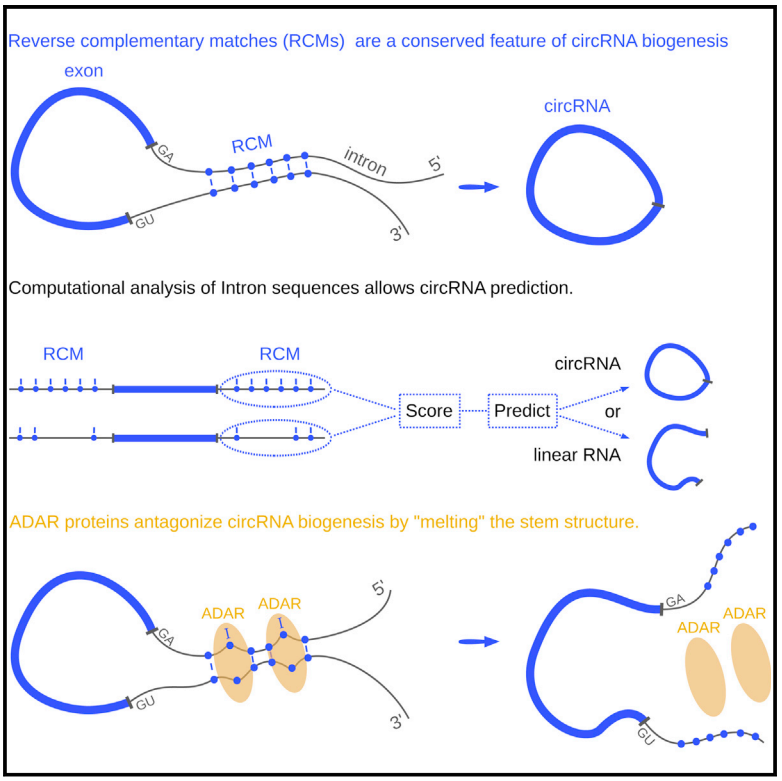
Report

Cell Reports

Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals

Graphical Abstract



Authors

Andranik Ivanov, Sebastian Memczak, ..., Christoph Dieterich, Nikolaus Rajewsky

Correspondence

rajewsky@mdc-berlin.de

In Brief

Ivanov et al. study the biogenesis of circRNAs and reveal its conserved features. Moreover, they show that circRNAs can be predicted based on the genomic sequence of their flanking introns. Finally, they demonstrate that the circRNA formation mechanism depends on the RNA-editing enzyme ADAR.

Highlights

- Reverse complementary matches are a conserved feature of circRNA biogenesis
- Computational analysis of intron sequences allows successful prediction of circRNAs
- The circRNA formation mechanism depends on the RNA-editing enzyme ADAR

Accession Numbers

GSE63823
SRP050149

Ivanov et al., 2015, Cell Reports *10*, 170–177
January 13, 2015
http://dx.doi.org/10.1016/j.celrep.2014.12.019

OPEN ACCESS
CellPress

Cell Reports
Report

Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals

Andranik Ivanov,[1] Sebastian Memczak,[1,5] Emanuel Wyler,[2,5] Francesca Torti,[1,5] Hagit T. Porath,[3] Marta R. Orejuela,[1] Michael Piechotta,[4] Erez Y. Levanon,[3] Markus Landthaler,[2] Christoph Dieterich,[4] and Nikolaus Rajewsky[1,*]

[1]Laboratory for Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical Systems Biology, Max Delbruck Center for Molecular Medicine, Robert Roessle Straße 10, 13125 Berlin-Buch, Germany

[2]Laboratory for RNA Biology and Posttranscriptional Regulation, Berlin Institute for Medical Systems Biology, Max Delbruck Center for Molecular Medicine, Robert Roessle Straße 10, 13125 Berlin-Buch, Germany

[3]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 5290002, Israel

[4]Max Planck Institute for Biology of Ageing, Cologne, Joseph Stelzmann Straße 9B, 50931 Köln, Germany

[5]These authors contributed equally to this work

*Correspondence: rajewsky@mdc-berlin.de

http://dx.doi.org/10.1016/j.celrep.2014.12.019

SUMMARY

Circular RNAs (circRNAs) are a large class of animal RNAs. To investigate possible circRNA functions, it is important to understand circRNA biogenesis. Besides human ALU repeats, sequence features that promote exon circularization are largely unknown. We experimentally identified circRNAs in *C. elegans*. Reverse complementary sequences between introns bracketing circRNAs were significantly enriched in comparison to linear controls. By scoring the presence of reverse complementary sequences in human introns, we predicted and experimentally validated circRNAs. We show that introns bracketing circRNAs are highly enriched in RNA editing or hyperediting events. Knockdown of the double-strand RNA-editing enzyme ADAR1 significantly and specifically upregulated circRNA expression. Together, our data support a model of animal circRNA biogenesis in which competing RNA-RNA interactions of introns form larger structures that promote circularization of embedded exons, whereas ADAR1 antagonizes circRNA expression by melting stems within these interactions.

INTRODUCTION

Recently, several studies have revealed that the transcriptome of animals contains many single-stranded exonic circular RNAs (circRNAs) (Jeck et al., 2013; Jeck and Sharpless, 2014; Memczak et al., 2013; Salzman et al., 2012; Wang et al., 2014). Although circRNAs have tissue- and stage-specific expression (Memczak et al., 2013), the function of circRNAs is altogether unknown. The human circRNA *CDR1as* (Hansen et al., 2011;) can act as a miRNA sponge (Hansen et al., 2013; Memczak et al., 2013). However, we and others have proposed that generally circRNAs might function in assembly of complexes, in transport, in *trans* (Memczak et al., 2013), by competing with linear splicing, or as regulators of the local concentration of RNA-binding proteins (Ashwal-Fluss et al., 2014). In vitro studies (Braun et al., 1996; Pasman et al., 1996) and a recent in vivo study (Ashwal-Fluss et al., 2014) provided evidence that circRNAs are often generated cotranscriptionally by "head-to-tail" splicing. In humans, circularized exons are typically bracketed by unusually long introns (Jeck et al., 2013). Moreover, the circRNA *SRY* (Capel et al., 1993) is bracketed by very long (~15,000 nt), almost perfectly complementary intronic sequences, and these reverse complementary matches (RCMs) are required for *SRY* circularization (Dubin et al., 1995). Therefore, it has been proposed that RCMs promote hairpin formation of the transcript. This would explain how the 5′ and 3′ ends of an exon can be in spatial proximity, perhaps thereby inducing "head-to-tail" splicing (Figure 1A). Jeck et al. (2013) reported that in humans, introns bracketing circRNAs are highly enriched in ALU repeats. The fact that ALU repeats contain RCMs supports this model; however, ALU repeats are specific to a small branch of vertebrates, and thus the widespread existence of circRNAs in other animals remains to be explained.

Caenorhabditis elegans, a well-annotated genome that is not rich in repeats, offers the possibility of identifying conserved features of circRNA biogenesis outside of vertebrates. We first sequenced RNA from several life stages of *C. elegans* and were able to boost the number of annotated exonic circRNAs from ~300 (Memczak et al., 2013) to ~1,100. Computational analysis of the bracketing introns revealed that these circRNAs are significantly enriched for RCMs. We developed a simple model for scoring circRNA biogenesis from intronic sequence analysis and asked whether our model could predict novel human circRNAs. We successfully validated the predicted human circRNAs.

The RNA-editing factor ADAR binds double-stranded RNA. Thus, the model of circRNA biogenesis predicts that circRNAs should be flanked by intronic sequences that are enriched in adenosine to inosine (A-to-I) editing. Moreover, it is known that ALU elements are edited, often by ADAR1 (Athanasiadis et al., 2004; Levanon et al., 2004; Osenberg et al., 2010; Ramaswami

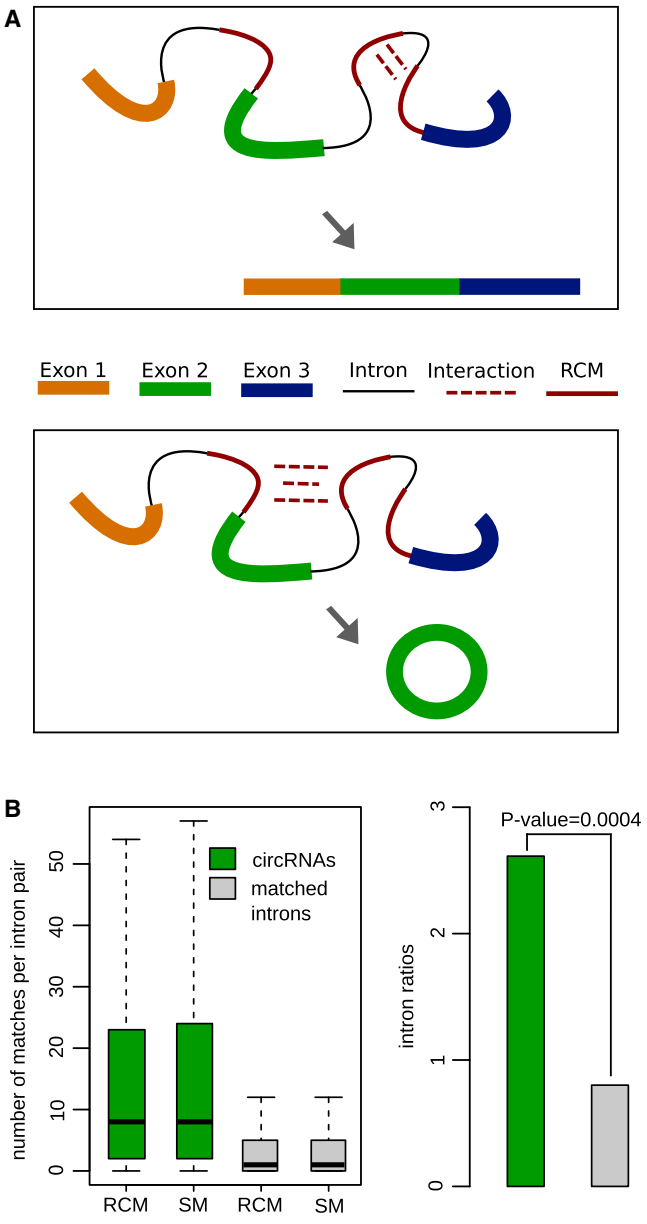


Figure 1. RCMs between or within Introns and circRNA Biogenesis

(A) Model for the possible influence of RCMs on circRNA biogenesis. RCMs between different introns (lower panel) in competition with intron internal RCMs (upper panel) may promote hairpin formation and circularization of the embedded exon.

(B) Enrichment of RCMs and SMs in introns bracketing *C. elegans* circRNAs. Left: distribution of RCM and SM counts per intron pair that flanks circRNAs (green) and length-matched controls (gray). Right: ratio of the number of intron pairs bracketing circRNAs that contain only RCMs but no SMs to the number of intron pairs bracketing circRNAs with only SMs but no RCMs (p value: Fisher's exact test).

See also Figures S1, S4, and S5.

et al., 2012). Also in *C. elegans*, long inverse repeats are enriched in A-to-I editing (Morse et al., 2002). Indeed, we discovered significant A-to-I editing, including hyperediting (Carmi et al., 2011), in introns bracketing circRNAs. To test whether ADAR1 and ADAR2 knockdown in human cells affects circRNA expression, we performed RNA sequencing (RNA-seq) and quantitative RT-PCR (qRT-PCR). We observed a significant and specific upregulation of most circRNAs, whereas their linear host transcripts were less perturbed. Thus, our data reveal a surprising function for ADAR as an antagonist of circRNA production. We discuss the implications of these findings.

RESULTS

Identification and Characterization of circRNAs in *C. elegans*

Recently, hundreds of circRNAs were reported in samples from very early developmental stages of *C. elegans* (one/two-cell embryo, oocytes, and sperm) (Memczak et al., 2013). To include circRNAs expressed in later stages, we sequenced ribosomal-depleted RNA from major life stages, including adulthood (see Experimental Procedures). We computationally identified circRNA candidates by applying our pipeline (Glažar et al., 2014; Memczak et al., 2013; http://www.circbase.org) and ensuring that the head-to-tail junctions precisely overlapped the annotated canonical splice sites (Experimental Procedures). This extended the published set of exonic circRNAs from ~300 to 1,111. Analogously to human circRNAs (Jeck et al., 2013), the circRNA flanking introns were much longer (median ~10-fold) than all of the *C. elegans* introns (Figure S1A). Therefore, we asked whether intron length alone is sufficient for circularization or additional sequence features are needed. More precisely, in *C. elegans* we tested the idea that RCMs between introns bracketing circRNAs may induce larger hairpin structures that promote the circularization of embedded exons (Figure 1A).

In *C. elegans*, Introns Bracketing circRNAs Are Highly Enriched for RCMs

For each intron pair that flanked a circRNA, we aligned the respective introns using Basic Local Alignment Search Tool (BLAST; Experimental Procedures). RCMs were strongly enriched compared with length-matched introns (Figure 1B), with a median of eight RCMs per intron pair (control pairs: one RCM). We observed the same number and significance of matches on the same strand ("sense matches" [SMs]) of intron pairs (Figure 1B). However, the exclusive occurrence of RCMs on intron pairs bracketing circRNAs was significantly enriched compared with the exclusive occurrence of SMs (Fisher's exact test, p value < 0.0004; Figures 1B and S1C). This result suggests that exon circularization is promoted by RCMs that can induce basepairing between flanking introns.

Predicting circRNAs from RCMs

We asked whether the occurrence of RCMs between circRNA flanking introns is significant enough to predict circularized exons. Therefore, we developed a simple probabilistic score *H* to capture the likelihood that introns will basepair via RCMs (Supplemental Experimental Procedures) and therefore promote hairpin formation. Competition between RCMs is naturally taken into account by this model. For each intron pair in a transcript that contained one of the 1,111 circRNAs, we computed *H*.

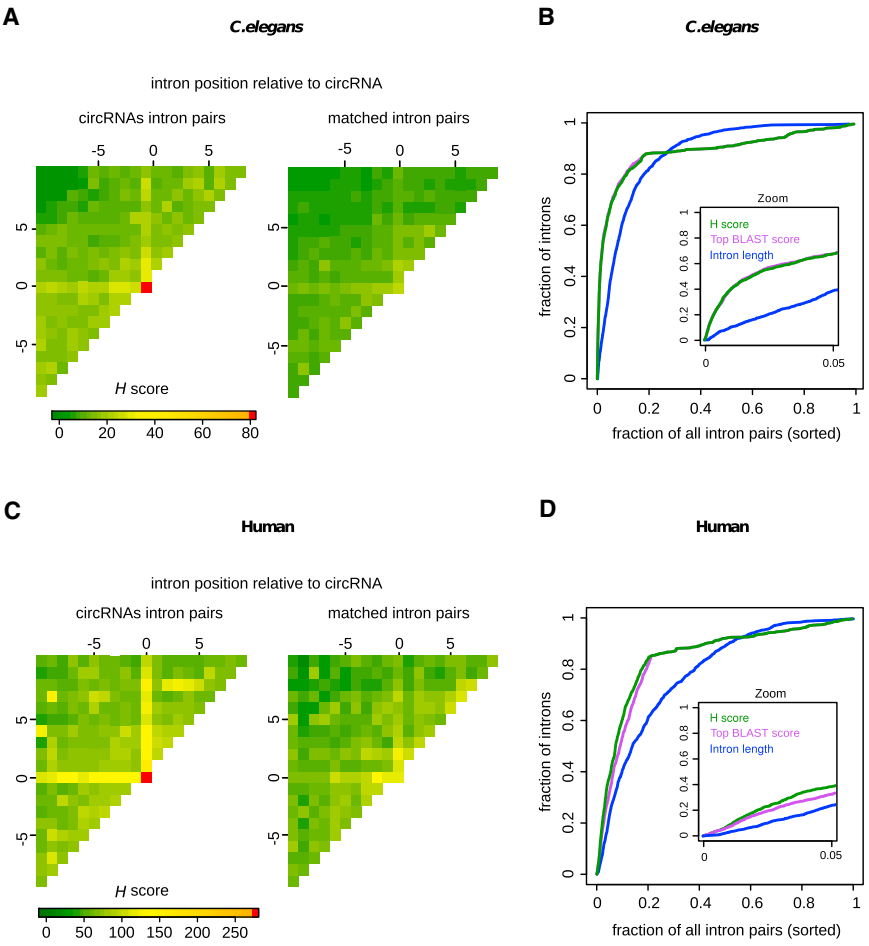


Figure 2. Prediction of circRNAs Based on Sequence Analysis of Introns

(A) *C. elegans*. The matrix represents the average circularization scores for intron pairs of circRNA producing genes. (0,0) coordinates correspond to circRNA flanking introns. Upstream and downstream introns are enumerated with decreasing and increasing numbers, respectively (with 0 being directly adjacent, −1/+1 being the preceding/next intron, etc.; see Supplemental Experimental Procedures).

(B) *C. elegans*. Different prediction methods are used. The x axis is the fraction of all introns pairs per gene sorted by *H* scores. Small (high) x values denote intron pairs with high (low) scores for bracketing a circRNA. The y axis is the fraction of 1,111 exonic circRNAs recovered by *H* score.

(C and D) Analyses for human circRNAs (Memczak et al., 2013), analogous to (A) and (B).

See also Figures S4 and S5, and Tables S1, S2, and S3.

The resulting symmetric matrix (Figure 2A; Supplemental Experimental Procedures) shows that the intron pairs flanking circularized exons stood out from all intron pairs. This was not the case for the intron length-matched control transcripts (Figure 2A). We then ranked, genome wide, all intron pairs of all transcripts annotated in *C. elegans* by *H* and calculated the cumulative fraction of known circRNAs as a function of intron-pair rank (Figure 2B). The top 4,200 (top ~1%) of *C. elegans* intron pairs precisely bracketed 430 (~38%) of the already annotated circRNAs, a highly statistically significant enrichment (hypergeometric test, p value $< 2.2 \times 10^{-16}$). Together, these results suggest that RCMs provide highly significantly improved accuracy in predicting circRNAs compared with intron length. However, we note that (for *C. elegans*) the top BLAST score of RCMs in a pair of introns yielded similar results.

To determine whether the presence of RCMs is a conserved feature of circRNA formation, we next asked whether we could predict circRNAs in human by intron sequence analysis.

The Presence of RCMs in Introns Flanking circRNAs Is a Conserved Feature of circRNA Biogenesis

For human circRNAs, we used 1,067 exons overlapping circRNAs (Memczak et al., 2013). As in *C. elegans*, RCMs were highly significantly enriched (Fisher's exact test, p value $< 4 \times 10^{-6}$; Figures S1B and S1C). Again, the number of intron pairs flanking circularized exons that contained only SM elements was much lower compared with intron pairs containing only RCMs (Figure S1C). The median length of human RCMs was 20-fold higher than that of *C. elegans* (Figures S1D and S1E). We found that 88% of the top-scoring RCMs overlapped with ALU elements, 4.5% of RCMs overlapped with L1/L2 repeats, and the remaining RCMs did not overlap any annotated repetitive element and had the highest sequence conservation (Figure S1F).

Further, we analyzed circRNA biogenesis conservation between mouse and human. We defined 71 circRNAs that are circularized in human and mouse ("conserved circular expression," Figure S1G). Since repetitive sequences such as ALU elements are rapidly evolving, it is difficult to align them between species. To circumvent this problem, we computed *H* scores by independently scoring the human and mouse introns (Figure S1G). As controls, we defined (1) circRNAs that are circularized in humans but have not been annotated as circular in mouse ("nonconserved circRNAs"), and (2) randomly selected exons that have a similar bracketing intron length as human circRNAs, and have positive *H* scores. Mouse *H* scores for conserved circRNAs were significantly higher (p value $< 2.2 \times 10^{-16}$) compared with the scores for exons that circularize only in human or the length-matched controls (Figure S1G). Thus, the set of circRNAs that are conserved between mouse and human are also conserved in their biogenesis as quantified by *H*.

Human circRNAs Can Be Predicted Based on the Sequence Composition of Their Flanking Introns

To predict circRNAs on a genome-wide level, we computed *H* for all possible intron pairs in the UCSC RefSeq database

(Experimental Procedures). As a control, we used transcripts with introns matched in length (Figure 2C; Supplemental Experimental Procedures) or predictions using the intron length or simply the top BLAST score between intron pairs (Figure 2D). For high-ranking intron pairs, the enrichment in circRNAs was highest when the *H* score was used, followed by BLAST and intron length. The top 20,000 (~1% of all) predictions comprised 610 (~9%) exonic circRNAs cataloged in the Memczak et al. (2013), Jeck et al. (2013), and Zhang et al. (2014) data sets, a highly significant enrichment (hypergeometric test, p value $< 2.2 \times 10^{-16}$). Predictions based on sequences associated with ALU repeats yielded similar results (data not shown), showing that in humans, ALU elements have likely contributed dominantly to circRNA formation (Experimental Procedures). We note that the number of false positives is very difficult to estimate because many human circRNAs have not yet been reported. For example, when we restricted the circRNA predictions to well-expressed mRNAs in HEK293 cells, the success rate of our predictions was much higher (Figure S1H). Therefore, experimental validation of predictions seems to be a better way to estimate false-positive rates.

Predicted circRNAs Are Experimentally Validated in HEK293 Cells

Since the circRNA predictions were not informed by expression data, we expected that we could only validate circRNAs that are isoforms of expressed transcripts. To test our predictions, we considered the top 1% (*H* ranked) predicted circRNAs and grouped them into three bins based on the expression of the linear host transcripts in HEK293 cells (top 1% expressed, medium, and bottom 50%). For experimental validation, we randomly selected six, ten, and five circRNAs from the high-, medium-, and low-expression bins, respectively. As negative controls, we used three well-expressed linear mRNAs and two exons that are well expressed in HEK293 cells and have flanking introns but an *H* score of 0. The linear controls were not circularized (RNase R negative) and the two exons with an *H* score of 0 could not be amplified with divergent primers. However, 12 out of 16 circRNAs with top *H* scores and high/medium expression of the host transcripts were (1) resistant to RNase R and (2) had the predicted head-to-tail junctions, as validated by Sanger sequencing (Experimental Procedures; Figure 3), suggesting that these 12 circRNAs exist in circularized form in HEK293 cells. Five candidates with decent *H* scores that we could not confirm were, as expected, from the third, low-expression bin. The predicted head-to-tail splicing of ten (out of 12) of the RNase R-resistant circRNAs was validated by Sanger sequencing (Experimental Procedures). We found that two circRNA candidates were circularized (by RNase R assay) but had an additional exon incorporated (as observed in Sanger sequencing; Figure 3A). We note that five of the tested circRNAs had reasonable expression (1%–10% of *VCL* expression; Experimental Procedures).

RNA Editing of Introns Flanking circRNAs

ADAR is a highly conserved RNA-editing enzyme that binds double-stranded RNA (Nishikura, 2010) and deaminates adenosine bases to inosine. In humans, ADAR1 and ADAR2 interact with double-stranded ALU repeats (Athanasiadis et al., 2004; Levanon et al., 2004; Nishikura, 2010; Ramaswami et al., 2012). We compared A-to-I conversions (Ramaswami and Li, 2014) in 1,500 bp regions flanking circRNA splice sites with (1) other splice sites in transcripts that produce circRNAs, and (2) length-matched introns (Figure 4A). A-to-I conversions nearby circRNA splice sites were enriched compared with controls. In general, ALU repeats in circRNAs flanking introns were edited significantly higher compared with expression- and length-matched controls (Figures S2A and S2B). Since A-to-I editing is a hallmark of basepaired RNA, we asked whether RCMs are preferentially located at sites of editing. We compared the position of the RCMs (defined as the nearest RCMs that match between the pair of introns bracketing circRNA) with the same controls as before (Figure 4B). These results suggest that indeed A-to-I editing preferentially occurs at regions that are basepaired and proximal (upstream and downstream 200–600 nt) to the splice sites of circularized exons.

To test whether ADAR proteins are involved in circRNA biogenesis, we codepleted ADAR1 and ADAR2 in HEK293 cells using RNAi (Supplemental Experimental Procedures). We used two controls: untreated total RNA and codepletion of three proteins of the APOBEC family (APOBEC3B, APOBEC3C, and APOBEC3F), which are known to bind mRNAs (Baltz et al., 2012). APOBEC enzymes are known to edit single-stranded DNA or RNA (Vasudevan et al., 2013), but not double-stranded RNA. The efficacy of the different knockdowns (KDs) was validated by western blotting and qRT-PCR (Figure S2C). Total RNA extracted from the different experiments was depleted from rRNAs and sequenced (Supplemental Experimental Procedures). We reproducibly observed that ADAR depletion resulted in significantly higher (p value $< 2.2 \times 10^{-16}$) circRNA expression compared with controls (Figure S2D), whereas the linear host transcripts were less strongly affected (Figure 4C). For example, 84 circRNAs and 11 linear hosts were upregulated more than 2-fold (Fisher's exact test, p value 5.7×10^{-13}). This effect was seen in independent biological replicates, as well as in a comparison of ADAR1/2 KD with APOBEC3 KD (Figures S2E and S2F).

We set out to validate these observations in independently carried out ADAR1 and ADAR2 KD experiments (using the previous and an independent small interfering RNAs [siRNAs] against ADAR1) followed by qRT-PCR assays (Experimental Procedures; Figures 4D, 4E, and S2G). CircRNA candidates were selected based on the increased fold changes observed in sequencing data sets. circRNA expression was compared with the expression of the respective linear host gene. Four out of eight circRNAs (*UBAC2*, *SMARCA*, *SPECC1*, and *HIPK2*) were upregulated upon ADAR1 depletion, whereas the linear host RNAs did not show consistent expression changes. *PUM1* and *CREBBP* circRNA were upregulated to the same level as their linear host transcripts. Consistent with sequencing data, *CDR1as* was expressed 4-fold higher in ADAR1 KD samples compared with control. Two circRNAs (*GAPVD1* and *PDS5B*) did not show the expression changes observed in the sequencing data (Figures 4E and S2G). Based on these findings, we conclude that ADAR1 depletion can induce upregulation of circRNAs independently of the expression level of the linear host circRNA.

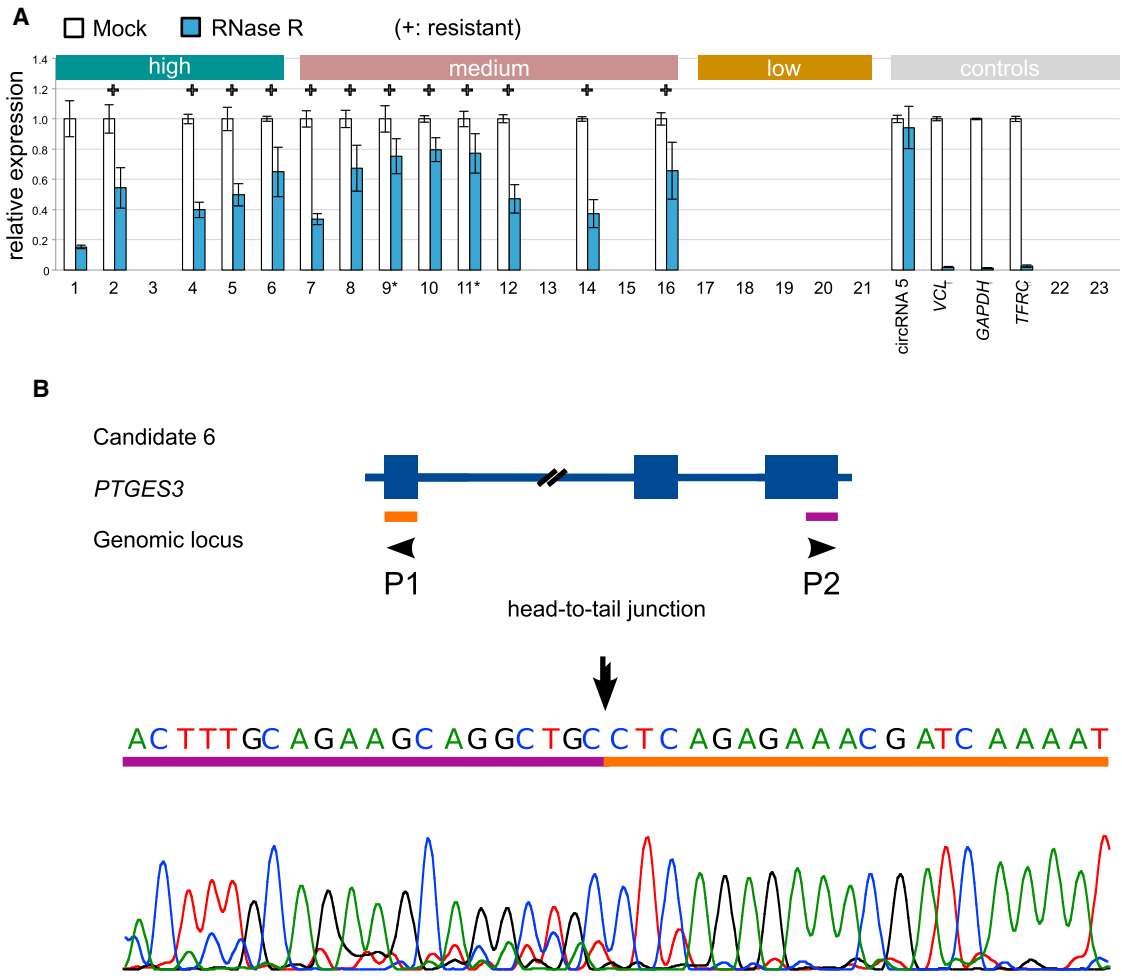


Figure 3. Experimental Validation of Human circRNAs Predicted by RCM Analyses

(A) CircRNA candidates from high-, medium-, and low-expression sets were assayed by qPCR with divergent primers and RNase R treatment. Linear control: VCL, GAPDH, TFRC; positive control: a known circRNA (Memczak et al., 2013). Sanger sequencing of amplicons confirmed in all tested cases the predicted head-to-tail junctions (candidates 9 and 11 contained an additional exon: marked with *). CircRNA candidates that were >10-fold resistant to RNase R treatment compared with Vinculin were counted as positive (+). As negative controls, we selected exons from highly expressed genes with *H* score = 0. Error bars, SEM; n = 4.

(B) The chromatogram of a Sanger sequencing experiment confirms the presence of the predicted head-to-tail junction of the candidate circRNA from the *PTGES3* gene locus.

See also Figure S5 and Tables S1, S2, and S3.

To explore whether circRNA flanking introns undergo extensive hyperediting (Carmi et al., 2011), we applied a computational pipeline that detects hyperediting events in the RNA sequencing data sets used by Memczak et al. (2013) to predict circRNAs, as well as in ADAR KD and control data sets. This analysis (for details, see Porath et al. 2014) identified ~165,000 unique hyperediting sites in the human genome (~72.000, ~49.000, ~45.000, and ~11.000 in Memczak et al. [2013] and two control and ADAR KD samples, respectively). Notably, the number of hyperedited sites identified in the ADAR KD sample was 4- to 5-fold smaller compared with controls. We found that 25% (19%) of circRNA upstream (downstream) introns had at least one conversion, whereas only 6% of other introns from the same genes had conversions. The number of conversions per base was also found to be elevated for the upstream (downstream) introns (Figures S3A–S3C). A similar highly significant enrichment of hyperediting events was detected for *C. elegans* circRNA upstream introns (Supplemental Experimental Procedures; Figure S3D). Together, our data suggest that A-to-I editing by ADAR is a "universal" hallmark of circRNA biogenesis in animals.

DISCUSSION

Our analyses support a model in which RCMs between introns that bracket an exon promote the circularization of that exon. We note that this model is powerful enough to enable us to successfully predict and experimentally validate circRNAs. Our data suggest that this model of circRNA biogenesis is conserved across animals.

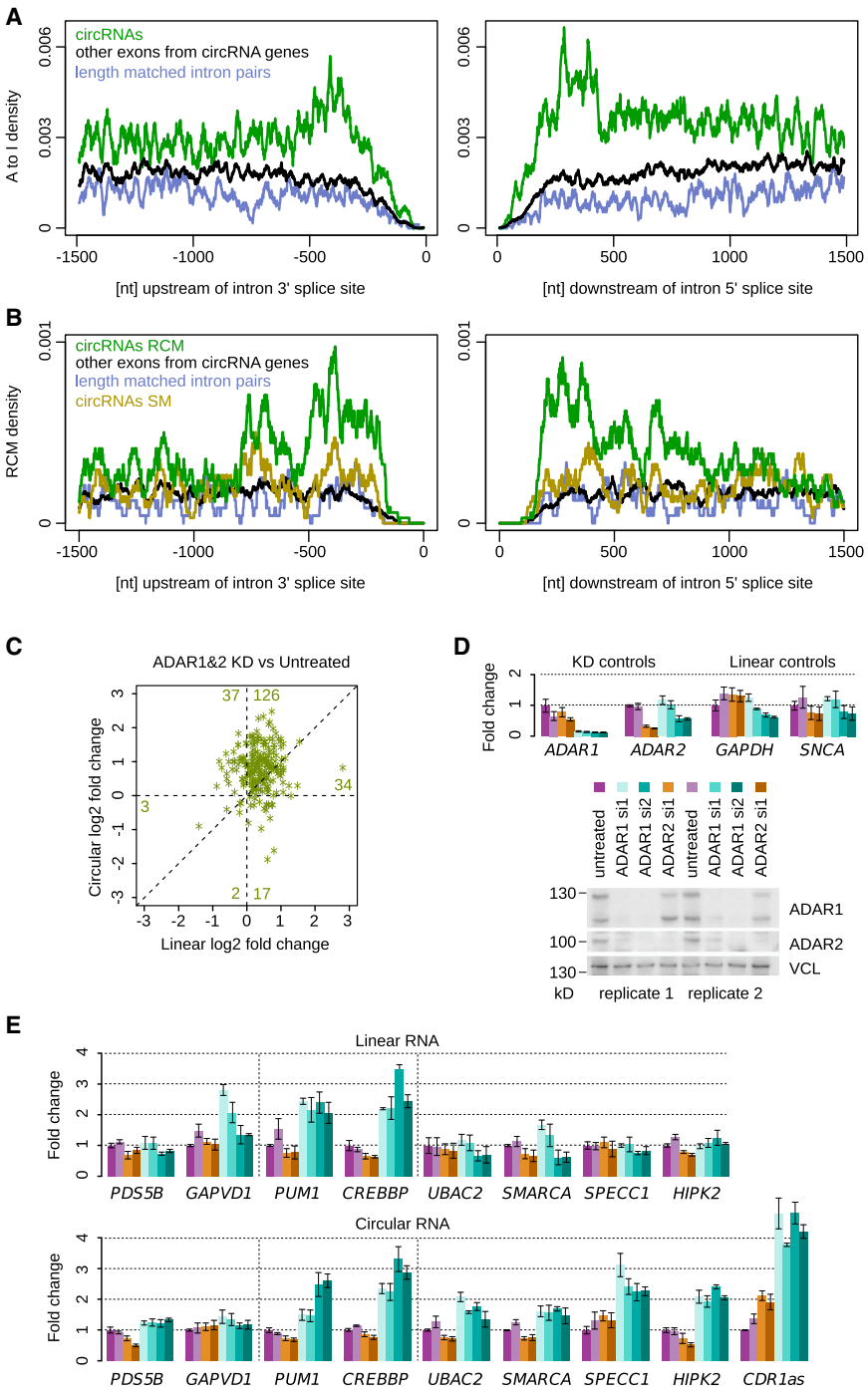


Figure 4. ADAR Antagonizes circRNA Expression

(A) Normalized distribution of A-to-I conversions upstream/downstream of the head/tail splice sites of circRNAs (Memczak et al., 2013). As controls, we selected length-matched introns and introns from the same genes that produce circRNAs. Introns smaller than 1.5 kb were removed from the analysis. The curves were smoothed within ±10 bp at each position and normalized to the total number of analyzed introns.

(B) Position of the nearest RCMs (or SMs) around circRNA splice sites. For each intron pair, we selected the top-scoring RCMs within the 1.5 kb region around the splice site. The curves were smoothed within a ±20 bp window and normalized to the total number of analyzed intron pairs.

(C) Comparison of the differential expression (ADAR1 and ADAR2 codepletion against untreated control) of linear RNA and circRNA in two independent replicates (merged). The y axis is the log2 fold change of the reads (+5) supporting the head-to-tail splice sites. The x axis is log2 fold change of the reads supporting linear splice sites of circRNA host genes. For the analysis, we selected only circRNAs with at least ten reads in ADAR1/2 KD or control samples. The numbers tally the count of circRNAs/corresponding linear hosts in each segment of the graph.

(D) Depletion of ADAR1 by RNAi. ADAR1 was depleted from HEK293 cells for 96 hr using two different siRNAs and compared with untreated and ADAR2-depleted cells. Upper panel: qRT-PCR of *ADAR1* and *ADAR2* transcripts, as well as *GAPDH* and *SNCA* as controls. Lower panel: western blot using ADAR1- and ADAR2-specific antibodies. Error bars, SD; n = 3.

(E) Quantification of circRNAs and host transcript levels. Using circRNA-specific primers (upper panel) or exon-spanning primers that were not part of the circRNA (lower panel), we quantified RNA abundances relative to untreated cells by qRT-PCR (Delta-Delta Ct values were normalized to *C. elegans* spike-in). Error bars, SD; n = 3.

See also Figures S2, S3, and S5, and Tables S1, S2, and S3.

An immediate follow-up question is, do RCMs have sequence specificity? In general, our data do not support sequence specificity in *C. elegans*, since the most prevalent motif in RCMs between introns bracketing circRNAs was clearly present in only 11% of all such RCMs. The equivalent motif analysis in humans yielded (perhaps as expected) the ALU element. For mouse circRNAs, we also found an ALU-like element. Sequence alignment between these motifs (Figure S4) suggests that there may be an ancient RNA sequence that promotes circularization. This motif awaits experimental testing.

RCMs can be generated by simple random matches under neutral evolution. One possible scenario is that relatively quickly evolving RCMs constantly generate a pool of circRNAs that can subsequently be selected for and fixed. Finding circRNAs that are under negative selection seems possible because, as we have shown, RCMs between introns can be analyzed across species. Thus, we think that cross-species comparisons of competing RCMs will provide a way to find circRNAs that may be functionally important. Indeed, our results show that our *H* score is already able to link conserved circular expression to conserved biogenesis.

In humans, 88% of circRNAs have ALU repeats in their flanking introns, which, as we have shown, likely promote circularization by RCMs. Therefore, it is interesting to speculate about the possible functions of circRNAs, since ALU repeats have expanded relatively recently in vertebrate evolution. We noticed that in fly, genes with neuronal functions often have long introns and express circRNAs at relatively high levels (Ashwal-Fluss et al., 2014). This also holds true for human brain tissues, where dozens of circRNAs appear to be more highly expressed than their (well-expressed) linear hosts (N.R., unpublished data). For human circRNAs, we found that A-to-I editing events had a tendency to occur at intronic positions that were proximal (upstream and downstream 200–600 nt) to the splice sites of circularized exons. These results link A-to-I editing to circRNA biogenesis and predict that A-to-I editing events "melt" the stems that are formed across introns that bracket a circRNA. Our KD experiments showed that indeed ADAR1 antagonizes circRNA biogenesis. The upregulation of circRNAs was stronger compared with upregulation of their linear host transcripts, suggesting that ADAR1 has a specific effect on circRNA biogenesis. However, we cannot rule out that indirect effects explain upregulation of circRNAs upon ADAR1 KD.

It is very interesting to think about the implications of these findings. As a class, circRNAs may become more important in systems where ADAR1 expression is temporarily low. For example, ADAR1 expression decreases in human embryonic stem cells that are differentiating into the neuronal lineage (Osenberg et al., 2010). Since circRNAs are unusually stable, this might be a mechanism to generate a long-term "memory" of past states. However, one should not forget that other factors, such as muscleblind, have been shown to promote the biogenesis of circRNAs (Ashwal-Fluss et al., 2014). Therefore, circRNAs may also be well expressed in systems in which ADAR1 expression is high. Finally, it has been shown that circular splicing and linear splicing can compete with each other (Ashwal-Fluss et al., 2014). Thus, it is possible that regulation of circRNA biogenesis by ADAR1 serves as a mechanism to regulate expression of the linear isoforms.

Note: while this paper was under review, two studies were published describing RCM-dependent circularization in human cells (Liang and Wilusz, 2014; Zhang et al., 2014). We also acknowledge Starke et al. (2014), which presents related content in this issue of *Cell Reports*.

CONCLUSIONS

In summary, in this work, we explored a specific model of circRNA biogenesis and showed that this model seems to be conserved across animals and is powerful enough to successfully predict circRNAs. Our data also link RNA editing to circRNA biogenesis and suggest a function for ADAR1. These results will enable the future detection and understanding of possible circRNA functions, particularly in neuronal tissues.

EXPERIMENTAL PROCEDURES

Identification of *C. elegans* circRNAs

We first mapped worm sequencing data to rRNA to reduce the ribosomal reads. The remaining reads were analyzed as described in Memczak et al. (2013) with the additional filtering step of requiring unique alignments for both of the read anchors prior to extension. Thereafter, we retained circRNAs that overlap annotated internal splice sites.

Intron Alignments

We carried out intron alignments using BLAST (Altschul et al., 1990) with the parameters "-task blastn -word_size 6" for *C. elegans* and "-task blastn -word_size 11" for human. For further analysis, we considered only alignments that exceeded BLAST score cutoffs of 20 and 100 for *C. elegans* and human, respectively. For each intron pair, we calculated circularization *H* as a top BLAST score multiplied by the probability of forming at least one stem around the exon (Supplemental Experimental Procedures).

Nonrepetitive RCMs were defined as BLAST matches that did not overlap with any sequence present in UCSC RepeatMasker. Putative circRNAs with nonrepetitive RCMs are shown in Table S3.

Conservation Analysis

To determine homologous exon groups, we used UCSC liftOver with minMatch = 0.1 option. We found that 71 out of 1,067 human circRNAs (Memczak et al., 2013) overlapped annotated splice sites and were present in circular form in mouse. We compared *H* score distributions using the Mann-Whitney two-sided test.

qRT-PCR

We performed qPCR using the Maxima SYBR-Green/ROX qPCR master mix (Thermo Scientific) and a StepOnePlus PCR system (Applied Biosystems). To detect putative head-to-tail junctions, we designed divergent primers for each circRNA candidate. Ct values for mock/RNase R-treated circRNAs were normalized to *C. elegans* spike-in RNA (for standard curves, see Figures S5A–S5C and Table S2). Amplicons were gel or bead purified (Zymoclean gel DNA recovery kit [Zymo Research]; Agencourt AMPure XP [Beckman Coulter]) and subjected to Sanger sequencing by LGC Genomics. Confirmed head-to-tail junctions are available in Figure S5D. A list of the oligos is given in Table S1.

For more details, see the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The data have been deposited in the NCBI Gene Expression Omnibus and are available under accession numbers GSE63823 and SRP050149. circRNAs identified in this study, as well as *H* scores, are available at http://www.circbase.org.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.12.019.

AUTHOR CONTRIBUTIONS

A.I. performed all computational analyses except hyperediting. S.M. contributed experimental validation assays. E.W. designed and carried out ADAR RNAi, RNA-seq, and qPCR experiments and initial analysis of the RNA-seq data. H.P. and E.L. performed hyperediting detection. F.T. collected and prepared worm samples, carried out initial ADAR KDs, and helped in the design of the validation experiments. M.O. annotated worm circRNAs. M.P. performed an initial analysis of the RNA-seq data and was supervised by C.D. A.I. and N.R. analyzed the data and wrote the paper with input from E.L.

ACKNOWLEDGMENTS

A.I. and N.R. thank Sebastian Kadener (Hebrew University), Albrecht Bindereif (University of Giessen), and Marvin Jens (N.R. lab) for helpful discussions. We thank all members of the N.R. lab for discussions and support. This work was

supported by German-Israeli-Foundation for Scientific Research and Development (G.I.F) and German Ministry for Education and Research (SatNet program).

Received: September 16, 2014
Revised: November 25, 2014
Accepted: December 9, 2014
Published: December 31, 2014

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

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.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. *2*, e391.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

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.

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.

Carmi, S., Borukhov, I., and Levanon, E.Y. (2011). Identification of widespread ultra-edited human RNAs. PLoS Genet. *7*, e1002317.

Dubin, R.A., Kazmi, M.A., and Ostrer, H. (1995). Inverted repeats are necessary for circularization of the mouse testis Sry transcript. Gene *167*, 245–248.

Glažar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. RNA *20*, 1666–1670.

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.

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.

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.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., et al. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. *22*, 1001–1005.

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

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.

Morse, D.P., Aruscavage, P.J., and Bass, B.L. (2002). RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA, *Volume 99* (USA: Proc. Natl. Acad. Sci), pp. 7906–7911.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. *79*, 321–349.

Osenberg, S., Paz Yaacov, N., Safran, M., Moshkovitz, S., Shtrichman, R., Sherf, O., Jacob-Hirsch, J., Keshet, G., Amariglio, N., Itskovitz-Eldor, J., and Rechavi, G. (2010). Alu sequences in undifferentiated human embryonic stem cells display high levels of A-to-I RNA editing. PLoS ONE *5*, e11173.

Pasman, Z., Been, M.D., and Garcia-Blanco, M.A. (1996). Exon circularization in mammalian nuclear extracts. RNA *2*, 603–610.

Porath, H.T., Carmi, S., and Levanon, E.Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. Nat. Commun. *5*, 4726.

Ramaswami, G., and Li, J.B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res. *42*, D109–D113.

Ramaswami, G., Lin, W., Piskol, R., Tan, M.H., Davis, C., and Li, J.B. (2012). Accurate identification of human Alu and non-Alu RNA editing sites. Nat. Methods *9*, 579–581.

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.

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.-H., and Bindereif, A. (2014). Exon circularization requires canonical splice signals. Cel Rep *10*, this issue.

Vasudevan, A.A.J., Smits, S.H.J., Höppner, A., Häussinger, D., Koenig, B.W., and Münk, C. (2013). Structural features of antiviral DNA cytidine deaminases. Biol. Chem. *394*, 1357–1370.

Wang, P.L., Bao, Y., Yee, M.-C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O., and Salzman, J. (2014). Circular RNA is expressed across the eukaryotic tree of life. PLoS ONE *9*, e90859.

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.