Reverse complementary matches simultaneously promote both back-splicing and exon-skipping

Dong Cao[1]

1 Information Processing Biology Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Kunigami, Okinawa, Japan, 904-0495

Correspondence: dong.cao@oist.jp

Summary

Circular RNAs (circRNAs) play diverse roles in different biological and physiological environments. Understanding circRNA expression profile and how circRNAs are regulated help make better use of them. Here, using large-scale neuron isolation from the first larval stage (L1) of *Caenorhabditis elegans* (*C. elegans*) followed by whole-transcriptome RNA sequencing (RNA-seq), we provide the first neuronal circRNA data in *C. elegans*. We show that circRNAs are highly expressed in the neurons of *C. elegans* and are preferably derived from neuronal genes. More importantly, reverse complementary matches (RCMs) in circRNA-flanking introns are not only required for back-splicing but also promote the skipping of exon(s) to be circularized. Finally, through one-by-one mutagenesis of all the splicing sites and branch points required for exon-skipping and back-splicing, we show that exon-skipping is not absolutely required for back-splicing, neither the other way. Instead, the coupled exon-skipping and back-splicing are happening at the same time and the two pathways work together to regulate the levels of the circular and the skipped transcripts.

Keywords

circular RNA; back-splicing; exon-skipping; reverse complementary matches; neuron sorting

Introduction

cicRNAs are regulatory RNA molecules that are covalently closed by back-splicing, in which a downstream splice donor is joined with an upstream splice acceptor (Chen, 2020). circRNAs have been reported to bind to microRNAs (the so-called "miRNA sponges") (Hansen et al., 2013; Memczak et al., 2013) and to act as decoys for proteins to regulate the expression and function of genes (Du et al., 2020; Xia et al., 2018; Zhu et al., 2019). Some circRNAs are translated to functional proteins/peptides through cap-independent mechanisms (Legnini et al., 2017; Pamudurti et al., 2017; Yang et al., 2017; Zhang et al., 2018a; Zhang et al., 2018b; Zheng et al., 2019). circRNAs are always expressed tissue-specifically. Especially, circRNAs are enriched in the brain of several species (Gruner et al., 2016; Rybak-Wolf et al., 2015; Westholm et al., 2014; You et al., 2015), like humans, mice, *Drosophila*, etc. In *C. elegans*, circRNAs are also identified (Ivanov et al., 2015; Memczak et al., 2013) and accumulate during aging (Cortes-Lopez et al., 2018), but such neuronal circRNA profile has not yet been reported.

Back-splicing is a well-regulated process (Chen, 2020). The reverse complementary matches (RCMs) that locate in the flanking introns of circRNA-producing exons promote circRNA formation by bringing the splice sites for back-splicing to proximity. This model was first proposed when a circular transcript was identified in *Sry* in mice, in which a pair of > 15,500 nt RCMs are present in the introns flanking the ~1,200 nt exon to be circularized (Capel et al., 1993; Gubbay et al., 1992). Subsequent *in vitro* experiments showed that as less as 400 nt of complementary sequences are sufficient enough for the production of *circSry* (Dubin et al., 1995). Genome-wide analysis of RNA-seq data in humans revealed that Alu repeats, which contain RCMs, are enriched in the flanking introns of circRNAs (Jeck et al., 2013). Zhang et al proved that the orientation-opposite Alu elements as well as other non-repetitive RCMs sufficiently promote circRNA formation in cultured human cells (Zhang et al., 2014). More importantly, disturbance of RCMs is shown to be an efficient method for circRNA knockout with little effect on cognate linear mRNA (Xia et al., 2018; Zhang et al., 2016). In *C. elegans*, RCMs are abundant in circRNA-flanking introns (Cortes-Lopez et al., 2018; Ivanov et al., 2015), but their roles in circRNA production have not been experimentally tested.

circRNA is found to be correlated to exon-skipping (Barrett et al., 2015; Jeck et al., 2013; Kelly et al., 2015; Surono et al., 1999; Zaphiropoulos, 1996, 1997). In early years, sporadic examples showed that some circRNA-producing genes have transcripts that skip the exons to be circularized (Surono et al., 1999; Zaphiropoulos, 1996, 1997). Later, systematic analysis of RNA-seq data in human cells found a global correlation of exon-skipping with exon circularization (Kelly et al., 2015). In *Schizosaccharomyces pombe*, Barrett et al showed that circRNA can be produced from an exon-containing lariat intermediate produced by exon-skipping (Barrett et al., 2015). Given that the correlated exon-skipping and circRNA formation use the same pair of introns, it is possible that RCMs in circRNA-flanking introns also regulate exon-skipping.

Here, using *C. elegans* as the model organism, we provide the first circRNA profile in its neurons by large-scale neuron isolation followed by whole-transcriptome RNA-seq. Hundreds of novel circRNAs were annotated, mainly from the sorted neuron samples. As in other organisms, circRNAs are also abundant in the neurons of *C. elegans*. Dozens of neuron-enriched and neuron-

depleted circRNAs are identified. Interestingly, the differentially expressed circRNAs show strong correlation to their cognate linear mRNAs. By deleting one of the RCMs in multiple circRNA genes, we experimentally validated that RCMs are required for circRNA formation in *C. elegans*. Moreover, RCMs also promote the correlated exon-skipping. Through *in vivo* one-by-one mutagenesis of splicing sites and branch points of exon-skipping and back-splicing, we show that RCM-promoted back-splicing and RCM-promoted exon-skipping are happening together to produce the circular and the skipped transcripts.

Results

1. Successful neuron isolation for circRNA detection by RNA-seq

Currently, there is no available neuronal circRNA data in *C. elegans*, mainly due to challenges in obtaining enough neuron samples from the tiny worms which have no obvious compartmentalized "brain" tissue. The most common method to obtain neuron cells from *C. elegans* is by the "labeling-dissociation-sorting" method (Figure 1A) (Christensen et al., 2002; Deffit et al., 2017; Fox et al., 2005; Kaletsky et al., 2016; Spencer et al., 2014; Zhang et al., 2011; Zhang et al., 2002), in which target neurons are labeled by fluorescent protein and, after mild dissociation of the worms, labeled neurons are collected by fluorescence activated cell sorting (FACS). This method can obtain target neurons in high purity and is used for the detection of gene expression in single neurons to all the neurons. However, due to the low efficiency of the dissociation (Spencer et al., 2014; Zhang et al., 2011), total RNA obtained from sorted cells is limited, hence this method is only used for mRNA detection, either by microarray or RNA-seq (Deffit et al., 2017; Kaletsky et al., 2016; Spencer et al., 2014; Von Stetina et al., 2007; Watson et al., 2008). By optimization of previous protocols (Zhang et al., 2011), we aim to improve the final total RNA yield to hundreds of nanograms for circRNA detection by whole-transcriptome RNA-seq.

Here, using a strain (NW1229) with pan-neuronal green fluorescent protein (GFP) expression, we found that by shortening the time of SDS-DTT treatment (from 2 min to 1.5 min) and washing (from 5 x 1.0 min to 5 x 40 sec) as well as increasing the time of mechanical disruption (from 10 min to 15 min) (Figure S1A), cell yield could be increased. After dissociation, cell suspension was stained with propidium iodide (PI) to label dead/damaged cells and then subjected to FACS. GFP positive singlet cells were sorted (Figure 1B). Majority of sorted cells showed GFP fluorescence under confocal microscope (Figure 1C). Consistent with previous findings, neurites can grow out after culture of sorted neuron cells (Figure S1B) (Spencer et al., 2014; Zhang et al., 2011). To further confirm the effectiveness of sorting, the levels of two marker genes (*myo-3* and *unc-64*) were quantified by digital droplet PCR (ddPCR). As expected, the neural syntaxin *unc-64* was highly enriched in the sorted cells, whereas the muscle gene *myo-3* was depleted (Figure 1D).

Using this optimized protocol (see "Materials and methods"), we obtained 200 - 500 ng total RNA from cells sorted from ~ 1.5 - 5 million L1 worms (the sort group). RNA samples from dissociated worms before sorting were also prepared for comparison (the whole group, Figure 1A). For RNA-seq, 150 ng total RNA from three independent trials of the sort group and the whole group was used as input for library preparation with ribosomal RNA removal and first strand cDNA synthesis using random hexamers. More than 45 million 150 nt paired-end reads were obtained for each sample. Differential expressed genes between the two groups were analyzed by DESeq2 (Love et al., 2014). Consistent with the ddPCR results (Figure 1D), *myo-3* was significantly depleted, while *unc-64* was significantly enriched in the sort group compared with the whole group (Figure S1C). The significantly upregulated genes (Table S4) in the sort group were searched in WormExp (Yang et al., 2016) (https://wormexp.zoologie.uni-kiel.de/wormexp/) to identify whether these genes overlap with previous results of neuronal expressed genes. As expected, the resulted top three datasets were all pan-neural enriched genes determined by microarray analysis of sorted neurons

(Von Stetina et al., 2007; Watson et al., 2008), indicating our RNA-seq results from sorted samples successfully revealed the gene expression pattern in the neurons.

2. circRNAs are highly expressed in the neurons

circRNA annotation from the RNA-seq data was carried out using the DCC pipeline (Cheng et al., 2016). 6475 circRNAs were identified with at least one junction read across six samples, with 4786 novel circRNAs when compared with circRNAs of *C. elegans* in two databases: circBase (Glazar et al., 2014) and CIRCpedia v2 (Dong et al., 2018) (Figure S2A). The results were filtered with a cutoff of at least three junction reads in each group, which gave 1452 circRNAs derived from 1010 genes and 29 not-annotated loci (Figure 2A, Table S5). Of the filtered circRNAs, majority of the identified junction reads were from exon-to-exon joining (Figure S2B). The filtered circRNAs were compared with published dataset of circRNAs in aging worms (Cortes-Lopez et al., 2018), which showed 450 overlapped circRNA (Figure S2C). The novel circRNAs identified in our dataset were mainly from the sorted group, suggesting that sequencing form sorted neuron samples was helpful to identify circRNAs that may not be detected using whole-worm samples. Gene ontology (GO) enrichment analysis of the circRNA-producing genes showed that terms related to the neuronal functions were significantly enriched, like neurogenesis, synaptic signaling, etc (Figure 2B).

To validate the annotated circRNAs by DCC, two strategies were used: 1) Amplification of the back-spliced junction sequences by RT-PCR using divergent primers followed by Sanger-sequencing (Figure 2C). Eighteen out of 19 selected circRNAs, including seven novel circRNAs, were confirmed with the back-splicing junction (BSJ) sequences (Figure S2D). 2) Enrichment quantification by RT-qPCR after RNase R treatment. Since there are no ends in circRNAs, they show RNase R resistance. As expected, while two linear mRNAs (*pmp-3* and *cdc-42*) were depleted after RNase R treatment, all the circRNAs were enriched (Figure 2D). The resistance to RNase R was also confirmed by northern blot, which showed that while linear transcript was not detected after RNase R treatment, circRNA from *Y20F4.4* was still detected (Figure 2E). Taken together, circRNA annotation is of high accuracy.

Of the 1452 circRNAs, more circRNAs (1125/1452) were found in the sort group, with 539 found in both groups and 586 found only in the sort group (Figure 2A). Then, the abundances of circRNAs in the sort group and the whole group were compared to check whether circRNAs were highly expressed in the neurons of *C. elegans* or not. TPM (transcripts per million reads) values were used for comparison. The principal component analysis (PCA) plot of circRNA TPM showed a clear separation between the two groups (Figure S2E), suggesting different circRNA profiles between them. For all the circRNAs in both groups, circRNAs in the sort group showed significantly higher TPM values than in the whole group (Figure 2F, $p < 2.2e-16$, paired Wilcoxon test), indicating circRNAs were enriched in the sort group. The same trend was also observed for the shared 539 circRNAs in both groups (Figure 2G, $p = 4.7e-14$, paired Wilcoxon test).

Next, we analyzed differentially expressed circRNAs between the sort and the whole group, trying to identify neuron-enriched circRNAs. Using BSJ read numbers as input for DESeq2 and adjusted

p value < 0.05, 31 circRNAs were found significantly enriched and 35 circRNAs were significantly depleted in the sort group (Figure S2F, Table S6). We asked whether these circRNAs were also derived from neuronal genes or not. The fold changes of circRNAs between the sort and the whole group were plotted against the fold changes of their cognate linear mRNAs. Here, a cutoff of baseMean (given by DESeq2) bigger than 3 was used, which contained 268 circRNAs including all the significantly differentially expressed circRNAs (Figure 2H). The results showed a strong positive correlation (Figure 2H, Pearson's correlation coefficient $R = 0.72$, $p < 2.2e\text{-}16$), which suggested that at the L1 stage of *C. elegans*, neuronal circRNAs were more likely to be derived from neuronal genes. When all circRNAs were considered, it still showed a moderate strong positive correlation (Figure S2G, Pearson's correlation coefficient $R = 0.51$, $p < 2.2e\text{-}16$).

In summary, the first neuronal circRNA profiles in *C. elegans* have been obtained, which show that circRNAs are highly expressed in the neurons and positively correlated with their cognate mRNAs.

3. RCMs are required for circRNA production

Next, we analyzed features of circRNA-flanking introns. Similar to previous findings (Cortes-Lopez et al., 2018; Ivanov et al., 2015), introns that flank circRNA-producing exons were much longer than average and much more RCMs were identified when compared with flanking introns of non-circular exon controls (exon 2 and exon 8) (Figure 3A and 3B). Lengths of the best-matched RCMs in circRNA introns were also much longer than those in introns flanking control exons (Figure S3A). Although previous study showed that RCMs can predict the existence of circRNAs (Ivanov et al., 2015), the role of RCMs in circRNA formation has not been experimentally confirmed in *C. elegans*. Here, six circRNA genes with RCMs in flanking introns were chosen and one RCM in each gene was deleted using CRISPR-Cas9 (Figure 3C and Figure S3B). As expected, all the circRNAs were either undetectable or reduced to very low level after the removal of one RCM in the flanking introns (Figure 3D and Figure S3C), proving that RCMs in *C. elegans* strongly promote circRNA formation. Of note, in some of the chosen genes, the linear mRNA levels were altered in RCM deletion mutants.

4. RCMs promote both back-splicing and exon-skipping

circRNA production has been correlated with exon-skipping that skips the circularized exon(s) (Kelly et al., 2015; Surono et al., 1999; Zaphiropoulos, 1996, 1997). In our RNA-seq data, reads mapped to the skipped junctions can be identified in some circRNA genes (Figure S4A and S4B), suggesting the existence of skipped transcripts. In *zip-2* and *Y20F4.4*, RT-PCR using primers that bracket circRNA-producing exon(s) gave 2 bands, of which the longer ones were full-length transcripts, and the shorter ones were confirmed to be the skipped transcripts (Figure 4A and Figure S4E). In some other circRNA genes, the skipped transcripts can be amplified in a 2-round PCR, in which the corresponding skipped transcripts were gel-cut purified after first round PCR, which

were used as templates for a second round PCR (Figure S4C and S4D). In total, skipped transcripts were confirmed in 6 out of 7 chosen circRNA genes (Figure S4E).

Previous studies have shown that conserved complementary sequences in introns are associated with exon-skipping (Miriami et al., 2003), and Alu sequences in introns regulate alternative splicing, including exon-skipping (Lev-Maor et al., 2008). We then checked whether RCMs are also required for exon skipping. In *Y20F4.4*, the skipped transcript was strongly reduced after RCM removal (Figure 4B and Figure S5A). In *arl-13*, the skipped transcript was downregulated in the RCM-deleted mutant (Figure 4C and Figure S5B). In *zip-2*, two pairs of perfectly matched RCMs, 7 nt and 13 nt in length, respectively, were identified (Figure 4D and Figure S5C). Deletion of the RCMs in intron 1 or intron 4 was achieved by CRISPR-Cas9 (Figure S5C). It showed that canonical splicing of intron 1 and intron 4 were not affected by RCM deletions (Figure 4E, Exon 1-2 & Exon 4-5). However, although the circRNA and skipped transcript can be detected in the RCM-deleted strains, their levels seemed much lower than those in wild type (Figure 4E, circ & skip). Quantification of the levels of the 3 transcripts of *zip-2* (circular, skipped, full-length linear) showed that while full-length linear *zip-2* was only slightly affected, the production of both the circRNA and the skipped transcript were dramatically reduced in both RCM-deleted mutant strains (Figure 4F and 4G). Together, these findings suggest that RCMs in the flanking introns of circRNA-producing exon(s) also promote the skipping of these exon(s).

5. RCMs simultaneously promote back-splicing and exon-skipping

Current knowledge suggests that RCMs promote circRNA formation by bringing the splicing sites for back-splicing to close proximity. Since the correlated back-splicing and exon-skipping use the same pair of introns, it is possible that RCMs also bring the splice sites for exon-skipping together. In principle, the y-shaped intermediate of back-splicing could be further spliced to form the corresponding skipped transcripts. Moreover, previous study has shown that circRNA can be produced through a lariat intermediate produced by exon-skipping. This gives 3 possibilities: 1). RCMs promote back-splicing first; 2). RCMs promote exon-skipping first; 3). RCMs promotes both back-splicing and exon-skipping at the same time (Figure 6). To clarify the 3 possibilities, the 4 splice sites (ss) and 2 branch points (BP) in intron 1 and intron 4 of *zip-2* were mutated one by one. The 5'ss in intron 1, BP and 3'ss in intron 4 are needed for exon-skipping, hence these sites are named as skip-5'ss, skip-BP and skip-3'ss, respectively. Similarly, BP and 3'ss in intron 1 and 5'ss in intron 4 were named as circ-BP, circ-3'ss and circ-5'ss, respectively (Figure 5B). For ss mutation, we deleted the conserved AG or GT and mutated some possible cryptic ss nearby (Figure S6A and S6B). For BP mutation, since there is little information about BP sites in *C. elegans* (Zahler, 2012), we changed all A nucleotides to G nucleotides in the 3' half of intron 1 and intron 4, without disturbing the RCM sequences. (Figure S6A and S6B). The results showed that mutation of ss and BP for exon-skipping sufficiently abolished *zip-2-skip*. However, *circ-zip-2* was still produced in these mutant strains (Figure 5A). Similarly, mutation of ss and BP required for back-splicing blocked circRNA production, but exon-skipping still can be detected (Figure 5A). These results suggest that in *zip-2*, exon-skipping is not absolutely required for back-splicing and *vice versa*. RCMs can simultaneously promote both exon-skipping and back-splicing

Next, the levels of *circ-zip-2* and *zip-2-skip* in these mutant strains were quantified by RT-qPCR. As expected, *zip-2-skip* was either undetectable or reduced to very low levels in strains with skip-5'ss, skip-3'ss or skip-BP mutated. However, *circ-zip-2* in these 3 strains were either increased or decreased. Similar things were observed in strains with circ-BP, circ-3'ss or circ-5'ss mutated: while *circ-zip-2* was sufficiently blocked, *zip-2-skip* showed increased level in 2 strains (circ-BP and circ-3'ss) and decreased level in 1 strain (circ-5'ss). Interestingly, when the mutated sites were in the upstream intron (intron 1), the transcripts that do not use the mutated sites were all increased. This may be due to the competition of back-splicing and exon skipping with canonical splicing of intron 1. Since splicing is always happening during transcription (Herzel et al., 2017), mutation of ss or BP in intron 1, which results in the intron retention (Figure S6C), could increase the possibility to use the splicing sites in downstream introns, which could be either exon-skipping or back-splicing. However, when the mutated sites were in the downstream intron (intron 4), the transcripts that do not use the mutated sites were all decreased. Given that canonical splicing in intron 1 could be dominant, the chances of back-splicing or exon-skipping may be quite limited when transcription goes to intron 4. Hence, blocking either of them could result in the reduction of the final products: *circ-zip-2* or *zip-2-skip*.

Together, the ss/BP mutations suggest that RCM-promoted back-splicing and exon-skipping work together to produce the circular and skipped transcripts (Figure 6).

Discussion

In this study, we optimized a method for large-scale neuron isolation from L1 worms. The amount of obtained RNA from sorted neurons was increased for the first time ever to hundreds of nanogram scale, making the detection of circRNA by whole-transcriptome RNA-seq more reliable. Using this method, we provided the first neuronal circRNA profiles in *C. elegans* and found that circRNAs were abundant in the neurons. Interestingly, circRNAs showing higher levels in the neurons tend to be derived from genes that also show higher expression in the neurons (Figure 2H). The time between egg to L1 is the first main period of neuron development and at the time of hatching, the majority (222/302) of neurons are already formed (Altun and Hall, 2011). The high levels of these circRNAs may be due to the active expression of their parental genes for neuron development at the L1 stage.

RCMs are abundant in circRNA introns of *C. elegans*. For the first time, we validated that RCMs are required for circRNA formation in multiple circRNA genes. This provides a sufficient method to knockout (KO) circRNAs in *C. elegans*, given the highly efficient genome editing by CRISPR-Cas9. However, the circRNA-KO strains generated in this work did not show any obvious phenotypes in several assays, like locomotion, chemotaxis, lifespan and Aldicarb resistance (data not shown). Even if some phenotypes were observed, it should be carefully interpreted, since the linear mRNA levels could also be changed (Figure 3D). In *zip-2*, we identified 2 short pair of RCMs, 7 nt and 13 nt in length. To our best knowledge, they are the shortest identified *cis* elements that promote circRNA formation.

Currently, two models have been proposed to explain the correlation between exon-skipping and circRNA formation (Jeck and Sharpless, 2014; Jeck et al., 2013): 1). RCM-promoted back-splicing produces circRNAs and y-shaped intermediates, which are further spliced to form skipped transcripts; 2). Exon-skipping produces skipped transcripts and lariat intermediates, which are further back-spliced to form circRNAs. For the first time we show that RCMs not only are required for back-spicing, but also promote exon-skipping. We further delineated that RCMs are not promoting exon-skipping through back-splicing, neither the other way. Rather, RCMs simultaneously promote both back-splicing and exon-skipping. We propose that RCMs in the introns not only bring the splice sites for back-splicing to proximity, but also bring the sites for exon-skipping together, facilitating both processes simultaneously. Since the RCMs still exist in the intermediates of back-splicing and exon-skipping, they may function twice to promote further splicing/back-splicing in these intermediates (Figure 6).

Previous studies of RCMs' role in circRNA regulation or splice sites required for back-splicing were mainly based on plasmids in cultured cells, which are usually overexpressed. In this work, we show that *C. elegans* is a good model for *in vivo* investigation of circRNA regulation.

Acknowledgement

The NW1229 strain was provided by CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We thank Mr. David Knupp and Dr. Pedro Miura from University of Nevada, Reno for instructions on northern blot. We thank Dr. Ichiro Maruyama and Dr. Pedro Miura for scientific discussions. We thank the members of Information Processing Biology Unit (Maruyama Unit) for their discussions and feedbacks. We thank DNA Sequencing Section of OIST for the help on RNA sequencing. We are grateful for the help and support provided by the Scientific Computing and Data Analysis section of Research Support Division at OIST. We thank Okinawa Institute of Science and Technology, Graduate School for financial support.

Author Contributions

D.C. designed and conducted all the experiments, performed all the analysis, and wrote the paper.

Declaration of Interests

The author declares no competing interests.

Figure Legends

Figure 1. Large scale neuron isolation from *C. elegans* for circRNA detection.

(A) Workflow of neuron isolation and circRNA detection by RNA-seq.

(B) Gating strategy for FACS, live and GFP-positive cell singlets are sorted.

(C) Confocal images of sorted GFP-positive neurons. Scale bar: 50 μm.

(D) ddPCR results showing the relative levels of two genes (*myo-3* and *unc-64*) in the sort group compared with those in the whole group. Error bars stand for standard deviations of three biological replicates. *P* values are ratio paired *t* test.

Figure 2. circRNAs are enriched in the neurons of *C. elegans*.

(A) Overlap of circRNAs detected in the "sort group" and the "whole group".

(B) Top 10 enriched gene ontology (GO) terms of circRNA-producing genes.

(C) Scheme showing amplification of back-splicing junction of a circRNA from *glr-2* using divergent primers. Amplified sequences are confirmed by Sanger sequencing. Red triangle denotes the joint site.

(D) RT-qPCR results of fold change of circRNAs and two linear mRNAs (*pmp-3* and *cdc-42*, inset) after RNase R treatment. Blue dashed line shows one-fold change. Error bars are the standard deviations of three biological replicates.

(E) Northern blot detection of *Y20F4.4* transcripts in total RNA (20 μg) without or with RNAse R treatment, using probes that hybridize to both linear and circular transcripts.

(F, G) TPM (transcripts per million reads) comparison of all circRNAs (F) and shared circRNAs (G) between the "sort group" and the "whole group"; *p* values are paired Wilcoxon test.

(H) Scatter plot showing the fold changes of 268 circRNAs with baseMean > 3 versus fold changes of their corresponding linear mRNAs. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown. Significantly differentially expressed circRNAs are shown by colored dots. Names of several circRNA genes are labeled.

Figure 3. RCMs are required for circRNA production.

(A) Length distributions of introns flanking circRNA-producing exon(s), compared with the lengths of all introns.

(B) Average number of RCMs in one pair of flanking introns of circRNA compared with those in non-circRNA control exons (exon 2 and exon 8). Values are shown as mean ± SEM. Numbers in

the brackets are numbers of intron pairs used for analysis. *p* values in A & B are from Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons. ****, $p < 0.0001$.

(C) Schematic plot showing that RCMs promote circRNA production and the strategy to disturb one of the RCMs by CRISPR-Cas9.

(D) Quantification of linear mRNA and circRNA in wildtype N2 strain and in RCM deletion mutant strains of six circRNA genes. Error bars are the standard deviations of three biological replicates. n.d.: not detected (Ct values not determined or bigger than those in no-template controls (NTC)).

Figure 4. RCMs promote both back-splicing and exon-skipping.

(A) RT-PCR detection of full-length transcripts and skipped transcripts in *zip-2* and *Y20F4.4*.

(B) RT-qPCR quantification of *Y20F4.4* transcripts in wildtype and RCM-deleted *Y20F4.4* strain.

(C) RT-qPCR quantification of *arl-13* transcripts in wildtype and RCM-deleted *arl-13* strain.

(D) Illustration of gene structure of *zip-2*. P1-P6: position of primers. Black rectangles indicate coding regions and white parts are untranslated regions (UTRs). RCM areas are in red.

(E) RT-PCR detection of transcripts from *zip-2* gene in wildtype N2 strain and RCM-deleted strains.

(F, G) RT-qPCR quantification of *zip-2* transcripts (circular: *circ-zip-2*, skipped: *zip-2-skip*, full length linear: *L-zip-2*) in RCM-deleted strains compared with wildtype N2 strain.

(B,C F & G) Results are normalized to levels in N2 strain using *pmp-3* as the reference gene. Error bars are the standard deviations of three biological replicates. ***, $p < 0.001$, ****, $p < 0.0001$, two-tail Student's *t* test.

Figure 5. RCMs simultaneously promote back-splicing and exon-skipping

(A) RT-PCR detection of *zip-2* transcripts in wildtype N2 strain and strains with mutated ss or BP.

(B) Gene structure of *zip-2*. P1-P6: position of primers. Position of splicing sites (ss) and branch points (BP) that are required for back-splicing and exon-skipping are labeled.

(C) RT-qPCR quantification of levels *zip-2* circRNA (*circ-zip-2*) and skipped transcripts (*zip-2-skip*) in the indicated ss mutation strains and BP mutation strains. Results are normalized to levels in N2 strain using *pmp-3* as reference gene. Error bars are the standard deviations of three biological replicates. n.d.: not detected. ***, $p < 0.001$, ****, $p < 0.0001$, Two-tail Student's t test.

Figure 6. Proposed model that RCMs promote both back-splicing and exon-skipping at the same time.

(1) Canonical splicing to form full-length linear mRNA. (2) RCMs facilitate circRNA formation by bringing splice sites for back-splicing sites together. The y-shaped intermediate is further spliced to form the skipped transcript. (3) RCMs promote exon-skipping by bringing splice sites for exon-skipping together. The lariat intermediate is further backspliced to form circRNA. RCMs in the y-shaped intermediate and the lariat intermediate may help the 2nd splicing steps. The 3 pathways compete with each other to regulate the levels of the 3 types of transcripts.

Materials and methods

Worm maintenance

C. elegans Bristol N2 strain was used as the wild type. Worms were maintained using standard conditions on Nematode Growth Media (NGM) agar plates with *Escherichia coli* strain OP50 (Brenner, 1974) at 20°C or 25ºC. Strains used in this study are listed in Table1.

Worm synchronization

Worm synchronization was performed by bleaching for large-scale worm preparation (L1 worms for dissociation). Briefly, worms were washed off plates using M9 buffer when a lot of eggs were laid and most of the worms were gravid adults. The worms were washed with M9 buffer and then bleached in bleach solution (1 M NaOH, 0.6% (m/v) NaClO) with ~5 minutes continuous shaking. Then eggs were pelleted and washed three times with 12 ml M9 buffer by centrifuging at 2000 rpm for 0.5 minutes. Finally, the egg pellet was re-suspended in ~5 ml M9 buffer and rocked at room temperature for 17-24 hours to hatch. For small-scale worm preparation (locomotion assay), worms were synchronized by egg-laying. Briefly, 10-15 gravid adult worms were placed onto NGM plate for four hours and worms were removed after egg-laying. The eggs were then cultured to desired stage.

L1 worm dissociation

To ~80 µl of L1 worm pellet, 200 µl SDS-DTT solution (200 mM DTT, 0.25% SDS, 20 mM HEPES pH 8.0, 3% sucrose) was added, followed by a 1.5-minute incubation at room temperature. Then, the worm pellet was washed 5x with 1 ml egg buffer (25 mM HEPES pH 7.3, 118 mM NaCl, 48 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 0.340 ± 0.005 Osmolarity) and centrifuged at 10,000 ×g for 30 seconds. The washing steps and centrifugation should be performed quickly so that one round of washing and centrifugation is done in 40-50 seconds. The washed worm pellet was re-suspended in 100 µl pronase (15 mg/ml in egg buffer) from *Streptomyces griseus* (Sigma-Aldrich). Worms were dissociated by periodic mechanical disruption by pipetting for 15 minutes. 200 µl tips were used for mechanical disruption by the method mentioned in Zhang et al's protocol ("Pipette the larvae suspension with a 200 µl tip during the digestion. Adjust the pipetting volume to the approximate volume of the suspended pellet. Slowly pull suspended larvae into the pipette tip. Then, press down to force the pipette tip against the bottom of the microcentrifuge tube and slowly eject the contents") (Zhang et al., 2011). Do as many times as possible. When most worm bodies were dissociated, 900 µL L-15/FBS medium (10% FBS in Leibovitz's L-15 medium (Gibco), 0.340 ± 0.005 Osmolarity adjusted by sucrose) were added. Cells were collected and washed twice with 1 ml egg buffer by centrifuging at 9600× g for 5 min at 4ºC. Cells were suspended in the appropriate amount of egg buffer and allowed to sit on ice for at least 30 min. The upper volume of cell suspension was used for FACS. For whole worm control, after dissociation and washing, the cell suspension was put on ice in the whole procedure of sorting.

Fluorescence-Activated Cell Sorting (FACS)

Sorting was performed on a FACS AriaII flow cytometer (Becton Dickinson) equipped with a 70 µm nozzle. 2 µm and 3.4 µm polystyrene beads (Spherotech) were used for size calibration. Before

sorting, propidium iodide (PI) was added to the cell suspension to a final concentration of 0.2 - 0.5 µg/ml. Then, profiles of dissociated cells from GFP-labeled strains were compared to profiles of cells from N2 worms to exclude auto-fluorescent cells. Sorted cells were collected in 3 ml L-15/FBS medium in a 15 ml conical tube chilled on ice. For RNA extraction, sorted cells and whole worm control samples were collected by centrifugation in a swing-bucket centrifuge at 4400 rpm, 4ºC for 10 min. The supernatant was removed and 0.3 ml Trizol solution (Invitrogen) were added and stored at -80ºC. For culture, sorted cells were seeded onto a poly-D-lysine coated glass-bottom dish (MatTec) with daily changes of L-15/FBS buffer. Cells were visualized by confocal microscopy (Carl Zeiss, LSM780) with 60 × oil lens.

Mutagenesis by CRISPR-Cas9

Mutation by CRISPR-Cas9 was based on the protocol published by Dokshin et al (Dokshin et al., 2018) with minor modifications. Briefly, Cas9 (Sigma-Aldrich; 0.5 µl, 10 µg/µl in supplied buffer), tracrRNA (Sigma-Aldrich; 5 µl, 0.4 µg/µl in 10 mM Tris-HCl, pH7.5) and designed crRNA (ThermoFisher or IDT; 0.4 µg/µl in Tris-HCl, pH7.5; 2.8 µl for single crRNA, 1.4 µl each for 2 crRNAs) were mixed and incubated at 37ºC for 10 min. Then recombinant dsDNA fragment (add to > 400 ng/µl final concentration) or recombinant single strand DNA (ordered from Invitrogen or IDT; 2.2 µl, 1 µg/µl in Tris-HCl, pH7.5), injection marker pRF4(*rol-6*) (2.7 µl, 300 ng/µl in Tris-HCl, pH7.5), KCl (0.5 µl, 1 M), HEPES (1 µl, 0.2 M, pH7.4) and H_2O were added to make a final 20µl injection mixture. Injected P0 worms were recovered at 20ºC overnight and then transferred to RT (25ºC). For F1 with obvious phenotypes, F1 worms with target phenotype were picked and homozygous progeny were kept. For mutagenesis with no obvious genotypes, ~10 F1 rollers were picked onto separated plates and their genotypes were checked by single worm PCR after laying eggs. Large indels were identified by amplicon size differences. Small indels were checked by enzyme digestion of amplicons. Non-roller homozygous progenies with target genotype were kept. CrRNAs, recombinant ssODNs, validation primers and restriction enzymes used in this work are listed in Table S3.

RNA extraction

RNA extraction from sorted samples and whole worm samples was performed using Direct-zol RNA MicroPrep kit (ZYMO Research) with on-column DNase I (ZYMO Research) digestion according to the manufacturer's protocol. RNA quality and quantity were measured by High Sensitivity RNA ScreenTape (Agilent) on TapeStation 4200 (Agilent). For RNA extraction from worms, worms were first flash frozen in Trizol solution (Invitrogen) in liquid N_2 and then homogenized in by vortexing with glass beads (ϕ 0.1 mm) in Beads Cell Disrupter MS-100 (TOMY). If not mentioned, all the RNA samples used in this study were from L1 worms of indicated genotypes.

RNA Sequencing

For RNA-seq of samples from sorted neurons (the sort group) and whole worms (the whole group), libraries were prepared using KAPA RNA HyperPrep kit with RiboErase (HMR) (KAPA biosystems) according to manufacturer's protocol. The RNA input was 150 ng and fragmentation

conditions were 85ºC for 5 min. Barcodes were introduced to each sample using KAPA duel-indexed adapters (KAPA biosystems). Length distribution of each library was determined by TapeStation 4200 (Agilent) using High Sensitivity DNA ScreenTape (Agilent). Libraries were quantified by KAPA library quantification kit (KAPA biosystems) and then multiplexed and sequenced on Illumina Hiseq 4000 platform to obtain 150 nt paired-end reads.

Droplet digital PCR (ddPCR)

cDNA was reverse transcribed from 10 ng total RNA using an iScript Advanced cDNA synthesis kit (Bio-Rad). ddPCR was performed by using ddPCR EvaGreen Supermix kit (Bio-Rad) on a QX200 Droplet Reader (Bio-Rad) based on the manufacturer's protocol. Results were analyzed using QuantaSoft software (Bio-Rad).

Real-time PCR

Real-time PCR reactions were performed using soAdvanced Universal SYBR Green Supermix (Bio-Rad) with cDNAs synthesized from iScript Advanced cDNA synthesis kit (Bio-Rad). 20 µl reaction mix with 2 µl cDNA (~1-10 ng) were monitored on StepOnePlus Thermal Cycler (Applied Biosystems) in "fast mode". Cycling conditions: 95ºC, 30', 40 or 45 cycles of 95ºC, 15' and 60ºC, 30'with plate reading, and a final melt curve stage using default conditions. Primers used are listed in Table S2.

RNase R treatment

Total RNA was treated with or without (Mock) RNase R (2 U/µg) in the presence of Ribolock(2 U/µg) (ThermoFisher Scientific). The reaction was incubated at 37ºC for 30 min. Then RNA was purified with an RNA Clean and Concentrator kit (ZYMO Research) according to the manufacturer's protocol. For fold change quantification, RNA was quantified by Nanodrop and an equal amount of RNA input was used for cDNA synthesis. For northern blot, 20 µg total RNA with or without RNase R treatment was used for loading.

Northern blot

Northern blot was performed using NorthernMax kit (ThermoFisher Scientific) and the probes were labeled by α-^{32}P-deoxycytidine 5'-triphosphate (PerkinElmer) using Random Primer DNA Labeling Kit Ver. 2 (Takara, #6045) according to manufacturer's protocols. Briefly, RNA samples (10 µg or 20 µg) were resolved in 1% agarose gel by electrophoresis at 5 V/cm in 1× MOPS buffer for ~2 hours. Then RNA was transferred onto an Amersham Hybond-N+ membrane (GE Healthcare) by capillary blot for 2.5 hours using the transfer buffer supplied in NorthernMax kit. Transferred RNA was crosslinked by 254 nm UV at 1200 × 100 µJ/cm^2(Analytik Jena CL-1000). Prehybridization was performed in ULTRAHybe buffer at 50ºC for one hour followed by hybridization with ^{32}P labeled probes overnight at 50ºC. The membrane was washed 2 × 5 min at room temperature using Low Stringency Washing Solution and 2 × 15 min at 50ºC using High Stringency Washing Solution. The membrane was sealed in kitchen wrap and exposed to a phosphorscreen for several hours to overnight, and the signals were detected by Typhoon FLA7000 (GE Healthcare). Quantification of bands was performed using ImageQuant software (GE Healthcare). Primers used for probe amplification are listed in Table S2.

circRNA prediction and RNA-seq data analysis

The DCC pipeline (Cheng et al., 2016) was used for prediction of circRNAs with RNA-seq data. Briefly, raw reads were aligned to reference genome (WBcel235/ce11) using STAR (Dobin et al., 2013) (https://github.com/alexdobin/STAR) with the following options: --outSJfilterOverhangMin 15 15 15 15 –alignSJoverhangMin 15 –alignSJDBoverhangMin 15 --outFilterScoreMin 1 --outFilterMatchNmin 1 --outFilterMismatchNmax 2 --chimSegmentMin 15 --chimScoreMin 15 --chimScoreSeparation 10 --chimJunctionOverhangMin 15. Then the output files from STAR, chimeric.out.junction, were used for circRNA annotation with DCC (https://github.com/dieterich-lab/DCC). Predicted circRNAs from DCC were filtered with at least three junction reads in each group. Differential expression analysis of mRNAs and circRNAs were performed using DESeq2 (Love et al., 2014) package in R with the gene count output from STAR or the BSJ junction count output from DCC, respectively. The plots (PCA plots, boxplots, scatter plots) were generated using ggplot2 package (https://ggplot2.tidyverse.org/), and ggpubr (http://www.sthda.com/english/rpkgs/ggpubr) package in R.

RCM analysis

RCM analysis in flanking introns of circRNAs or non-circular control exons were performed using IntronPicker and autoBLAST scripts (https://github.com/alexandruioanvoda/) described in (Cortes-Lopez et al., 2018).

Microscopy

Confocal images were obtained using a Zeiss LSM780 confocal microscope and images were processed using ZEISS ZEN3.1 software.

Gene ontology analysis

Gene ontology enrichment analysis was performed using WormBase Enrichment Suite web server (Angeles-Albores et al., 2018; Angeles-Albores et al., 2016) (https://wormbase.org/tools/enrichment/tea/tea.cgi).

Statistical analysis

Statistical analysis was performed using R or Prism (GraphPad).

Data availability

Raw FASTQ files from the RNA-seq data were deposited at the NCBI Sequence Read Archive (BioProject: PRJNA669379). All strains and other materials are available upon request.

References:

Altun, Z.F., and Hall, D.H. (2011). Nervous system, general description. In Worm Atlas.
Angeles-Albores, D., Lee, R., Chan, J., and Sternberg, P. (2018). Two new functions in the WormBase Enrichment Suite. MicroPubl Biol *2018*.
Angeles-Albores, D., Lee, R.Y.N., Chan, J., and Sternberg, P.W. (2016). Tissue enrichment analysis for C. elegans genomics. Bmc Bioinformatics *17*.
Barrett, S.P., Wang, P.L., and Salzman, J. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. Elife *4*, e07540.
Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71-94.
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.
Chen, L.L. (2020). The expanding regulatory mechanisms and cellular functions of circular RNAs. Nat Rev Mol Cell Biol.
Cheng, J., Metge, F., and Dieterich, C. (2016). Specific identification and quantification of circular RNAs from sequencing data. Bioinformatics *32*, 1094-1096.
Christensen, M., Estevez, A., Yin, X., Fox, R., Morrison, R., McDonnell, M., Gleason, C., Miller, D.M., 3rd, and Strange, K. (2002). A primary culture system for functional analysis of C. elegans neurons and muscle cells. Neuron *33*, 503-514.
Cortes-Lopez, M., Gruner, M.R., Cooper, D.A., Gruner, H.N., Voda, A.I., van der Linden, A.M., and Miura, P. (2018). Global accumulation of circRNAs during aging in Caenorhabditis elegans. BMC Genomics *19*, 8.
Deffit, S.N., Yee, B.A., Manning, A.C., Rajendren, S., Vadlamani, P., Wheeler, E.C., Domissy, A., Washburn, M.C., Yeo, G.W., and Hundley, H.A. (2017). The C. elegans neural editome reveals an ADAR target mRNA required for proper chemotaxis. Elife *6*.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
Dokshin, G.A., Ghanta, K.S., Piscopo, K.M., and Mello, C.C. (2018). Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in Caenorhabditis elegans. Genetics *210*, 781-787.
Dong, R., Ma, X.K., Li, G.W., and Yang, L. (2018). CIRCpedia v2: An Updated Database for Comprehensive Circular RNA Annotation and Expression Comparison. Genom Proteom Bioinf *16*, 226-233.
Du, W.W., Yang, W., Wu, N., Li, X., Li, F., Chen, Y., He, Q., Liu, E., Yang, Z., Fang, L., *et al.* (2020). The Circular RNA circSKA3 Binds Integrin β1 to Induce Invadopodium Formation Enhancing Breast Cancer Invasion. Mol Ther.
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.
Fox, R.M., Von Stetina, S.E., Barlow, S.J., Shaffer, C., Olszewski, K.L., Moore, J.H., Dupuy, D., Vidal, M., and Miller, D.M., 3rd (2005). A gene expression fingerprint of C. elegans embryonic motor neurons. BMC Genomics *6*, 42.
Glazar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. Rna *20*, 1666-1670.
Gruner, H., Cortes-Lopez, M., Cooper, D.A., Bauer, M., and Miura, P. (2016). CircRNA accumulation in the aging mouse brain. Sci Rep *6*, 38907.

Gubbay, J., Vivian, N., Economou, A., Jackson, D., Goodfellow, P., and Lovell-Badge, R. (1992). Inverted repeat structure of the Sry locus in mice. Proc Natl Acad Sci U S A *89*, 7953-7957.
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.
Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat Rev Mol Cell Biol *18*, 637-650.
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.
Kaletsky, R., Lakhina, V., Arey, R., Williams, A., Landis, J., Ashraf, J., and Murphy, C.T. (2016). The C. elegans adult neuronal IIS/FOXO transcriptome reveals adult phenotype regulators. Nature *529*, 92-96.
Kelly, S., Greenman, C., Cook, P.R., and Papantonis, A. (2015). Exon Skipping Is Correlated with Exon Circularization. J Mol Biol *427*, 2414-2417.
Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., *et al.* (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. Mol Cell *66*, 22-37 e29.
Lev-Maor, G., Ram, O., Kim, E., Sela, N., Goren, A., Levanon, E.Y., and Ast, G. (2008). Intronic Alus Influence Alternative Splicing. Plos Genet *4*.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*.
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.
Miriami, E., Margalit, H., and Sperling, R. (2003). Conserved sequence elements associated with exon skipping. Nucleic Acids Res *31*, 1974-1983.
Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., *et al.* (2017). Translation of CircRNAs. Mol Cell *66*, 9-21 e27.
Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., *et al.* (2015). Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. Mol Cell *58*, 870-885.
Spencer, W.C., McWhirter, R., Miller, T., Strasbourger, P., Thompson, O., Hillier, L.W., Waterston, R.H., and Miller, D.M., 3rd (2014). Isolation of specific neurons from C. elegans larvae for gene expression profiling. Plos One *9*, e112102.
Surono, A., Takeshima, Y., Wibawa, T., Ikezawa, M., Nonaka, I., and Matsuo, M. (1999). Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. Hum Mol Genet *8*, 493-500.

Von Stetina, S.E., Watson, J.D., Fox, R.M., Olszewski, K.L., Spencer, W.C., Roy, P.J., and Miller, D.M. (2007). Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C-elegans nervous system. Genome Biology *8*.
Watson, J.D., Wang, S.L., Von Stetina, S.E., Spencer, C., Levy, S., Dexheimer, P.J., Kurn, N., Heath, J.D., and Miller, D.M. (2008). Complementary RNA amplification methods enhance microarray identification of transcripts expressed in the C. elegans nervous system. Bmc Genomics *9*.
Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R., and Lai, E.C. (2014). Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. Cell Rep *9*, 1966-1980.
Xia, P., Wang, S., Ye, B., Du, Y., Li, C., Xiong, Z., Qu, Y., and Fan, Z. (2018). A Circular RNA Protects Dormant Hematopoietic Stem Cells from DNA Sensor cGAS-Mediated Exhaustion. Immunity *48*, 688-701 e687.
Yang, W.T., Dierking, K., and Schulenburg, H. (2016). WormExp: a web-based application for a Caenorhabditis elegans-specific gene expression enrichment analysis. Bioinformatics *32*, 943-945.
Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L.L., Wang, Y., *et al.* (2017). Extensive translation of circular RNAs driven by N(6)-methyladenosine. Cell Res *27*, 626-641.
You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., *et al.* (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nat Neurosci *18*, 603-610.
Zahler, A.M. (2012). Pre-mRNA splicing and its regulation in Caenorhabditis elegans. WormBook, 1-21.
Zaphiropoulos, P.G. (1996). Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. Proc Natl Acad Sci U S A *93*, 6536-6541.
Zaphiropoulos, P.G. (1997). Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. Mol Cell Biol *17*, 2985-2993.
Zhang, M., Huang, N., Yang, X., Luo, J., Yan, S., Xiao, F., Chen, W., Gao, X., Zhao, K., Zhou, H., *et al.* (2018a). A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. Oncogene *37*, 1805-1814.
Zhang, M.L., Zhao, K., Xu, X.P., Yang, Y.B., Yan, S., Wei, P., Liu, H., Xu, J.B., Xiao, F.Z., Zhou, H.K., *et al.* (2018b). A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. Nat Commun *9*.
Zhang, S., Banerjee, D., and Kuhn, J.R. (2011). Isolation and culture of larval cells from C. elegans. Plos One *6*, e19505.
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.
Zhang, Y., Ma, C., Delohery, T., Nasipak, B., Foat, B.C., Bounoutas, A., Bussemaker, H.J., Kim, S.K., and Chalfie, M. (2002). Identification of genes expressed in C. elegans touch receptor neurons. Nature *418*, 331-335.
Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.L., Yang, L., and Chen, L.L. (2016). The Biogenesis of Nascent Circular RNAs. Cell Rep *15*, 611-624.

Zheng, X., Chen, L., Zhou, Y., Wang, Q., Xu, B., Hu, W., Wu, C., Jiang, J., Zheng, X., Chen, L., *et al.* (2019). A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. Mol Cancer *18*, 47.
Zhu, Y.-J., Zheng, B., Luo, G.-J., Ma, X.-K., Lu, X.-Y., Lin, X.-M., Yang, S., Zhao, Q., Wu, T., Li, Z.-X., *et al.* (2019). Circular RNAs negatively regulate cancer stem cells by physically binding FMRP against CCAR1 complex in hepatocellular carcinoma. Theranostics *9*, 3526-3540.

Figure 1

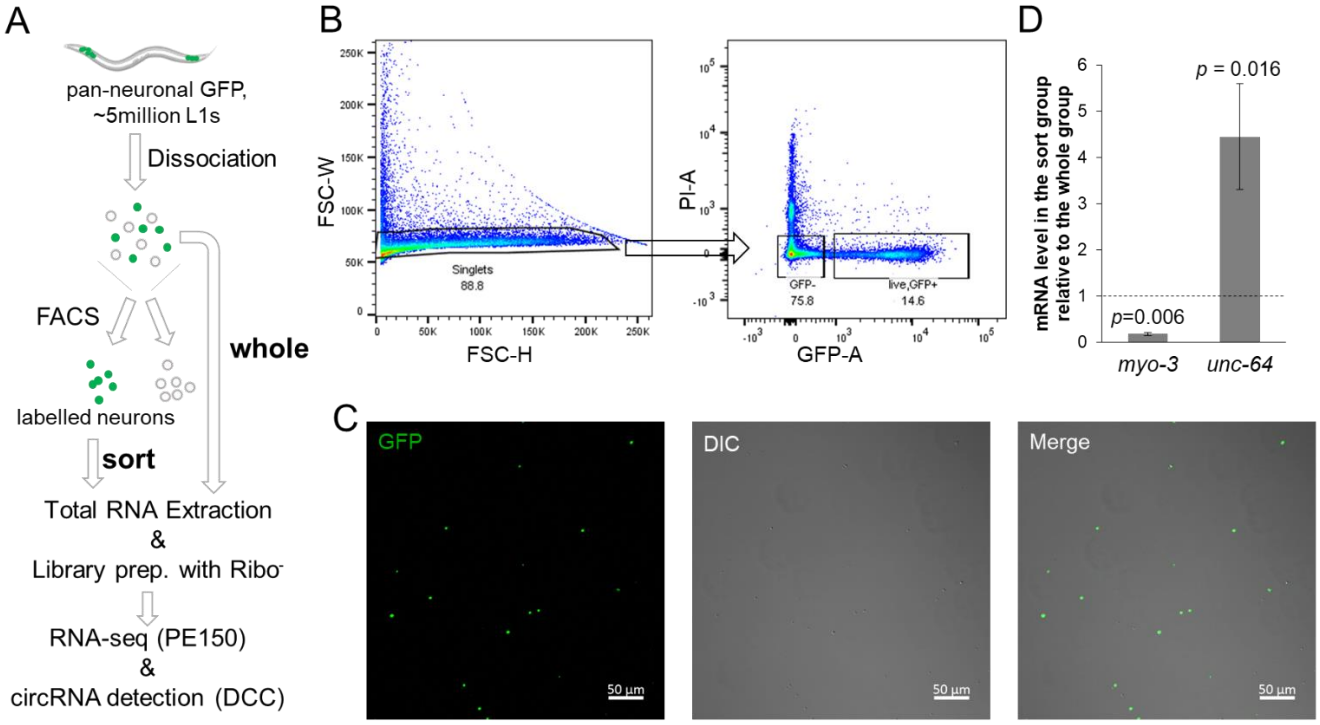

A
pan-neuronal GFP,
~5million L1s
Dissociation
FACS
whole
labelled neurons
sort
Total RNA Extraction
&
Library prep. with Ribo-
RNA-seq (PE150)
&
circRNA detection (DCC)
B
FSC-W
FSC-H
Singlets
88.8
PI-A
GFP-A
GFP-
75.8
live,GFP+
14.6
D
mRNA level in the sort group
relative to the whole group
p = 0.016
p=0.006
myo-3
unc-64
C
GFP
DIC
Merge
50 µm

Figure 2

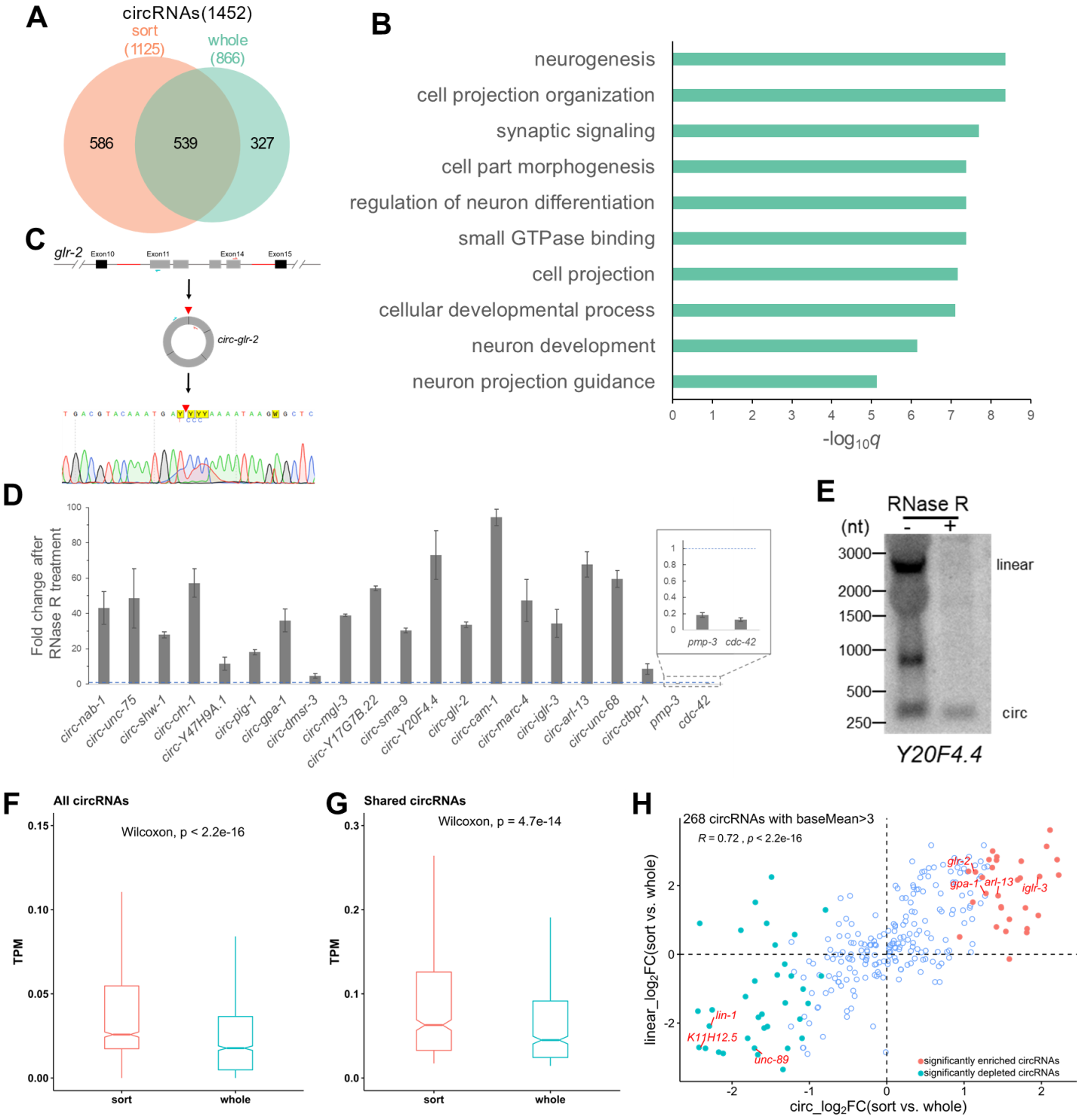

A
circRNAs(1452)
sort (1125)
whole (866)
586
539
327
B
neurogenesis
cell projection organization
synaptic signaling
cell part morphogenesis
regulation of neuron differentiation
small GTPase binding
cell projection
cellular developmental process
neuron development
neuron projection guidance
-log10q
C
glr-2
Exon10
Exon11
Exon14
Exon15
circ-glr-2
D
Fold change after RNase R treatment
circ-nab-1
circ-unc-75
circ-shw-1
circ-crh-1
circ-Y47H9A.1
circ-pig-1
circ-gpa-1
circ-dmsr-3
circ-mgl-3
circ-Y17G7B.22
circ-sma-9
circ-Y20F4.4
circ-glr-2
circ-cam-1
circ-marc-4
circ-iglr-3
circ-arl-13
circ-unc-68
circ-ctbp-1
pmp-3
cdc-42
E
RNase R
(nt)
-
+
3000
2000
1500
1000
500
250
linear
circ
Y20F4.4
F
All circRNAs
Wilcoxon, p < 2.2e-16
TPM
sort
whole
G
Shared circRNAs
Wilcoxon, p = 4.7e-14
TPM
sort
whole
H
268 circRNAs with baseMean>3
R = 0.72 , p < 2.2e-16
glr-2
gpa-1
arl-13
iglr-3
lin-1
K11H12.5
unc-89
significantly enriched circRNAs
significantly depleted circRNAs
linear_log2FC(sort vs. whole)
circ_log2FC(sort vs. whole)

Figure 3

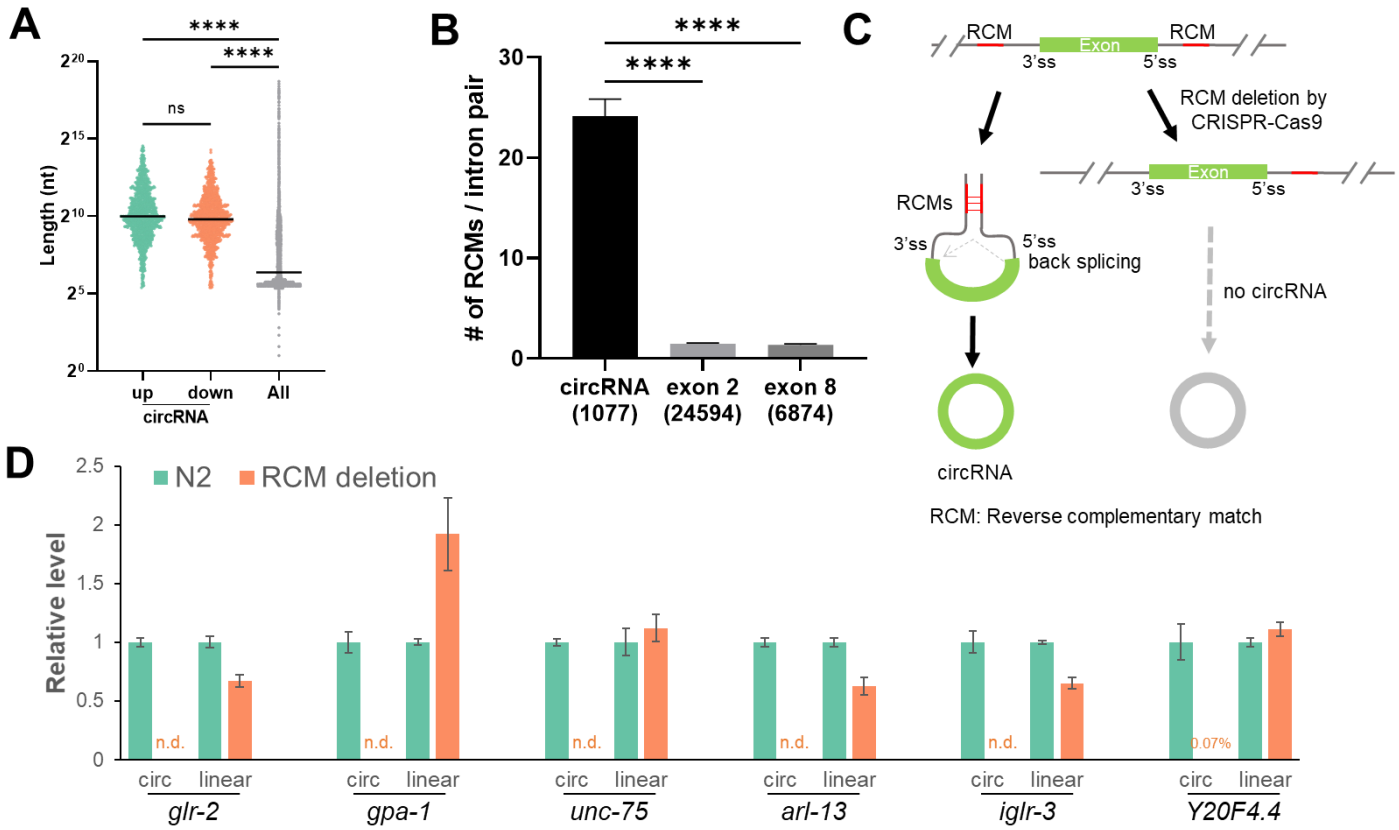

A
****
****
ns
Length (nt)
up
down
All
circRNA
B
****
****
# of RCMs / intron pair
circRNA (1077)
exon 2 (24594)
exon 8 (6874)
C
RCM
Exon
RCM
3'ss
5'ss
RCM deletion by CRISPR-Cas9
RCMs
3'ss
5'ss
back splicing
no circRNA
circRNA
RCM: Reverse complementary match
D
N2
RCM deletion
Relative level
n.d.
0.07%
circ
linear
glr-2
gpa-1
unc-75
arl-13
iglr-3
Y20F4.4

Figure 4

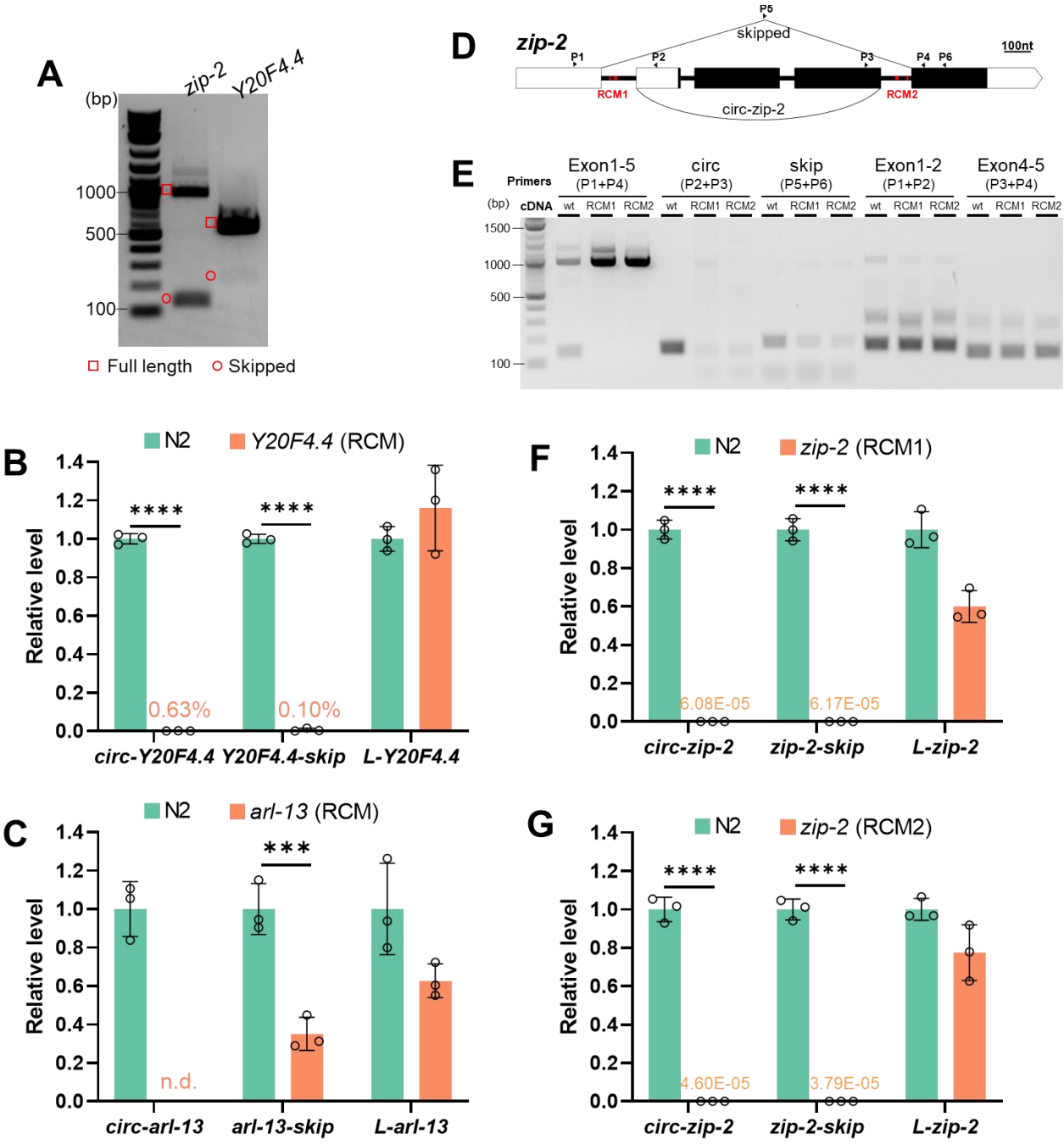

A
zip-2
Y20F4.4
(bp)
1000
500
100
Full length
Skipped
D
zip-2
P5
skipped
P1
P2
P3
P4
P6
100nt
RCM1
RCM2
circ-zip-2
E
Primers
Exon1-5 (P1+P4)
circ (P2+P3)
skip (P5+P6)
Exon1-2 (P1+P2)
Exon4-5 (P3+P4)
(bp)
cDNA
wt
RCM1
RCM2
1500
1000
500
100
B
N2
Y20F4.4 (RCM)
Relative level
****
****
0.63%
0.10%
circ-Y20F4.4
Y20F4.4-skip
L-Y20F4.4
C
N2
arl-13 (RCM)
***
n.d.
circ-arl-13
arl-13-skip
L-arl-13
F
N2
zip-2 (RCM1)
****
****
6.08E-05
6.17E-05
circ-zip-2
zip-2-skip
L-zip-2
G
N2
zip-2 (RCM2)
****
****
4.60E-05
3.79E-05
circ-zip-2
zip-2-skip
L-zip-2

Figure 5

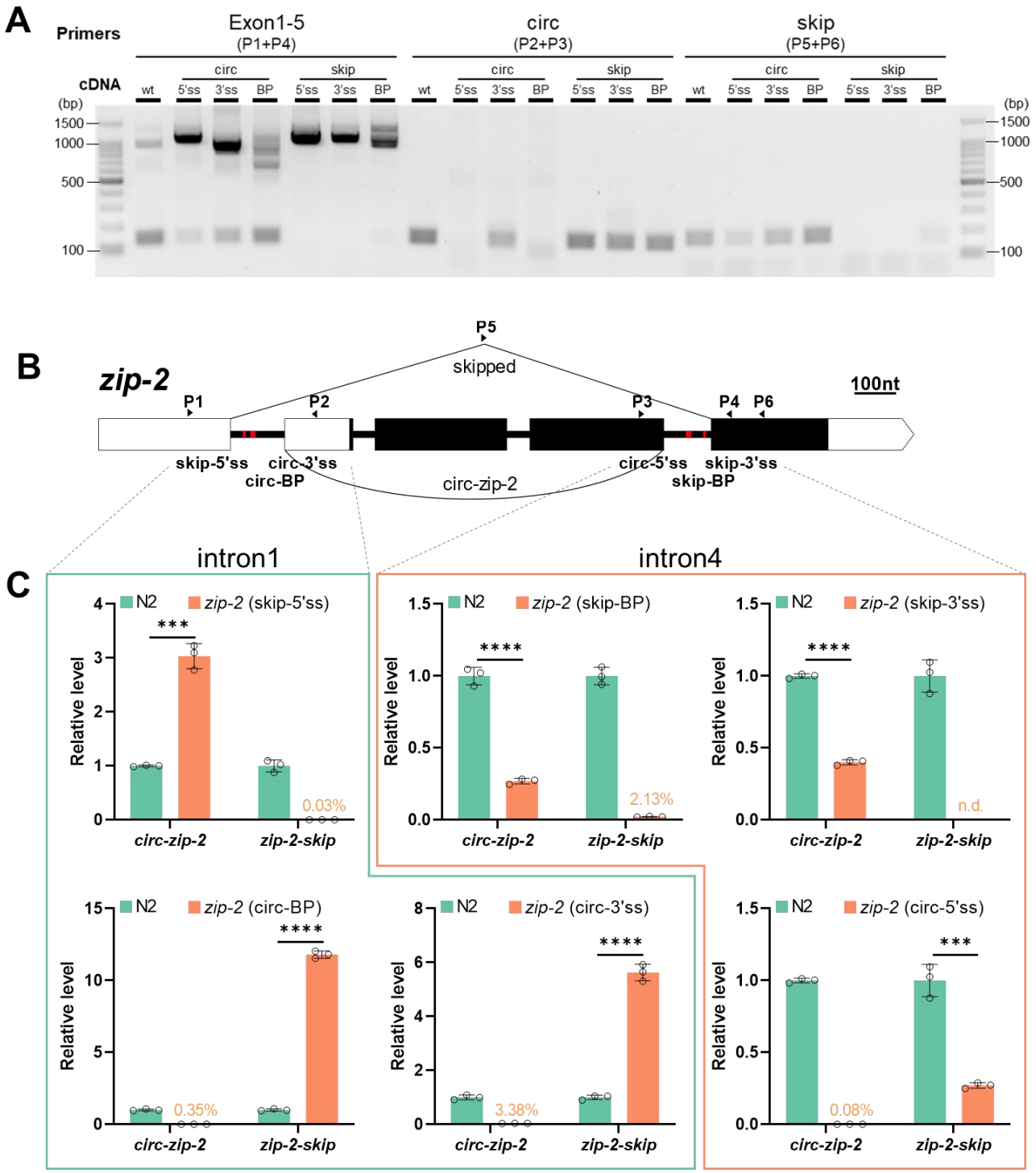
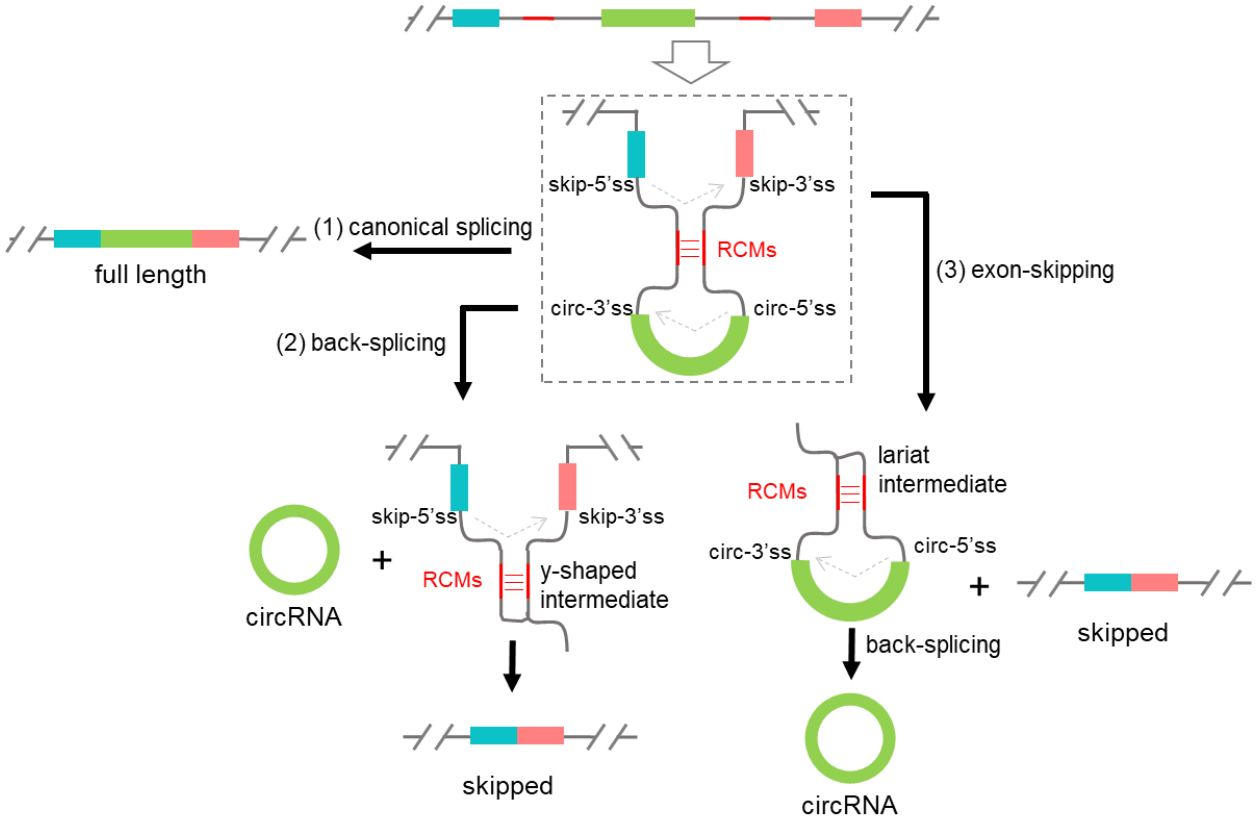
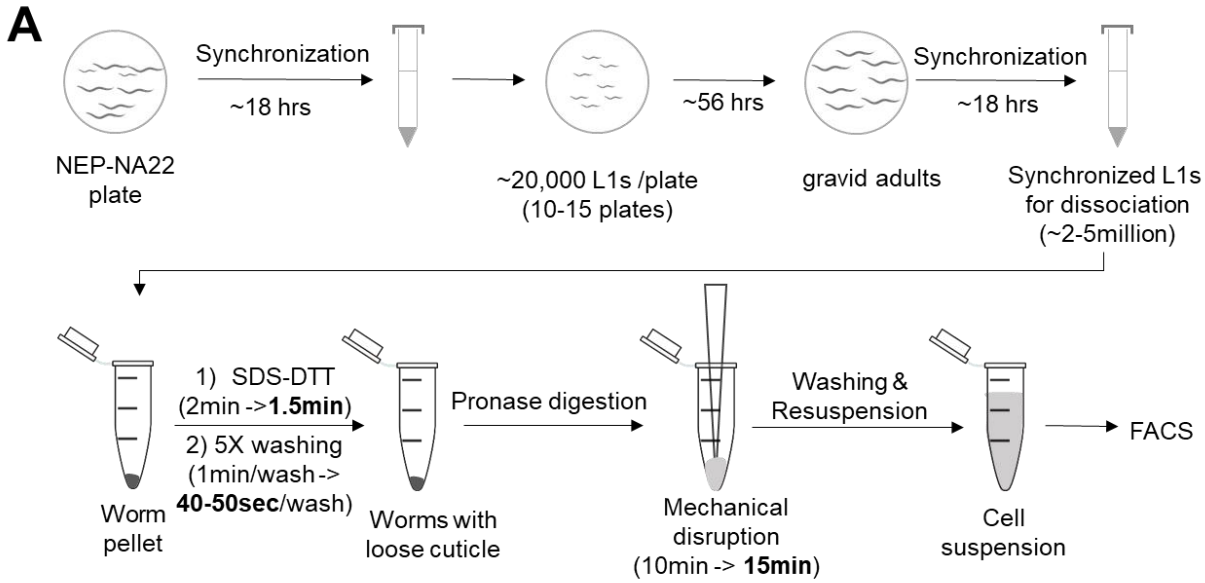
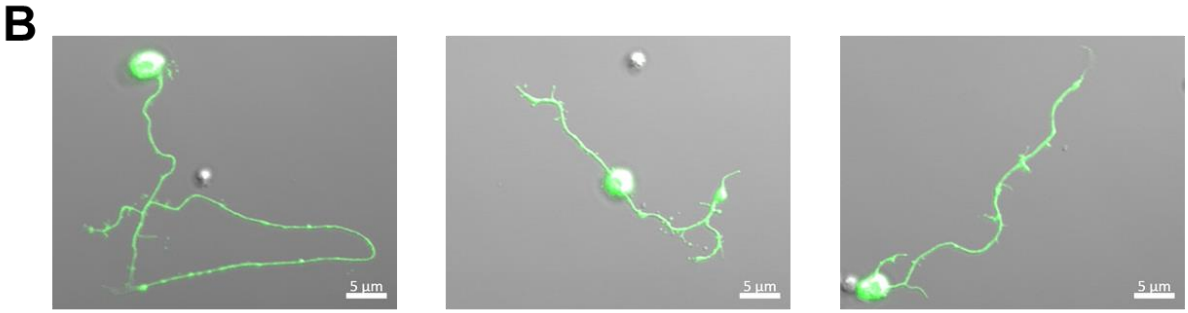

A
Primers
Exon1-5
(P1+P4)
circ
(P2+P3)
skip
(P5+P6)
cDNA
circ
skip
wt
5'ss
3'ss
BP
(bp)
1500
1000
500
100
B
zip-2
P5
skipped
P1
P2
P3
P4
P6
100nt
skip-5'ss
circ-3'ss
circ-BP
circ-zip-2
circ-5'ss
skip-3'ss
skip-BP
C
intron1
intron4
N2
zip-2 (skip-5'ss)
zip-2 (skip-BP)
zip-2 (skip-3'ss)
zip-2 (circ-BP)
zip-2 (circ-3'ss)
zip-2 (circ-5'ss)
Relative level
circ-zip-2
zip-2-skip
0.03%
2.13%
n.d.
0.35%
3.38%
0.08%

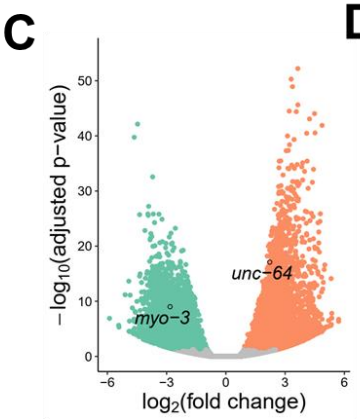
Figure 6


skip-5'ss
skip-3'ss
RCMs
circ-3'ss
circ-5'ss
(1) canonical splicing
full length
(2) back-splicing
(3) exon-skipping
circRNA
+
skip-5'ss
skip-3'ss
RCMs
y-shaped
intermediate
skipped
RCMs
lariat
intermediate
circ-3'ss
circ-5'ss
+
skipped
back-splicing
circRNA

Supporting Figures







D

Category	Term(select/unselect all)	Counts	ListSize	PopHit	Pop Size	Pvalue	Bonferroni	FDR
Tissue	Larval Pan-neural enriched	970	4990	1512	29469	0.0	0.0	0.0
Tissue	WT-Pico Pan-neural Enriched Genes	773	4990	1579	29469	4.2e-202	3.9e-200	2.0e-200
Tissue	Larval A-class enriched	215	4990	398	29469	9.6e-64	8.9e-62	3.0e-62
Tissue	panneural-enriched-core	88	4990	100	29469	2.9e-53	2.7e-51	6.7e-52
Tissue	Dopaminergic neuron enriched core	100	4990	149	29469	3.2e-41	3.0e-39	6.0e-40
Tissue	A-class Neuron enriched core	81	4990	106	29469	4.3e-40	4.0e-38	6.6e-39
Tissue	Neural Enriched without EST confirmation	143	4990	310	29469	1.4e-32	1.3e-30	1.8e-31
Tissue	Male-enriched genes (Kim)	442	4990	1750	29469	1.0e-19	9.5e-18	1.2e-18
Tissue	GABA-enriched-core	37	4990	48	29469	1.3e-18	1.2e-16	1.3e-17
Tissue	unc-4 enriched, motor neurons	272	4990	967	29469	1.5e-18	1.4e-16	1.4e-17
Tissue	L1 DD neurons	84	4990	201	29469	2.7e-16	2.5e-14	2.3e-15
Tissue	muscle depeleted 0h only	244	4990	888	29469	1.9e-15	1.8e-13	1.5e-14
Tissue	Muscle depleted genes	555	4990	2462	29469	6.3e-14	5.9e-12	4.5e-13
Tissue	Muscle Depleted 0h and 24H	214	4990	797	29469	1.3e-12	1.2e-10	8.8e-12
Tissue	Ectoderm genes	143	4990	479	29469	2.9e-12	2.7e-10	1.8e-11
Tissue	Sperm enriched	145	4990	568	29469	1.9e-07	1.8e-05	1.1e-06
Tissue	Sex-regulated neuronal genes	19	4990	38	29469	1.3e-05	0.0012	7.3e-05
Tissue	Enriched ASER vs. ASEL neurons	32	4990	100	29469	0.00038	0.036	0.0020
Tissue	Muscle Depleted 24H	311	4990	1574	29469	0.0018	0.16	0.0086
Tissue	Over-expressed in blastomeres compard to whole embryo	21	4990	69	29469	0.0089	0.83	0.041

Figure S1: related to Figure 1

(A) Steps of L1 worm preparation and dissociation for FACS. Optimized conditions are in bold.
(B) Confocal images showing sorted neurons after five-day culture at 20°C. Scale bars are 5 μm.
(C) Volcano plot showing differentially expressed genes between the sort group and the whole group. *myo-3* and *unc-64* are labeled.
(D) Output from WormExp for gene set enrichment search using upregulated genes in the neurons in our dataset. Red rectangle highlights the top 4 hits.

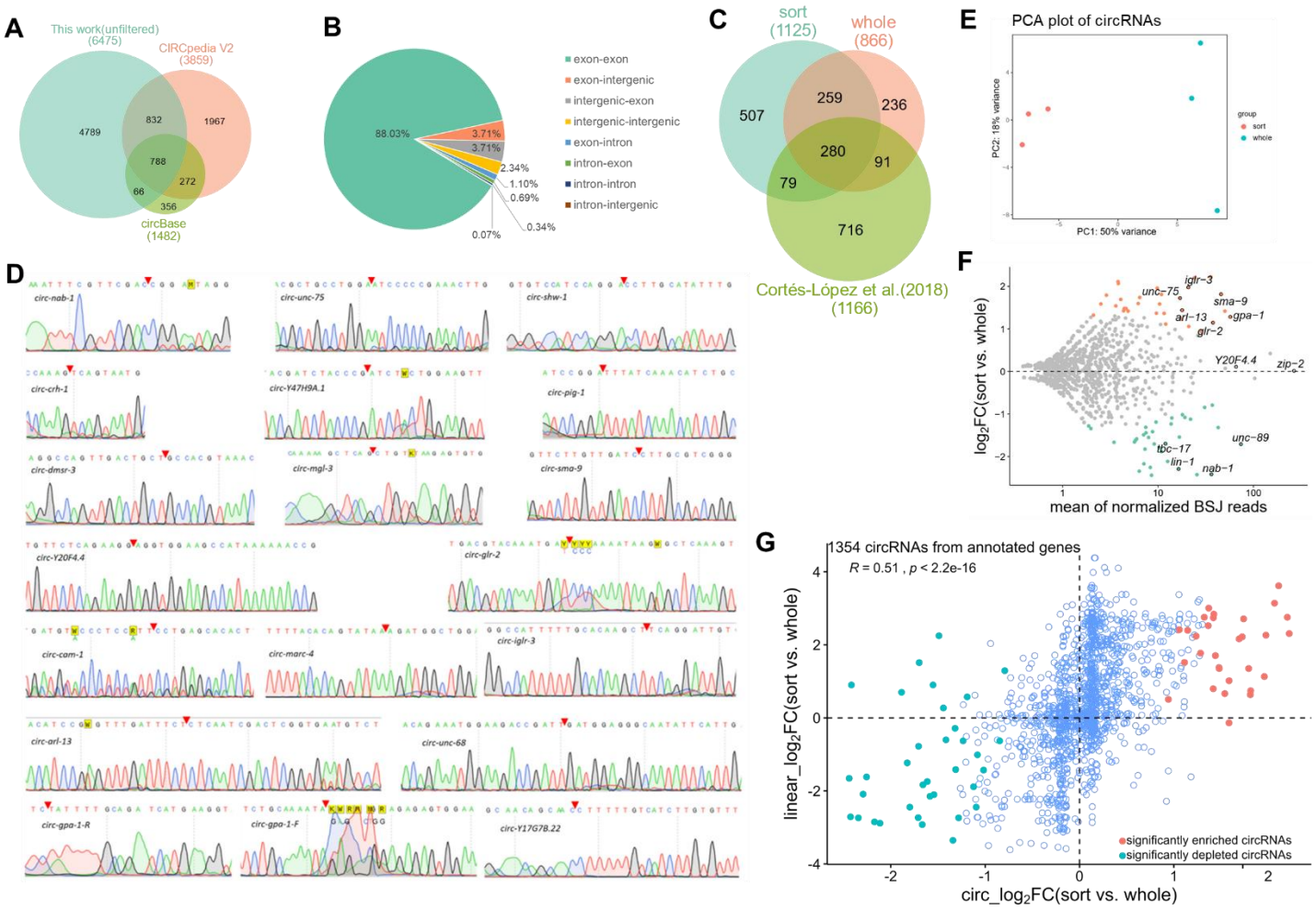


Figure S2: related to Figure 2.

(A) Overlap of circRNAs identified in this work and circRNAs of *C. elegans* in two databases (CIRCpedia V2 and circBase).

(B) Ratios of junction types of filtered circRNAs.

(C) Overlap of filtered circRNAs in this work and filtered circRNAs in work of Cortés-López et al (*1*).

(D) Sanger sequencing results of the BSJ sequences of selected circRNAs. Red triangles denote the joint sites.

(E) PCA plot of circRNAs in the sort group and the whole group.

(F) MA plot showing differentially expressed (DE) circRNAs between the sort group and the whole group. Significantly DE circRNAs are highlighted by colors. The gene names of some circRNA genes are labeled.

(G) Scatter plot showing the correlation of log2 fold change of circRNAs and their cognate linear RNAs in the sort group and whole group. The Pearson correlation coefficient (*R*) and *p* value (*p*) are shown. Significantly DE circRNAs are shown by colored dots.

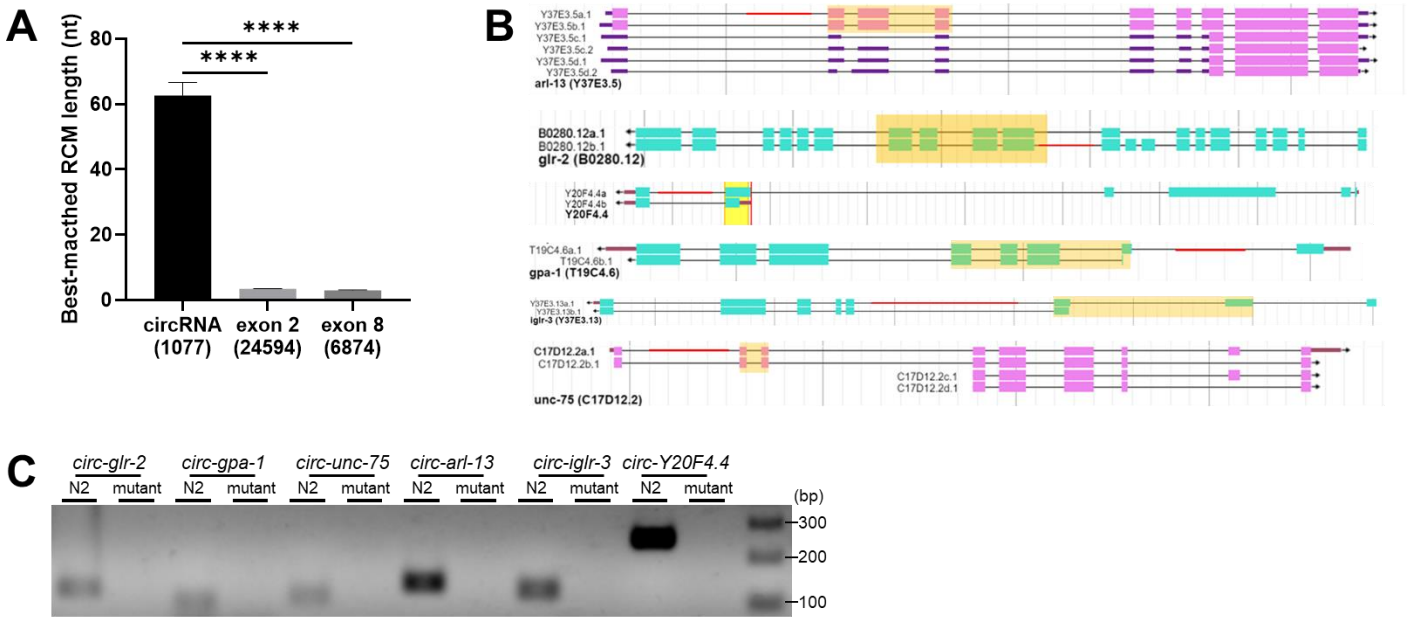
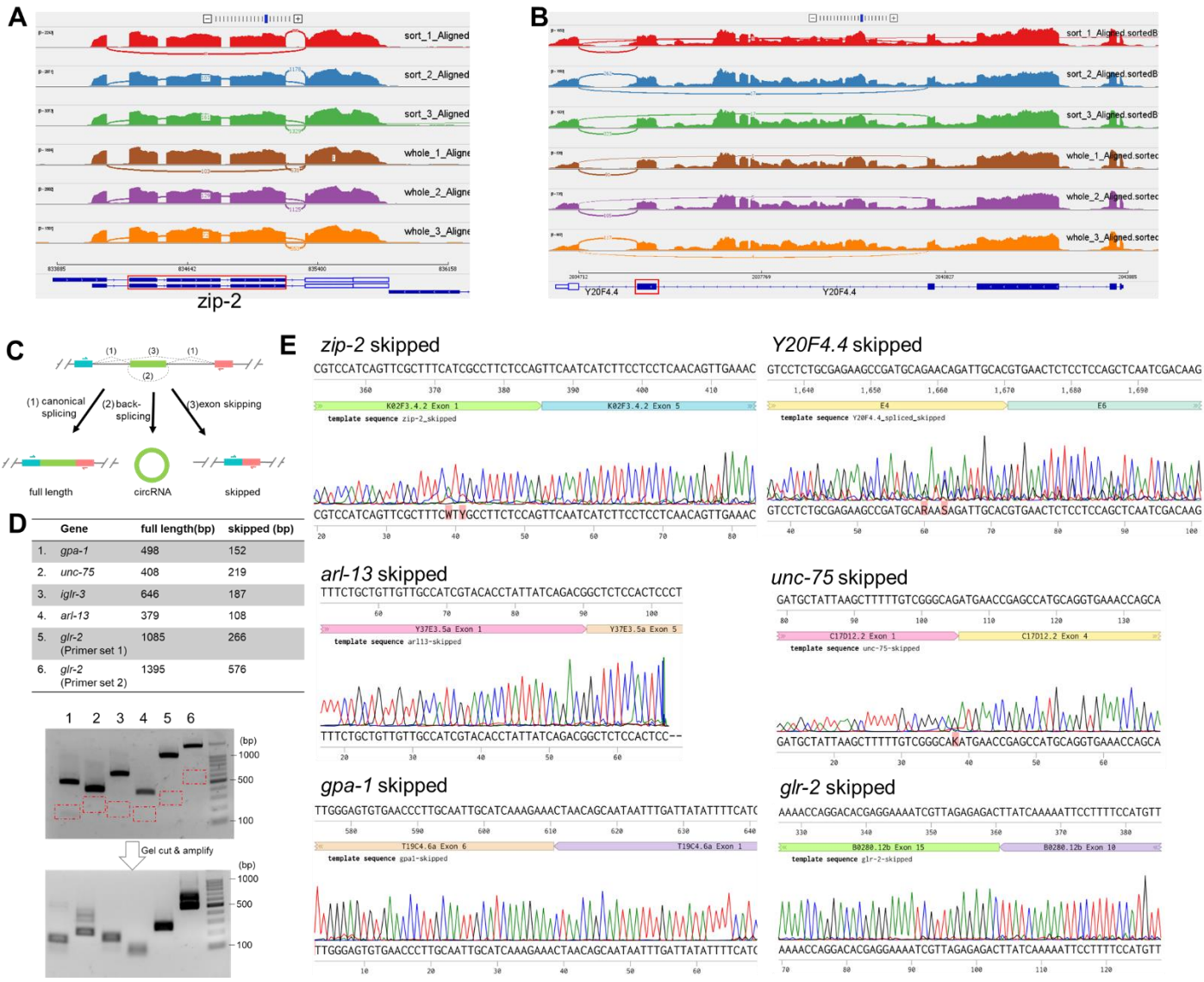


Figure S3: related to Figure 3.

(A) Lengths of best-matched RCMs in one pair of circRNA-flanking introns compared with those in non-circRNA control exons (exon 2 and exon 8). Values are shown as mean ± SEM. Numbers in the brackets are numbers of intron pairs for analysis. *p* values are from Kruskal-Wallis test with Dunn's post-hoc test for multiple comparisons. ****, $p < 0.0001$.

(B) Positions of deleted RCMs (red line) of the 6 circRNA genes. Exons in orange shadows are to form circRNAs.

(C) Amplification of circRNA using divergent primers in wildtype N2 strain and RCM-deletion strains (mutant) of 6 circRNA genes.



	Gene	full length(bp)	skipped (bp)
1.	*gpa-1*	498	152
2.	*unc-75*	408	219
3.	*iglr-3*	646	187
4.	*arl-13*	379	108
5.	*glr-2* (Primer set 1)	1085	266
6.	*glr-2* (Primer set 2)	1395	576

Figure S4: related to Figure 4.

(A & B) Sashimi plots showing the number of exon-skipping junction reads in *zip-2* and *Y20F4.4* in our RNA-seq dataset. Exon(s) in red rectangles are circularized to form circRNAs.

(C) Illustration of a circRNA-producing gene producing 3 transcripts: full-length, circular and skipped. Primers used to detect both the full length and the skipped transcripts are shown.

(D) Amplification of the skipped transcripts from several circRNA genes by 2-round PCRs. Red rectangles mark the gel areas to be cut.

(E) Confirmation of sequences of the skipped transcripts in 6 circRNA genes.

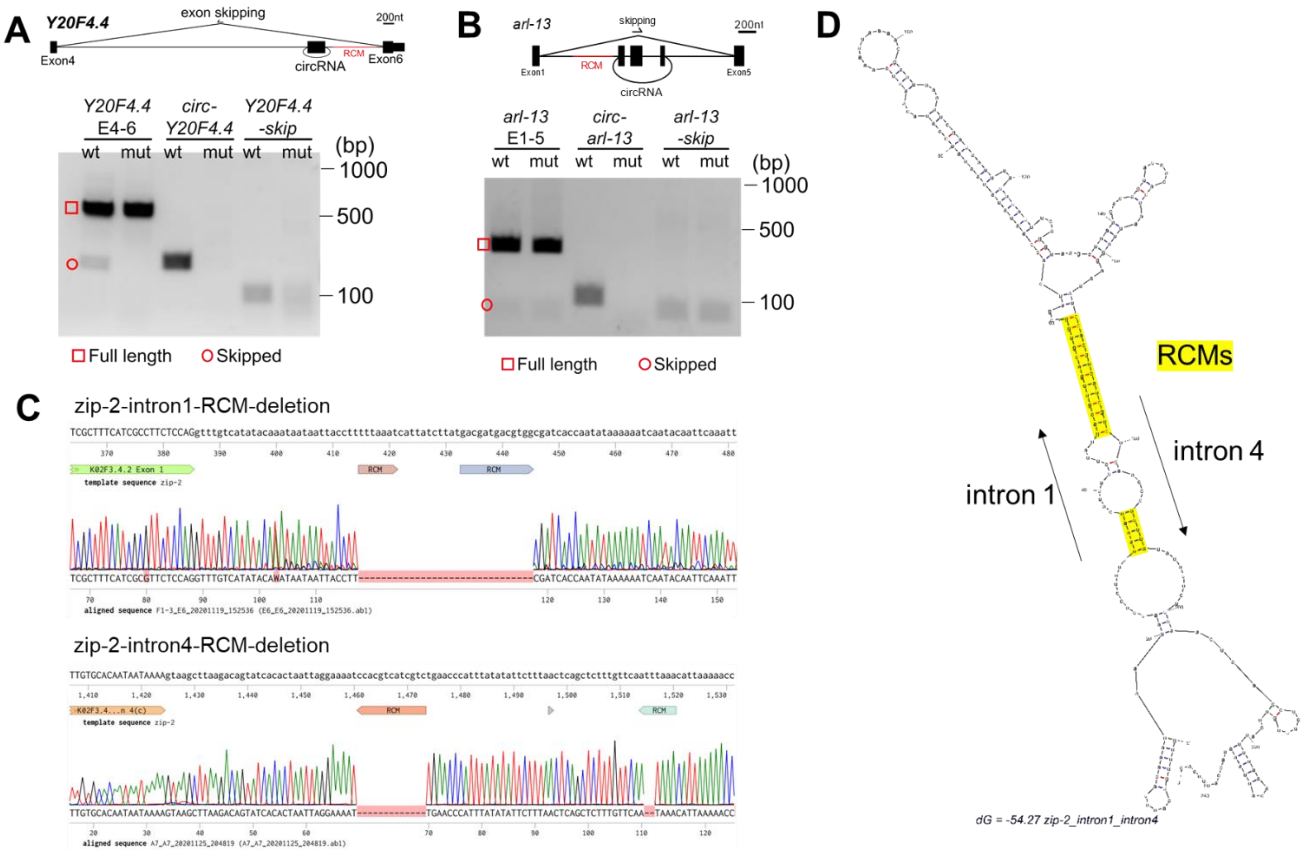


Figure S5: related to Figure 4.

(A) RT-PCR detection of *Y20F4.4* transcripts in wildtype N2 strain (wt) and the RCM-deleted *Y20F4.4* strain (mut).

(B) RT-PCR detection of *arl-13* transcripts in wildtype N2 strain (wt) and the RCM-deleted *arl-13* strain (mut).

(C) Deleted RCM sequences in intron 1 and intron 4 of *zip-2*.

(D) Folding prediction of intron 1 and intron 4 of *zip-2* by Mfold (http://www.unafold.org/mfold/applications/rna-folding-form.php). RCM sequences are highlighted.

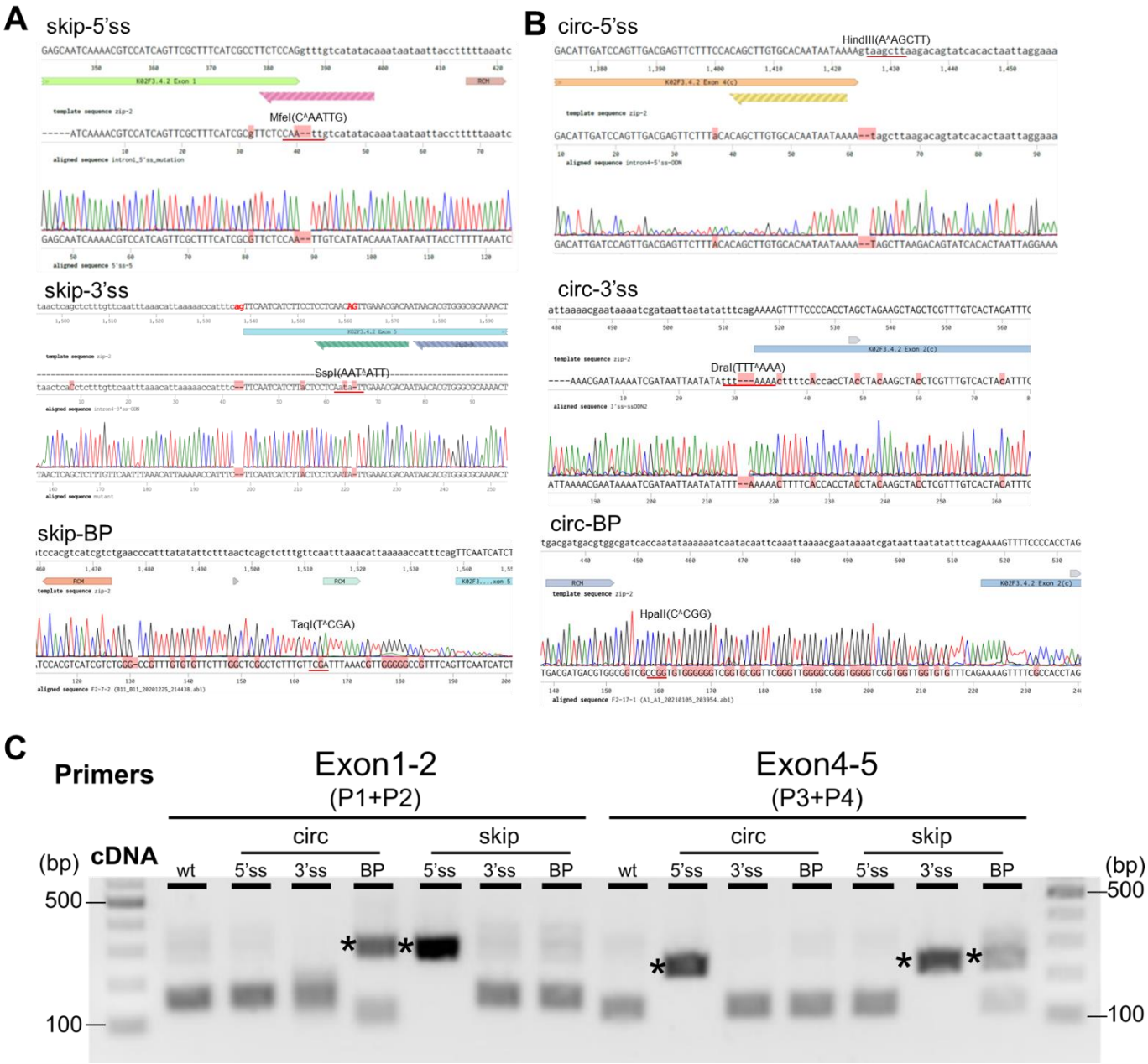


Figure S6: related to Figure 5.

(A, B) Sanger sequence results of splicing sites and branch points mutation in intron 1 and intron 4 of *zip-2*. The enzyme digestion sites used to distinguish wildtype sequences and mutated sequences are labeled.

(C) RT-PCR results of zip-2 intron 1 splicing and zip-2 intron 4 splicing in wildtype and the ss or BP mutated strains. *: intron retention.

References:

1. M. Cortes-Lopez *et al.*, Global accumulation of circRNAs during aging in Caenorhabditis elegans. *BMC Genomics* **19**, 8 (2018).