon, even after RNase digestion, appeared to be lariat RNA (L1), suggesting that ciRS-7 is produced *via* a lariat precursor (Figure 1A; 'Intra-lariat splicing pathway'). However, we could not rule out the possible detection of the looped form of precursor (P′) due to the incomplete RNase R digestion. If this was the case, the ciRS-7 is likely produced by direct back-splicing (Figure 1A; 'Back-splicing pathway').

Splice site targeting experiments demonstrates ciRS-7 generation by back-splicing

These two pathways can be distinguished by blocking of specific splicing events using antisense oligoribonucleotides (ASOs). The ASOs 1+2 targeting the flanking 5′ and 3′ splice sites of ciRS-7 exon (Figure 1A, C; blue bars) can prevent circular ciRS-7 generation *via* both pathways, so these ASOs 1+2 were used as the positive control (Figure 1A). In contrast, the

ASOs 3–6 targeting alternative 5′ and 3′ splice sites outside of ciRS-7 exon (Figure 1A, C; red bars) can prevent the exon-skipping and the following circular ciRS-7 generation in the 'Intra-lariat pathway' (Figure 1A; left panel). In the 'Back-splicing pathway', these same ASOs 3–6 cannot interfere with the back-splicing event, and thus allow circular ciRS-7 generation (Figure 1A; right panel).

We transfected human SH-SY5Y cells with these ASOs and we found that the ASOs 1+2 targeting 5′ and 3′ splice sites of ciRS-7 exon significantly repressed ciRS-7 production, given the efficiencies of transfection and annealing of ASOs (Figure 1D; blue lane). However, none of the ASOs 3–6 targeting alternative 5′ and 3′ splice sites outside of the ciRS-7 exon inhibited ciRS-7 production (Figure 1D; red lane). These data thus imply that the 'Back-splicing pathway', but not 'Intra-lariat pathway', produces circular ciRS-7 (Figure 1A).

To further validate this finding, we made a genomic deletion of the same relevant splice sites by CRISPR/Cas9-mediated technology (Supplemental Figures S3A and S3B). The deletion of either the alternative 5′ splice sites (between gRNA2 and gRNA3) or the 3′ splice sites (between gRNA4 and gRNA5) outside of ciRS-7 exon barely prevented the generation of ciRS-7 (Figure S3C), confirming that circular ciRS-7 is biosynthesized through the 'Back-splicing pathway' rather than 'Intra-lariat pathway'.

The ciRS-7 locus is flanked by inverted MIR sequences

Several human circRNAs are biosynthesized through the 'back-splicing pathway' facilitated by RNA pairing between inverted repeats of *Alu* elements, which are the most abundant transposable elements in the human genome (Jeck et al., 2013; Liang and Wilusz, 2014; Zhang et al., 2014; Zheng et al., 2016). Since we found that ciRS-7 is made by 'Back-splicing pathway', we searched for the prerequisite inverted *Alu* elements on the UCSC RepeatMasker track. There are two human *Alu* elements upstream of ciRS-7 within exon 5 and exon 6 (Supplemental Figure S1). However, there is no *Alu* element downstream that could generate the required base-pairing around the ciRS-7. Therefore, flanking inverted *Alu* elements cannot generate ciRS-7.

Despite this fact, we did find two relatively conserved regions inside the last exon 6 of ciRS-7 precursor (Figure 2A). These regions overlapped with mammalian-wide interspersed repeats (MIRs), which is an ancient short interspersed nuclear elements (SINE) conserved among mammals and marsupials (Jurka et al., 1995; Krull et al., 2007). Remarkably, these MIRs were located upstream and downstream of ciRS-7 exon in an inverted orientation that is conserved both in human and mouse (Figure 2A). This observation led us to postulate that ciRS-7 is generated *via* 'Back-splicing pathway' using inverted MIRs, instead of *Alu*s.

Inverted MIRs promote ciRS-7 biosynthesis

To test the above hypothesis, we used a doxycycline (DOX)-inducible circRNA expression system in a HEK293 stable cell line that efficiently generates a 5-kb transcript containing ciRS-7 exon flanked by upstream and downstream inverted MIRs (Figure 2B).

The ciRS-7 circular product and its precursor RNA were both detected, 24 h after induction with DOX (Figure 2C, left panel). We verified that the final ciRS-7 product was indeed circular by RNase R treatment prior to RT–PCR (Suzuki et al., 2006). RNase R degraded the linear control GAPDH, whereas the circular ciRS-7 product remained intact (Figure 2C, right panel). These results demonstrate that our mini-gene, covering the ciRS-7 exon and its inverted MIRs, recapitulates the endogenous generation of circular ciRS-7.

To confirm whether ciRS-7 generation from the mini-gene system was splicing-dependent, we used the general pre-mRNA splicing inhibitor Spliceostatin A (SSA) (Kaida et al., 2007; Yoshimoto et al., 2017). SSA inhibited ciRS-7 production, suggesting that ciRS-7 is indeed generated by splicing (Figure 2D). The observed modest inhibition by SSA is consistent with previous data in fruit fly cells where back-splicing was largely unaffected by splicing inhibition, either by SF3b knockdown or with splicing inhibitor Pladienolide B (Liang et al., 2017). Both SSA and Pladienolide B bind to SF3b sub-complex and interfere with branch site recognition (Corrionero et al., 2011; Folco et al., 2011; Yoshimoto et al., 2017). We still cannot rule out unknown splicing mechanism involved specifically in back-splicing on the looped structure,

which might be distinct from the established mechanism of canonical splicing.

Using this stable cell line system, we next made a series of MIR-deleted mini-genes to analyze the role of MIRs in human ciRS-7 biosynthesis (Figure 3A). The wild-type mini-gene efficiently generated ciRS-7 (Figure 3B; WT), whereas the deletion of either upstream or downstream MIR element seriously impaired ciRS-7 production (Figure 3B; Δ5′MIR, Δ3′MIR, Δ5′3′MIR), demonstrating that ciRS-7 generation depends on the inverted MIRs. We conclude that the ciRS-7 is biosynthesized *via* the 'Back-splicing pathway' promoted by its flanking inverted MIR sequences.

Other MIR-dependent and MIR-independent circRNAs were identified

Since MIRs are ancient SINEs and ubiquitous in mammal genomes (Jurka et al., 1995), it would be expected that the utilization of MIRs in promoting back-splicing is a globally conserved strategy rather than a gene-specific event.

Since we found that MIRs neighboring ciRS-7 are conserved in both human and mouse (Figure 2A), we searched for other MIR-dependent circRNAs. Using the RepeatMasker track in human hg19 coordinates, we found a total 595,094 of MIRs. Comparing with mouse mm9 coordinates, 104,074 are conserved between human and mouse. Previously, 7,771 circRNAs were identified from RNA-Seq data of foreskin cells (Jeck et al., 2013). Using these data sets, we screened 170 circRNA candidates that contains inverted MIRs within 1,000 nt upstream and downstream of the circularized exons.

We selected five circRNAs as representatives; circCDK8, circSPNS1, circTMEM109, circZYMND8, and circSRGAP2C, according to the size of precursors (<5 kb) and their ubiquitous expressions. With these five circRNAs, we searched for the located inverted MIRs in the flanking introns (Supplemental Figure S4). To test for MIR-dependent biosynthesis of these circRNAs, we made a series of MIR-deleted mini-genes as we had done for ciRS-7 (Figure 4A). With these human mini-genes, we transfected mouse N2A cells, in which ectopically expressed human circRNAs could be discriminated from the endogenously expressed mouse circRNAs. All the circRNAs, whose circular structures were verified by RNase R digestion, were successfully detected from the wild-type mini-genes (Figure 4B). Deletion of either upstream or downstream MIRs prevented the production of circCDK8 and circSPNS1 (Figure 4C; MIR-dependent circRNAs), but had no effect on the production of circTMEM109, circZYMND8 and circSPGAP2C (MIR-independent circRNAs).

MIR-MIR base-pairing stability is critical for circRNA generation

MIRs are older SINEs than *Alu*s (~130 vs. ~60 million years ago), and MIR elements have often become truncated during this time (Jurka et al., 1995). We assumed that such truncations would cause instability of MIR-MIR base-paring, which is required for efficient circRNA production.

To examine this hypothesis, we evaluated the integrity of MIRs located around MIR-dependent and MIR-independent circRNAs by comparing them to the MIR consensus sequence using a Smith-Waterman (SW) alignment score (Smith and Waterman, 1981). As we predicted, flanking MIRs of MIR-dependent circRNAs had markedly higher alignment scores than those of MIR-independent circRNAs (Figure 5A; Supplemental Figure S4 for the individual score of all MIRs). We also observed that the MIRs of MIR-dependent circRNAs are significantly longer than those of MIR-independent circRNAs (Figure 5B).

Using these criteria of SW alignment score, we estimated the number of MIR-dependent circRNAs from our identified 170 circRNAs with flanking inverted MIRs. We found 118 and 19 circRNAs that have higher SW scores than 250 and 500, respectively (Figure 5C). Because of the high MIR-SW scores (573–1077) of MIR-dependent circRNAs (Supplemental Figure S4), these 19 circRNAs (>500 MIR-SW scores) are very likely to be MIR-dependent circRNAs. We calculated the thermal stability of MIR-MIR pairing of these groups of circRNAs. The results indicated that the inverted MIRs with higher SW alignment scores can form more stable MIR-MIR paring (Figure 5D).

Finally, we tried to predict the complementarity of inverted MIRs of all six selected circular RNAs together with three *Alu*-dependent circRNAs. The MIR-MIR base-paring of

MIR-dependent circRNAs are apparently more stable than those of MIR-independent circRNAs (Figure 6). Together, we conclude that well conserved MIRs close to the consensus sequence can promote back-splicing by forming adjacent stable RNA paring.

DISCUSSION

The circular RNA ciRS-7 was first characterized as a regulator of a specific miRNA, miR-7. Here we demonstrate the definitive pathway for the biosynthesis of ciRS-7, using human cells stably expressing ciRS-7 precursor. Although ciRS-7 precursor has available alternative splice sites for the 'Intra-lariat splicing pathway', our experimental evidence proves the 'Back-splicing pathway' that is facilitated by the newly found inverted sequences from MIR, a well conserved mammalian SINE. We found a new category of mammalian circRNAs, in which MIRs promote back-splicing and RNA circularization.

Previous results are consistent with the MIR-dependent back-splicing of ciRS-7

Previous ciRS-7 back-splicing assays used the reporter plasmid that includes the endogenous flanking genomic sequence, i.e., 1 kb upstream and 0.2 kb downstream from ciRS-7 exon (Hansen et al., 2013). This reporter could not produce ciRS-7 and this observation is consistent with the fact that the critical inverted MIRs for back-splicing are located ~3.0 kb and ~0.6 kb away from the 3′ and 5′ splice sites of ciRS-7 exon, respectively.

Recently, the genomic structure around the ciRS-7 precursor has also been independently reported (Barrett et al., 2017). In agreement with our analysis, they showed that the promoter region of ciRS-7 overlaps an upstream long non-coding transcript (LINC00632 locus), suggesting that ciRS-7 precursor is a long non-coding RNA (Barrett et al., 2017). Their obtained BAC clones (185 and 200 kb) cover from the promoter of this non-coding transcript (LINC00632) to far downstream of the ciRS-7 exon, and they showed that these two clones could produce mature ciRS-7. The rationale is that these two BAC clones includes upstream and downstream MIRs that are essential for ciRS-7 generation. Our evidence for the requirement of these MIRs using mini-gene reporters will be further confirmed by endogenous ciRS-7 genomic deletion of the MIRs by CRISPR/Cas9-mediated editing, which is in progress.

MIR-dependent back-splicing is widely utilized for mammalian circRNA biosynthesis

The direct back-splicing pathway is commonly used to generate circRNA in metazoans (Ivanov et al., 2015), however only flanking inverted *Alu* sequences have been identified to be essential for back-splicing. Indeed *Alus* are likely major players in the biosynthesis of human circRNAs (Jeck et al., 2013; Zhang et al., 2014) because of their abundance (~10% in human genome)(Price et al., 2004). However, the *Alu*-SINEs are relatively young (~60 million years) and they only exist in primates. MIR-SINEs are much older (~130 millon years ago) and globally functional in mammalian (Krull et al., 2007). Therefore, it is reasonable that MIRs are generally used in the biosynthesis of mammalian circRNAs.

Due to their age many MIR sequences lost their 5′ and 3′ regions that reduced their potential to form complementary duplexes, whereas the more recently appeared *Alu* sequences have remained more complementary with one another. It was reported that short complementary sequences (30–40 nt) in the inverted *Alu* elements are required to support efficient back-splicing (Liang and Wilusz, 2014). This complementarity criteria for back-splicing would be applicable for MIR elements, and therefore, we found the MIR-independent circRNAs that must not be fulfilled this criteria. Interestingly, the difference between MIR-dependent and MIR-independent circRNAs can be apparently reflected in the predicted secondary structures of their MIR sequences (see Figure 6).

MIR-derived circRNAs likely have important functions

Our bioinformatics analysis selected 19 circRNAs having conserved flanking MIRs with the highest complementarity, which are thus expected to be biosynthesized *via* a MIR-dependent pathway. Among these, we verified that circCDK8 and circSPNS1, as well as ciRS-7, were

indeed generated through flanking inverted MIR elements. Since circRNA often regulates the expression of the host gene (reviewed in Wilusz, 2018), these MIR-derived circRNAs likely have important functions.

The CDK8 kinase is a component of the mediator complex that regulates transcription by RNA Polymerase II (Galbraith et al., 2010). CDK8 acts as a coactivator in the β-catenin pathway, the p53 pathway, the serum response pathway, and the TGF-β signaling pathway. CDK8 functions as an oncogene in colon cancers through effects on Wnt/β-catenin signaling, and chemical inhibitors targeting CDK8 kinase could inhibit the growth of cancers (Clarke et al., 2016). The circRNA derived from the CDK8 transcript (circCDK8) that consists of coding exons 10–12 was previously described based on RNA-Seq data (Salzman et al., 2012; Jeck et al., 2013; Zhang et al., 2014; Song et al., 2016). We could find two and three target sites for miR-615-5p and miR-7162-3p, respectively (data not shown); however, whether the circCDK8 functions as so called 'sponges' of these miRNAs remain to be elucidated. It was demonstrated that circRNA generation *via* back-splicing and mRNA production *via* conventional splicing are mutually exclusive events (Ashwal-Fluss et al., 2014). Since the generated circCDK8 accounts for ~5% of the CDK8 transcripts (Jeck et al., 2013), it potentially competes under certain conditions with the production of CDK8 mRNA. The circCDK8 product may act as a modulator for these important CDK8 activities.

The *SPNS1* gene is a human homologue of fruit fly *spinster* gene. SPNS1 protein is a putative lysosomal H^+-carbohydrate transporter that is required for autophagy (Sasaki et al., 2014). *In cellulo* evidence suggested that SPNS1 could ameliorates dysfunction of autophagy in lysosomal storage disorder known as Niemann-Pick type C disease (Yanagisawa et al., 2017). The circSPNS1 consists of coding exons 4–8 that accounts for ~1% of the SPNS1 transcripts (Jeck et al., 2013).

To examine the function of such MIR-derived circRNA, it is essential to knock-out the target circRNA while retaining the host gene expression, or the normal spliced mRNA. Now, we can take advantage of the detected MIRs to specifically attenuate circRNA expression using CRISPR/Cas9 editing. To determine the biological and physiological functions of circCDK8 and circSPNS1 using this technology is underway.

EXPERIMENTAL PROCEDURES

Analysis of RNA-Seq datasets

We used RNA-Seq datasets from non-neoplastic brain tissue (GSE59612; Gill et al., 2014). Obtained data were aligned to reference human genome hg19 using HISAT2 (Kim et al., 2015) and the aligned sequence reads were assembled by StringTie (Pertea et al., 2015). Repeated sequences were downloaded from the RepeatMasker track in the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTables) and analyzed with BEDtools (http://bedtools.readthedocs.io).

The box plots were constructed using the R/Bioconductor package (http://www.bioconductor.org). For statistical comparisons of groups, Wilcoxon rank-sum tests were used to calculate P-values.

Possible RNA pairings between inverted MIRs were predicted using the RNAcofold program from the Vienna RNA package (Lorenz et al., 2011). Thermodynamic stability of the base-paired MIRs was calculated with the RNAup program in the Vienna RNA package using '–b' option to include the probability of unpaired regions (Lorenz et al., 2011). Exceptionally weak MIR-MIR base-pairings that the RNAup program could not calculate their thermodynamic stabilities were set to zero values ('All circRNAs' in Fig. 5D).

RT–PCR assays

RT–PCR analysis was performed essentially as previously described (Yoshimoto et al., 2017). Human cerebral cortex total RNA was purchased from Clontech (CLN 636561). From culture cells, total RNA was isolated with a Nucleospin RNA kit (Machery Nagel) according to the manufacturer's instructions. 'On-column' DNase I digestion was performed to remove contaminated DNA. Purified RNA was reverse-transcribed using PrimeScript II (Takara Bio)

with oligo-dT and random primers, and cDNA was amplified by PCR (20–30 cycles) with Ex-Taq (Takara Bio) and specific primers (Supplemental Table S1). PCR-amplified products were analyzed by 8% polyacrylamide gel electrophoresis.

For nested PCR, the first PCR reaction mixture (30 cycles) was purified with a PCR cleanup column (Takara Bio) and the eluate was used for the second PCR reaction (30 cycles). The purified RNase R (1 μg) was added to digest 1μg of total RNA to remove linear RNA in a 20 μL reaction mixture at 30ºC for 30 min as described previously (Suzuki et al., 2006).

The Eco Real-Time PCR system (Illumina) was used to perform quantitative PCR (qPCR) with the same primers (Supplemental Table S1) that were used in the regular PCR. For the qPCR reactions, KAPA Taq PCR kit was used according to the manufacturer's protocol (Kapa Biosystems).

Antisense oligoribonucleotide-mediated splicing repression

Antisense 2′-*O*-methyl-modified phosphorothioate oligoribonucleotides (ASOs) were purchased from Integrated DNA Technologies (Supplemental Table S1). These chemically modified ASOs were electroporated into SH-SY5Y cells (Gene Pulser Xcell, Bio-Rad). Fully confluent SH-SY5Y cells grown on 10 cm plate in D-MEM/Ham's-F12 medium (Fujifilm Wako) were trypsinized and the washed cell pellets were suspended in 1 mL OPTI-MEM medium (Thermo Fisher Scientific). This cell suspension (0.2 mL) plus ASO (2.5–10 μM final concentration) were transferred into 0.4 cm cuvette (BEX) and electroporated at 200 V for 20 ms square-wave pulses. After electroporation, cell suspensions were transferred into 6-well plates supplemented with 1.8 mL D-MEM/Ham's-F12 that was cultured for 24 h before RT–PCR analysis.

Construction and expression of circRNA-reporter plasmids

The expression plasmids for ciRS-7, circCDK8, circTMEM109, circZMYNDR8, and circSRGAP2C were constructed by subcloning the PCR-amplified fragments into FLAG-pcDNA3 vector using BamHI and XhoI sites (Yoshimoto et al., 2009). The expression plasmids for circSPNS1 was constructed by subcloning the PCR-amplified fragments into FLAG-pcDNA3 plasmid using an In-Fusion HD cloning kit (Takara Bio) according to the manufacturer's protocol. The deletion of MIR element of these expression plasmids were made by conventional extension PCR method.

To establish HEK293T cells stably expressing ciRS-7, PCR-amplified a ciRS-7 fragment was subcloned into pcDNA5 FRT-TO vector (Thermo Fisher Scientific) using BamHI and XhoI sites, and transfected into Flp-In T-REx 293 cells (Thermo Fisher Scientific) along with Flp recombinase vector pOG44. The transfected cells were treated with 1 μg/mL DOX for 24 h and total RNA was prepared for the RT–PCR analysis. To inhibit the splicing reaction, 100 ng/mL SSA was added 18 h after DOX treatment.

Mouse N2A cells were transiently transfected with these expression plasmids using Lipofectamine 2000 (Thermo Fischer Scientific) according to the manufacturer's instructions. The transfected cells were incubated for 24 h and total RNA was prepared for RT–PCR analysis.

Targeted genomic deletion of the ciRS-7 locus by CRISPR/Cas9

Cas9 RNP was introduced with the Alt-R CRISPR-Cas9 System (Integrated DNA Technologies) according to the manufacturer's instructions. Briefly, HEK-293 cells were seeded into a 96-well plate at a density of 4×10^4 cells/well. The cells were transiently transfected with 0.75 pmol of Cas9 RNP using Lipofectamine RNAiMAX (Life Technologies). After culture for 2 weeks, genomic DNA was extracted and PCR was performed to verify the deleted region. The gRNA and primer sequences are listed in Supplemental Table S1.

ACKNOWLEDGEMENT

We are grateful to Drs. Kinji Ohno and Hitomi Tsuiji for SH-SY5Y cells and N2A cells, respectively. Spliceostatin A (SSA) was a generous gift from Dr. Minoru Yoshida. We thank Dr. Shinichi Nakagawa for critical reading of the manuscript, Dr. Makoto K. Shimada for

helpful suggestions for bioinformatics analysis, Taiwa Komatsu and Mayuko Tanahashi for providing instruction of the gene editing techniques, and members of our laboratory for constructive discussion.

R.Y. was supported by a Research Grant from the Hori Sciences and Arts Foundation, a Research Grant from the Nitto Foundation, a Science Research Promotion Fund from the Promotion and Mutual Aid Corporation for Private Schools of Japan (PMAC), and Grants-in-Aid for Scientific Research (C) (JP18K05563). A.M. was partly supported by Grants-in-Aid for Scientific Research (B) (JP16H04705) and Grants-in-Aid for Challenging Exploratory Research (JP16K14659) from the Japan Society for the Promotion of Science (JSPS).

AUTHOR CONTRIBUTIONS

R.Y. and A.M. conceived and designed the experiments; R.Y. performed most of the experiments and analyses, organized the data and drafted the manuscript; K.R. and K.K.E. performed CRISPR/Cas9-mediated genomic deletion of the MIR elements; T.H. and J.K. provided useful information and revised the manuscript; R.Y. and A.M. analyzed the data and edited the manuscript. A.M. coordinated and supervised the whole project. All authors read, corrected and approved the final manuscript.

REFERENCES

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. Mol. Cell *56*, 55–66.

Barrett, S.P., Parker, K.R., Horn, C., Mata, M., and Salzman, J. (2017). ciRS-7 exonic sequence is embedded in a long non-coding RNA locus. PLoS Genet. *13*, e1007114.

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.

Clarke, P.A., Ortiz-Ruiz, M.J., TePoele, R., Adeniji-Popoola, O., Box, G., Court, W., Czasch, S., El Bawab, S., Esdar, C., Ewan, K.*, et al.* (2016). Assessing the mechanism and therapeutic potential of modulators of the human Mediator complex-associated protein kinases. Elife *5*.

Cocquerelle, C., Mascrez, B., Hetuin, D., and Bailleul, B. (1993). Mis-splicing yields circular RNA molecules. FASEB J. *7*, 155–160.

Corrionero, A., Minana, B., and Valcarcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev. *25*, 445–459.

Folco, E.G., Coil, K.E., and Reed, R. (2011). The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. Genes Dev. *25*, 440–444.

Galbraith, M.D., Donner, A.J., and Espinosa, J.M. (2010). CDK8: a positive regulator of transcription. Transcription *1*, 4–12.

Gill, B.J., Pisapia, D.J., Malone, H.R., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J.S., Banu, M.*, et al.* (2014). MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. Proc. Natl. Acad. Sci. USA *111*, 12550–12555.

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.

Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., and Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J. *30*, 4414–4422.

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.

Jeck, W.R., and Sharpless, N.E. (2014). Detecting and characterizing circular RNAs. Nat. Biotechnol. *32*, 453–461.

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.

Jurka, J., Zietkiewicz, E., and Labuda, D. (1995). Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. Nucleic Acids Res. *23*, 170–175.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H.*, et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat. Chem. Biol. *3*, 576–583.

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

Kleaveland, B., Shi, C.Y., Stefano, J., and Bartel, D.P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. Cell *174*, 350–362 e317.

Krull, M., Petrusma, M., Makalowski, W., Brosius, J., and Schmitz, J. (2007). Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). Genome Res. *17*, 1139–1145.

Liang, D., Tatomer, D.C., Luo, Z., Wu, H., Yang, L., Chen, L.L., Cherry, S., and Wilusz, J.E. (2017). The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. Mol. Cell *68*, 940–954 e943.

Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. Genes Dev. *28*, 2233–2247.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms. Mol. Biol. *6*, 26.

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.

Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W., and Vogelstein, B. (1991). Scrambled exons. Cell *64*, 607–613.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.

Piwecka, M., Glazar, P., Hernandez-Miranda, L.R., Memczak, S., Wolf, S.A., Rybak-Wolf, A., Filipchyk, A., Klironomos, F., Cerda Jara, C.A., Fenske, P.*, et al.* (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. Science *357*, eaam8526.

Price, A.L., Eskin, E., and Pevzner, P.A. (2004). Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. Genome Res. *14*, 2245–2252.

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.

Sasaki, T., Lian, S., Qi, J., Bayliss, P.E., Carr, C.E., Johnson, J.L., Guha, S., Kobler, P., Catz, S.D., Gill, M.*, et al.* (2014). Aberrant autolysosomal regulation is linked to the induction of embryonic senescence: differential roles of Beclin 1 and p53 in vertebrate Spns1 deficiency. PLoS Genet. *10*, e1004409.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. J. Mol. Biol. *147*, 195–197.

Song, X., Zhang, N., Han, P., Moon, B.S., Lai, R.K., Wang, K., and Lu, W. (2016). Circular RNA profile in gliomas revealed by identification tool UROBORUS. Nucleic Acids Res. *44*, e87.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. Nucleic Acids Res. *34*, e63.

Wilusz, J.E. (2018). A 360 degrees view of circular RNAs: From biogenesis to functions. WIREs RNA *9*, e1478.

Yanagisawa, H., Ishii, T., Endo, K., Kawakami, E., Nagao, K., Miyashita, T., Akiyama, K., Watabe, K., Komatsu, M., Yamamoto, D.*, et al.* (2017). L-leucine and SPNS1 coordinately ameliorate dysfunction of autophagy in mouse and human Niemann-Pick type C disease. Sci. Rep. *7*, 15944.

Yoshimoto, R., Kaida, D., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Kawamura, Y., Hayashizaki, Y., Mayeda, A., and Yoshida, M. (2017). Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. RNA *23*, 47–57.

Yoshimoto, R., Kataoka, N., Okawa, K., and Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. Nucleic Acids Res. *37*, 891–902.

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.

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

FIGURES

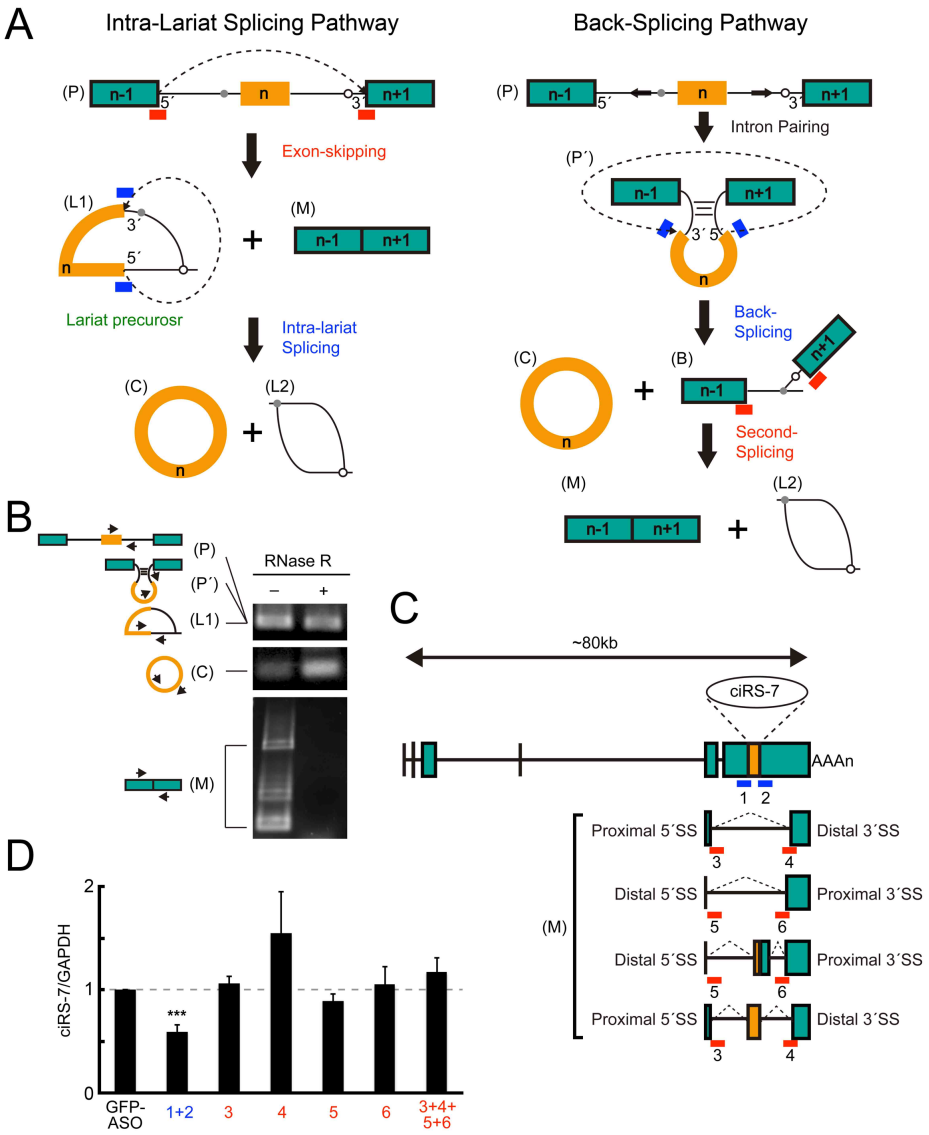

Figure 1. Evidence for 'Back-splicing' rather than 'Intra-lariat splicing' to produce ciRS-7
(A) Two proposed pathways for circRNA biosynthesis. Left panel: the generation by an intra-lariat splicing of lariat precursor (L1) that is excised by the preceding exon(s)-skipped splicing of pre-mRNA (P). Right panel: the circRNA (C) generation by a direct back-splicing event that is promoted by the looped pre-mRNA (P′) formed by base-paring of the flanking intronic complementary sequences. Blue and red bars indicate the antisense oligoribonucleotides (ASOs), numbered 1–6 in (C) and (D), annealed to prevent the following pathways (Exon-skipping, Intra-lariat splicing, Back-splicing, or Second-splicing).
(B) Detection of intermediates and products from the ciRS-7 precursor. Total RNA from human cerebral cortex was treated with (+) or without (–) RNase R to discriminate the circularize structure from the linear structure. The purified RNA was analyzed by RT–PCR with indicated primers (arrowheads) to detect ciRS-7 precursor (P), mature ciRS-7 (C), and the spliced isoforms (M), generated by alternative splicing in the last exon of the ciRS-7 precursor. Either the lariat precursor (L1) or the looped pre-mRNA (P′) was also detected.
(C) Schematic representation of the ciRS-7 precursor and the alternatively spliced isoforms generated in (B) (see also Supplemental Figure S1). ASOs (1–6) to block specific alternative splicing are indicated with bars.
(D) The quantified data of splicing prevention using ASOs (1–6) depicted in (C). SH-SY5Y cells were electroporated with each ASO. After 24 h incubation, RT–PCR analysis was performed with ciRS-7 primers and the control GAPDH primers. The detected ciRS-7 expression levels were normalized to the control expression level of GAPDH (ciRS-7/GAPDH) and plotted as ratios to the value of control cells treated with GFP-ASO. Means ± standard deviation (SD) are given for three independent experiments (***P < 0.001).

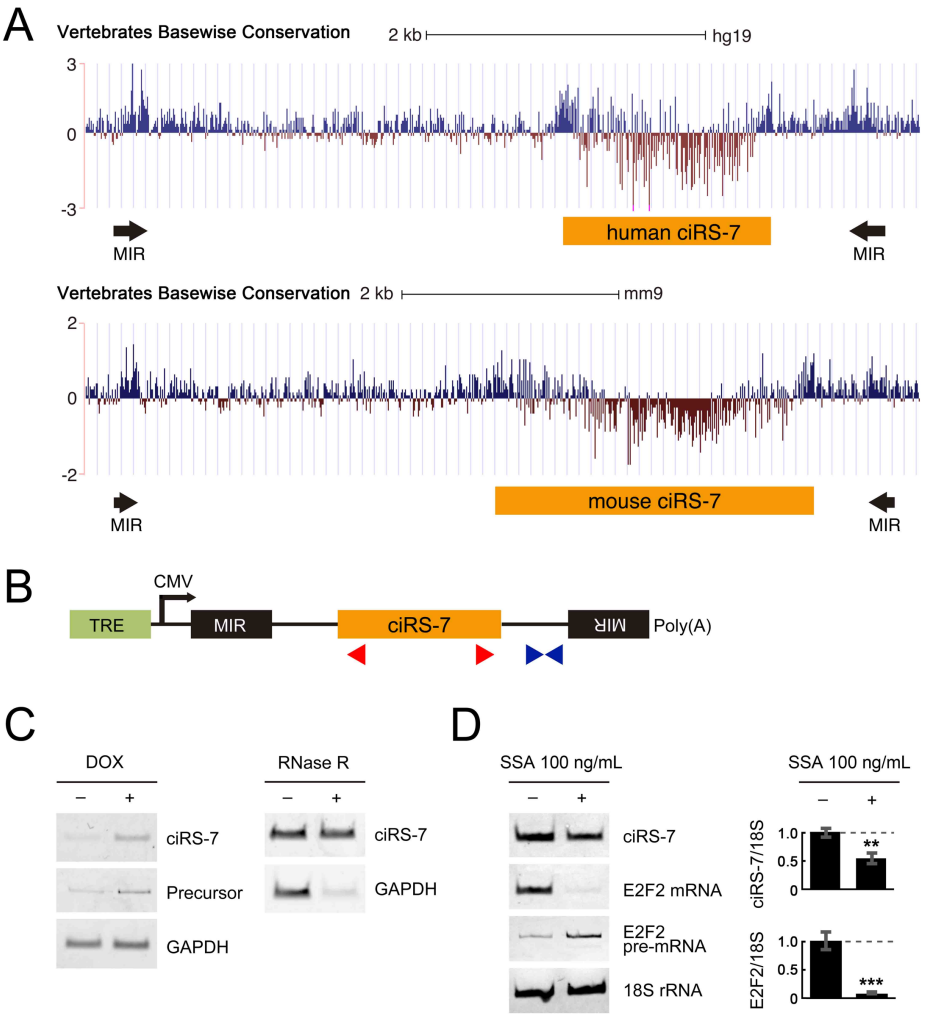

Figure 2. The production of ciRS-7 depends on the flanking inverted sequences of MIRs
(A) The ciRS-7 locus (human/mouse ciRS-7) and the identified conserved inverted MIR elements (MIRs) in the genome browser track. The vertebrate basewise conservations of human (upper) and mouse (lower) genomic sequences around the ciRS-7 precursor were displayed in the UCSC Genome Browser.
(B) Schematic structure of ciRS-7-expressing reporter plasmids. Transcription is driven by a CMV promoter and an arrow denotes the initiation site. The tetracycline response element (TRE), inverted MIR elements, ciRS-7 coding exon (ciRS-7) are indicated. Red and blue triangles indicate the positions of PCR primers to detect ciRS-7 and its precursor, respectively.
(C) The generation of ciRS-7 from reporter in stably transduced HEK293 cells. The cells were treated with DOX and extracted total RNA was analyzed by RT–PCR using primers against mature ciRS-7, precursor of ciRS-7, or a control, GAPDH (left panel). To validate the circular structure of generated ciRS-7, total RNA was also treated with RNase R prior to the RT–PCR analysis (right panel).
(D) The effect of splicing inhibitor, SSA, on the production of ciRS-7. The same stably transduced cells described in (C) were treated with DOX followed by SSA addition. After 24 h culture, extracted total RNA was analyzed by RT–PCR to detect ciRS-7. The E2F2 mRNA, as a positive control (Yoshimoto et al., 2017), was also detected (with primers against exon 3 and exon 4). The graphs show the relative values of ciRS-7 expression, normalized to the control level of 18S rRNA, and plotted as ratios to the value of control non-SSA-treated cells. Means ± SD are given for three independent experiments (**P < 0.01 for ciRS-7, ***P < 0.001 for E2F2).

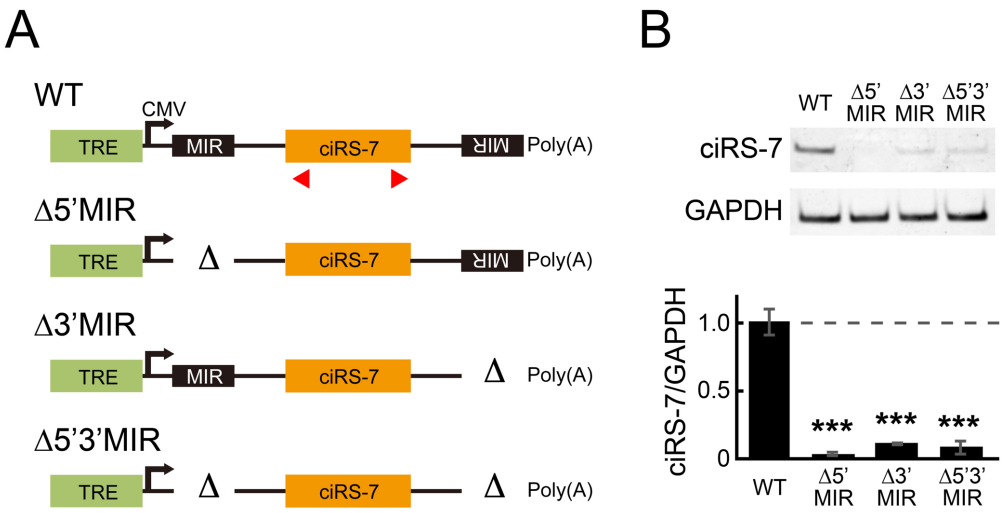

Figure 3. Deletion of either upstream or downstream MIRs aborts the production of ciRS-7
(A) Schematic structure of ciRS-7-expressing reporters with or without MIR sequence (see Figure 2B for the abbreviations). Red triangles indicate the positions of PCR primers to detect ciRS-7 in (B).
(B) Generation of ciRS-7 from reporters, with or without MIR sequences, in stably transduced HEK293 cells. Total RNA extracted from DOX-treated cells was analyzed by RT–PCR using primers against mature ciRS-7 and an internal control GAPDH. The graph shows quantification of the ciRS-7 generation. The ciRS-7 expression levels were normalized to the control expression level of GAPDH and plotted as ratios to the value of control wild-type (WT) plasmid-expressing cells. Means ± SD are given for three independent experiments (***P < 0.001).

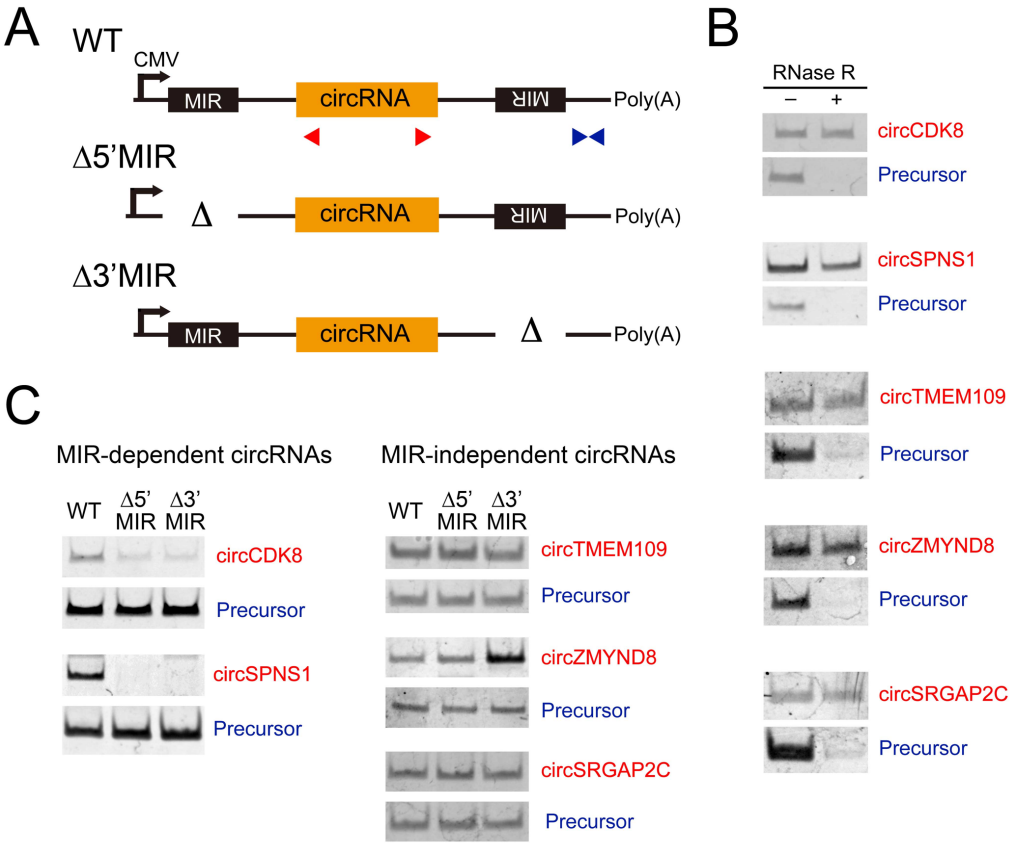

Figure 4. Several other human MIR-dependent circRNAs were identified
(A) Generalized schematic structures of three kinds of reporters, expressing wild-type (WT) or MIR-deletions (Δ5′MIR, Δ3′MIR) plasmids, which were constructed for five circRNAs. Red and blue triangles indicate the positions of PCR primers to detect these circRNA and the precursors, respectively.
(B) RT–PCR detections of indicated five circRNAs that were expressed in mouse N2A cells transfected with the plasmids in (A). Total RNA from the cells was treated with (+) or without (–) RNase R to discriminate the circularized structure from the linear structure.
(C) Identification of MIR-dependent and MIR-independent circRNAs. RT–PCR analysis of circRNAs and their precursors, which were expressed from plasmids described in (A).

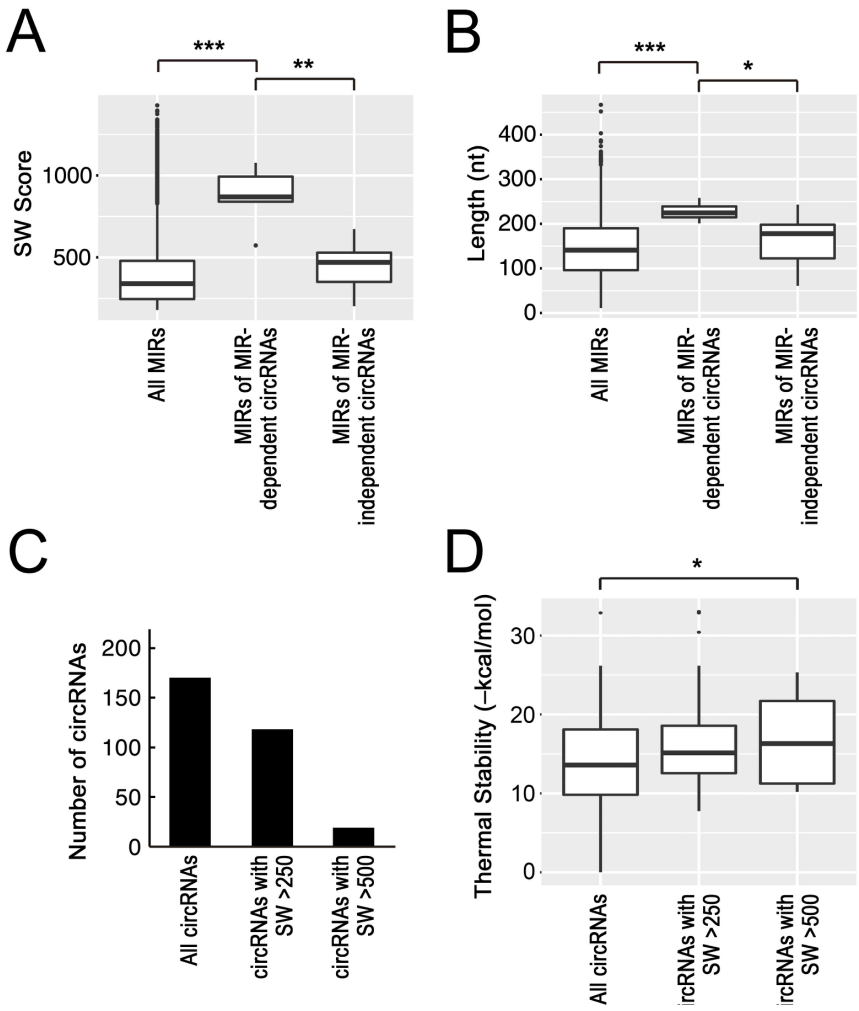

Figure 5. MIR-dependent circRNAs likely depend on stable paring between highly conserved MIRs
(A) The identified MIR-dependent circRNAs have MIRs with near-consensus features. Box plot of SW alignment scores of all reported MIRs, three MIR-dependent circRNAs, and three MIR-independent circRNAs (***P < 0.001; **P < 0.01). See Supplemental Figure S4 for each SW score of these MIRs.
(B) The identified MIR-dependent circRNAs have longer MIRs than those of MIR-independent circRNAs. Box plot of the MIR lengths of all reported MIRs, three MIR-dependent circRNAs, and three MIR-independent circRNAs (***P < 0.001; *P < 0.05).
(C) We identified 170 circRNAs (left bar) that have inverted MIRs within 1000 nt of the circRNA coding exon. Among these circRNAs, 118 circRNAs had MIRs with SW score >250 (middle bar) and 19 circRNAs had MIRs with SW score >500 (right bar).
(D) The inverted MIRs with higher SW scores have potential to form more stable MIR-MIR base-pairs. Box plot of the corrected thermal stabilities (–kcal/mol) of inverted MIRs from groups of the circRNA shown in (C), which are calculated by the RNAup program (*P < 0.05).

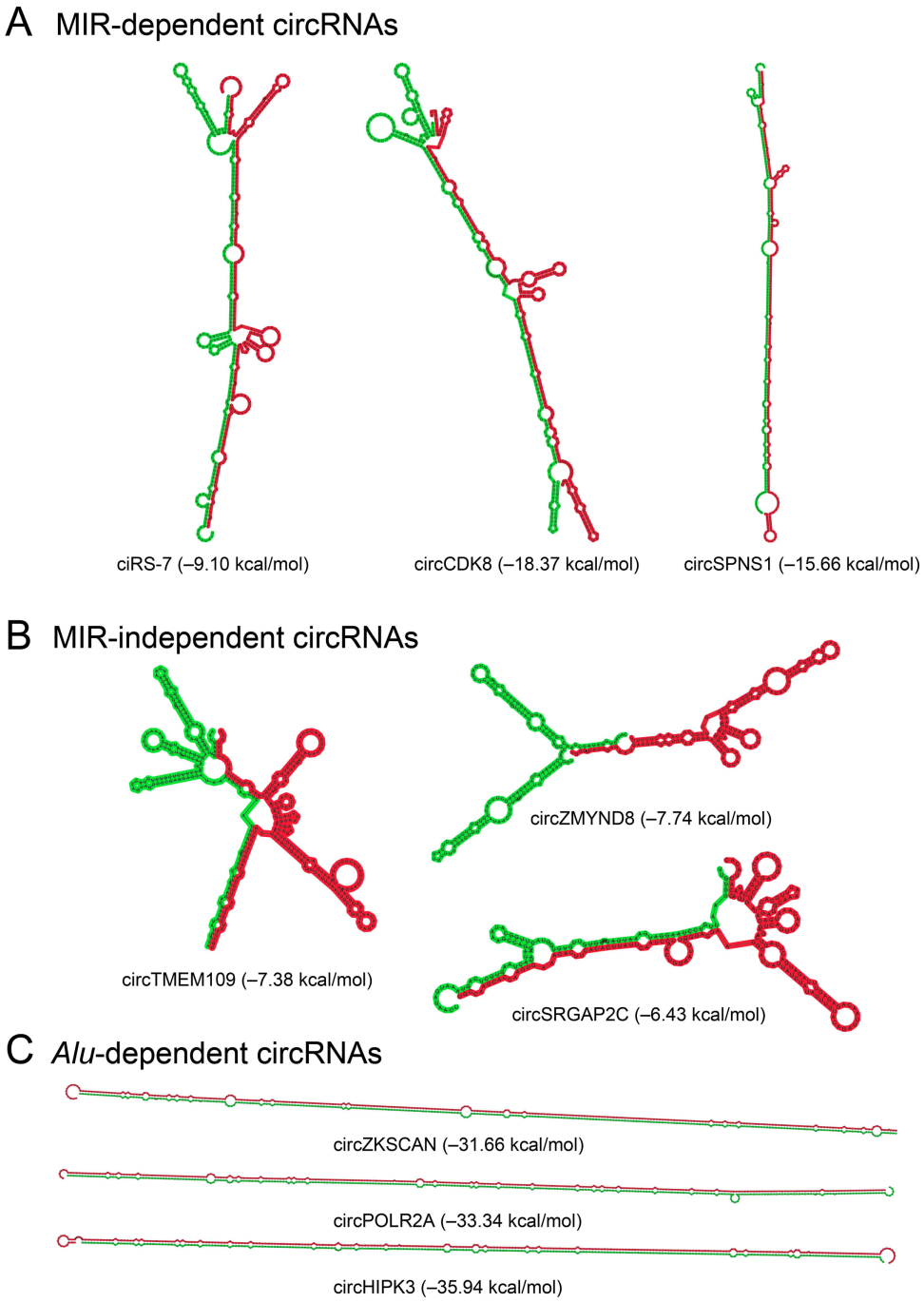

Figure 6. Inverted sequences are less complementary in MIRs than those in *Alu*s
RNA secondary structures were predicted by RNAcofold of vienna RNA package. The cofold structure of MIR-dependent circRNAs **(A)**, MIR-independent circRNAs **(B)**, and *Alu*-dependent circRNAs **(C)** are shown. The corrected thermal stability (–kcal/mol) of each MIR was calculated by the RNAup program.