ISCI, Volume *20*

Supplemental Information

Long and Repeat-Rich Intronic Sequences Favor Circular RNA Formation under Conditions of Reduced Spliceosome Activity

Mantian Wang, Jingyi Hou, Michaela Müller-McNicoll, Wei Chen, and Erin M. Schuman

Supplemental Figures

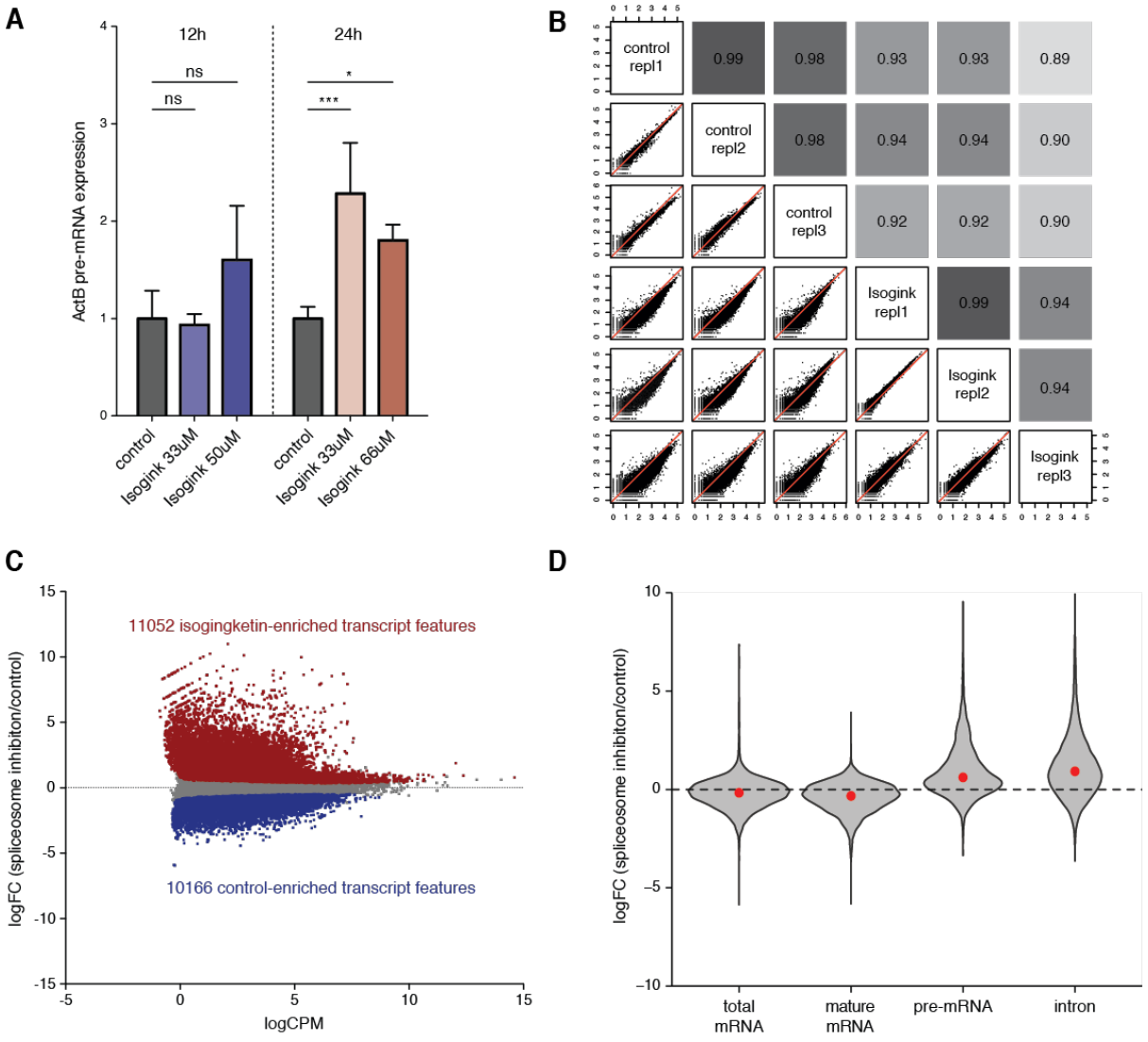

Figure S1 (related to Figure 1). Optimization of Isoginkgetin treatment conditions, reproducibility of RNA-seq and transcript feature expression. (A) qRT-PCR validation of the efficiency of spliceosome inhibition by treating primary neurons with different Isoginkgetin concentrations and durations. Strongest increase of ActB pre-mRNA measured after 24h treatment with 33μM Isogingketin (***p < 0.0001, *p < 0.05, one-way ANOVA followed by Bonferroni multiple comparisons test, n = 3). Error bars represent s.d. **(B)** Correlation of RNA-sequencing between replicates and conditions. Values on x- and y-axis denote the log10 transformed raw mRNA counts (Pearson's correlation). **(C)** Differential expression of transcript features in Isoginkgetin vs control condition. Red dots indicate significantly up-regulated and blue dots significantly down-regulated transcript features. Grey dots show transcript features that are un-changed. **(D)** Violin plot showing the fold-change distribution of total mRNA, mature mRNA, pre-mRNA and introns. Red dots indicate median.

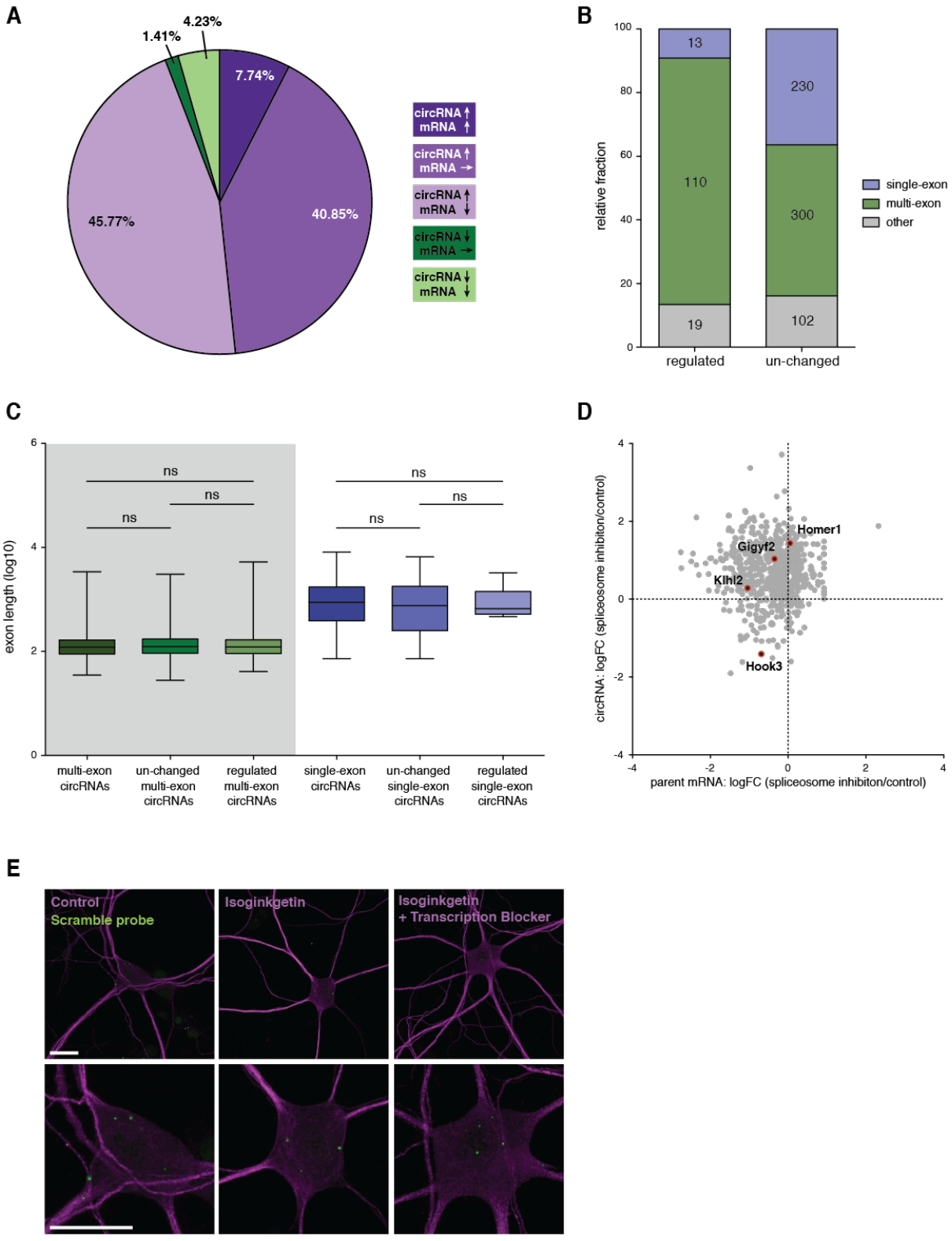

Figure S2 (related to Figure 2). Quantification of circRNA and parent mRNA expression changes, exon features of regulated circRNAs, identification of validation targets and FISH probe specificity. (A) Pie chart showing the fraction of significantly up- and down-regulated circRNAs grouped according to the direction of the circRNA and parent mRNA modulation. **(B)** Bar graph showing the fraction of circRNAs in regulated and un-changed groups that consists of a single exon, multiple exons or mixed introns and exons. **(C)** Exon length distribution of single- and multi-exon circRNAs between regulated and un-changed groups (ns > 0.05, Kruskal-Wallis test followed by Dunn's multiple

comparisons test). **(D)** Scatterplot of log2 fold-change between circRNA expression (x-axis) and parent mRNA expression (y-axis). Red dots indicate the four candidates choosen for qRT-PCR validation. **(E)** High-resolution *in situ* hybridization of scrambled negative control probe in control, Isoginkgetin, and Isoginkgetin and transcription-inhibitor treated neurons. Neuronal somata and dendrites were identified using MAP2-immunostaining. Scale bar = 25 µm. Box-plot whiskers show minimum, first quartile, median, third quartile, and maximum of a set of data **(C)**.

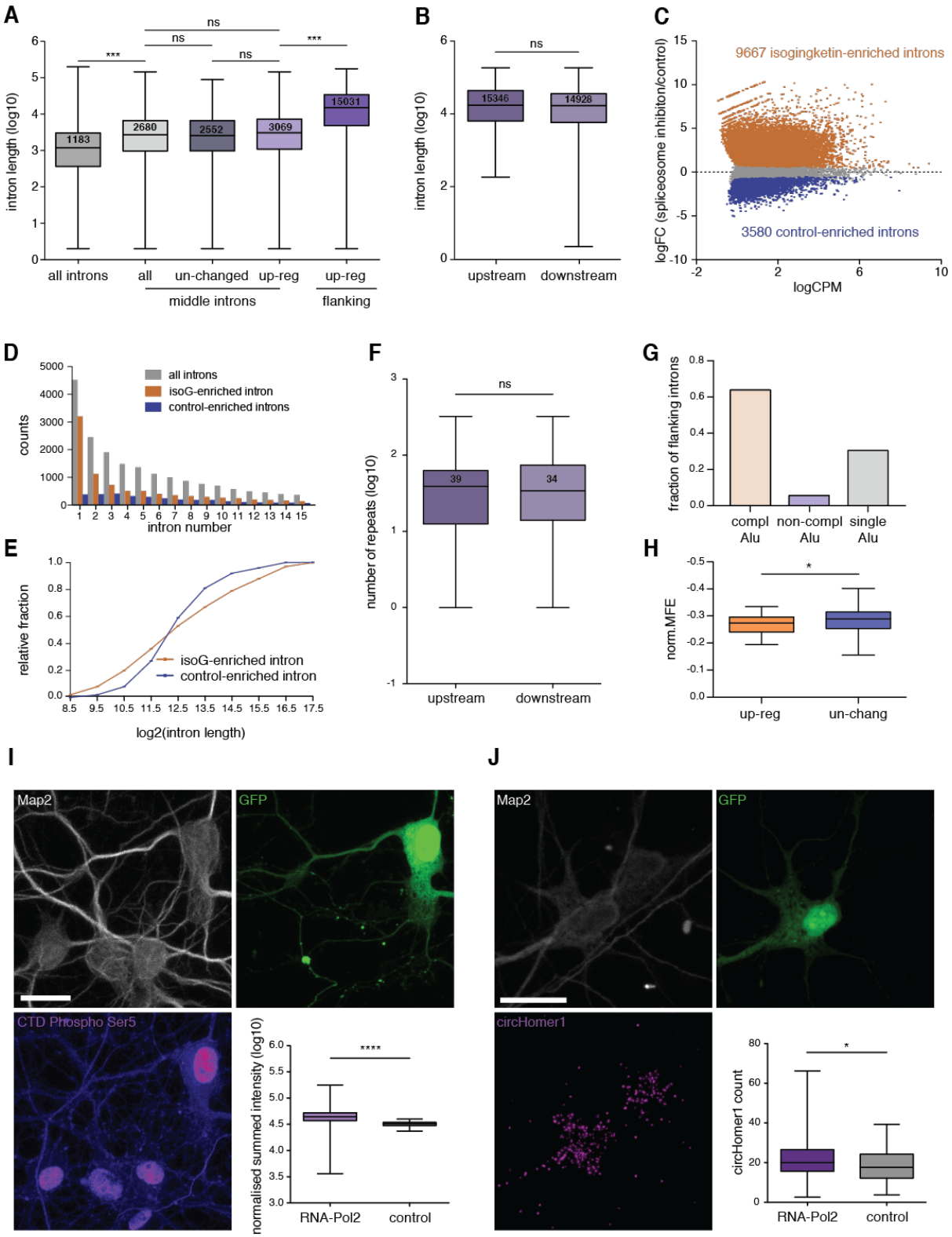

Figure S3 (related to Figure 3). Characteristics of circRNA flanking introns and differentially expressed introns, and RNA-Pol2 over-expression analysis (A) Length distribution of middle introns compared to all introns and flanking introns of up-regulated circRNAs. No significant difference in middle intron length was observed between all circRNA middle introns, middle introns of un-changed circRNAs and up-regulated circRNAs (*** $p < 0.0001$, ns > 0.05, Kruskal-Wallis test followed by Dunn's multiple comparisons test). **(B)** Up- and downstream flanking introns of up-regulated circRNAs share a similar length (ns > 0.05, Wilcoxon-matched pairs signed rank test). **(C)** Differential expression analysis of individual introns between Isoginkgetin and control condition. Orange dots indicate significantly up-regulated introns and blue dots significantly down-regulated introns, respectively. Grey dots are un-changed introns. **(D)** Frequency of intron number occurrence of Isoginkgetin- (orange bars) and control-enriched introns (blue bars). Frequency of intron occurrence of all detected introns is shown in grey. **(E)** Cumulative frequency plot of binned intron length for Isoginkgetin- and control-enriched introns. **(F)** Up- and downstream flanking introns of up-regulated circRNA harbor similar number repeat elements (ns > 0.05, Wilcoxon-matched pairs signed rank test). **(G)** Fraction of flanking intron pairs that possessed complementary, non-complementary or single Alu repeats. (**H**) Structural properties of flanking introns of up-regulated (orange) and un-changed (blue) circRNAs. Un-changed circRNAs showed lower norm.MFE compared to up-regulated circRNAs (*$p < 0.05$, unpaired t-test, two-tailed, n = 66 and 63). **(I)** CTD Ser5 phospho expression (magenta) was significantly increased in the nucleus 4 days post-transfection of positively transfected neurons (green). Neuronal somata and dendrites were identified using MAP2-immunostaining (***$p < 0.0001$, Mann Whitney U test, n = 98 and 98). Scale bar = 25 µm. **(J)** RNA-Pol2 over-expression caused increased levels of circHomer1 (*$p < 0.005$, Mann Whitney U test, n = 45 and 92). Scale bar = 25 µm. Box-plot whiskers show minimum, first quartile, median, third quartile, and maximum of a set of data **(A, B, F, H, I, J)**.

Supplemental Table

	DMSO			Isoginkgetin [33uM]		
	Replicate 1	*Replicate 2*	*Replicate 3*	*Replicate1*	*Replicate2*	*Replicate 3*
total number of reads	48775167	45505590	79203700	46558089	45724427	21831475
number of mapped reads	46902334	42075498	75172588	42708588	42337404	20194243
number of uniquely mapped reads	38566396	30894509	48283505	25809414	29450692	14387638
exon-body reads	17912355	14989485	22172891	9818742	11423787	5255304
exon-exon junction reads	7656382	6343313	9500829	4012615	4492982	2194226
intron body reads	3968750	2614698	4949182	3145322	3505727	2217062
exon-intron junction reads	547044	481658	867525	563174	572267	319903
back-splice junction reads	24155	18060	25692	23244	27633	15986
number of circRNAs detected	14382	11066	14984	14520	16603	11946
number of circRNA host genes	6284	5384	6426	6014	6481	5238

Table S1 (related to Figure 1). Details on RNA-seq samples

circRNA	**Primers**
cHomer1	AACACCCGATGTGACACAGA GCTCGAGTGCTGAAGATAGGTT
cKlhl2	TGGACCCTGAGGATGCTAAT TCTGATGACCCTGCTTTGTG
cGigyf2	AAAGATGTAGGCTCCGTGCT TCGGCCATATCGATAATCTGCT
cHook3	ACAAGAGACAGACTAGAAGCAG CATCGTTCTGTTGCCGAAGC
mRNA	**Primers**
pre-ActB (1)	GCCAGTGCTGAGAACGTTGTT CGCCCACGTAGGAGTCCTT
pre-ActB (2)	TGTGGCTTTAGGAGCTTGAC CTGGGGTGTTGAAGGTCTC
ActB	GGGTATGGGTCAGAAGGACT GGGTACTTCAGGGTCAGGAT
pre-Atp5i	CCGGTACTCCGCTCTGAT CCTCCCAATCCCCAAACT
Atp5i	GGCAGAGGAGGAGAGAAGAA TCTCTCAATCCGTTTCAACTCA
Homer1	GAGCTGGAAGAGACCCTAAAAG TCAAAGAGTCCCTCTGTTCTTG
Klhl2	TGTGAAGAAGACATGCTGTGAA TTTATTCAGGAGGTCTGTGCAT
Gigyf2	TTGCTGAAAACCTCTTGCTGTG TGCTGCCATTCTTCTCCGTA
Hook3	GACATTTGCAACTTCAGACCCA

	GCCATCCTGGCACCCTCTAT

Table S2 (related to Figure 1 and 2). List of circRNA and mRNA qRT-PCR primers

circRNA	**Probes**
cHomer1	TTTCACATAGGGAACAACCT
scramble control	GTGTAACACGTCTATACGCCCA

Table S3 (related to Figure 2). Sequence of circRNA FISH probes

Transparent Methods

Primary hippocampal neurons. Hippocampi from postnatal day 0-1 rat pups of either sex (Sprague-Dawley stain; Charles River Laboratories) were dissected and dissociated with papain (Sigma). For RNA isolation, neurons were plated at a density of 700K onto a 6cm poly(D-lysine)-coated petri dish. For fluorescence *in situ* hybridization (FISH), 30K cells were plated onto a poly(D-lysine)-coated glass-bottom Petri dish (MatTek). Hippocampal neurons were maintained at 37°C and 5% CO2 in growth medium (Neurobasal-A supplemented with B27 and GlutaMax-I, Life Technologies).

Pharmacological treatments. To test the efficiency of spliceosome inhibition, neurons grown in 6cm petri dishes were treated at DIV28-30 i) with 33μM or 50μM Isoginkgetin (Merck, dissolved 0.1% v/v in DMSO) for 12h or ii) with 33μM or 66μM Isogingketin for 24h. Neurons treated with 0.1% v/v DMSO (VWR) for the matching duration of 12h or 24h were used as negative controls. Cells were scraped using Trizol (Invitrogen) followed by RNA extraction and qRT-PCR. The treatment with 33μM Isoginkgetin for 24h resulted in the strongest spliceosome inhibition and was applied for the RNA-seq and *in situ* hybridization experiments. To inhibit transcription and splicing, neurons were incubated with Isoginkgetin and a cocktail of 40μM Actinomycin D (Sigma), 1μM Triptotide (Sigma) and 50μM 5,6-Dichlorobenzimidazole (Sigma) for 24h.

FLAG-Pol2 over-expression analysis. Cultured hippocampal neurons were co-transfected using Magnetofectamine (OZ Biosciences) at DIV 9-12 with 0.5ug of pAcGFP1-N1 (Clontech) and 1ug of FLAG-Pol2-WT plasmid construct (a gift from Benjamin Blencowe (Addgene plasmid #35175, Rosonina and Blencowe, 2004)). Immuno-staining against RNA-Polymerase II CTD (Abcam ab5131, 1:2000 dilution) and high-resolution *in situ* hybridization were performed 4 days post-transfection to quantify RNA-Polymerase II and circHomer1 expression, respectively.

SF3A2 and SF3B1 knock-down analysis. Cultured hippocampal neurons were transfected at DIV 9-12 with 1ug of SF3A2, SF3B1 or scrambled control shRNA-plasmid constructs (OriGene) using Magnetofectamine (Oz Biosciences). Knock-down efficiency was quantified after 4 days by immunostaining against SF3A2 (Thermofisher PA5-61969, 1:500 dilution) or SF3B1 (Abcam ab172634, 1:500 dilution). To quantify circHomer1 expression high-resolution *in situ* hybridization was performed.

RNA isolation and RNA-sequencing. Neurons were scraped using Trizol (Invitrogen) and RNA was extracted following the manufacturer's protocol. RNA clean up including the on-column DNase I digest was performed using the RNeasy Mini kit (QIAGEN). Ribosomal RNA was depleted using the RiboZero Gold kit (Epicentre Bio-technologies). RNA-seq library was then generated using Illumina stranded RNA Sample Prep kit according to the manufacturer's instructions and sequenced for 150nt from single end on an Illumina HiSeq2000.

Transcript feature identification and differential expression. After removing the Illumina sequencing adaptor, the reads were aligned to the rat (rn5.0) genome reference sequences using STAR aligner (Dobin et al., 2013). We used the BEDTool software suit (Quinlan and Hall, 2010) to annotate each read by intersecting its genomic coordinates with the Ensembl gene annotation for rat (rno5.0). To identify and count the reads that mapped to the exon- and intron-body, as well as exon-exon and exon-intron junction, we applied a custom script. To evaluate the differential expression of total mRNA, pre-mRNA, mature mRNA and introns between the Isoginkgetin and control conditions, we input the summed exon-exon and exon-intron junction reads, exon-intron junction reads, exon-exon junction reads and intron-body reads per transcript into the edgeR software (Robinson et al., 2010). Transcript features were filtered (CPM >1, detected in 4 out of 6 replicates) and library size adjusted. Model fitting and testing was performed using the exactTest function. A false discovery rate (FDR) of < 0.05 was used to determine differentially expressed events.

circRNA identification and differential expression. Using a custom script, chimeric aligments from a Chimeric.out.junction file were converted into BED12 files. The BEDTools software suit was employed to annotate the reads by intersecting their genomic coordinates with the Ensembl gene annotation for rat (rno5.0). If partial aligned segments within a chimeric read were i) at least 20bp long, ii) mapped to the same chromosome and strand, iii) showed the presence of a splice junction and iv) originated from the same gene but in reversed order, they were kept as back-splice junction read supporting the expression of a circRNA. Back-splice junction reads mapping to the X and Y chromosome, as well as mitochondrial chromosome were removed. Only circRNAs that were present

with at least 1 unique read in all replicates and conditions were used for further analyses. To evaluate the differential expression of circRNAs between the Isoginkgetin and control conditions, we input the back-splice junction reads and exon-exon junction reads of the parent mRNAs into the edgeR software. Model fitting and testing was performed using the exactTest function. Due to the low number of reads that were used as input, we decided to use a *P value* instead of false discovery rate (FDR) to determine differentially expressed events. The significance threshold value was set to 0.05.

Exon and intron feature analyses. The length of all rat exons and introns were determined using the Ensembl gene annotation for rat (rno5.0). The identity of the exons spanning the back-splice junction was used i) to distinguish between circRNAs comprised of a single exon or multiple exons, and ii) to identify the upstream and downstream flanking introns. A separate list of all introns located between the upstream and downstream flanking intron (a.k.a middle introns) was generated. The length of circRNA exons, flanking introns and middle introns were determined by cross-referencing their identity with the list containing the length of all rat exons and introns. Differentially expressed introns were identified as described above. However, instead of using intron-body reads per transcript, we applied the counts of each intron, together with all other transcript features as an input into the edgeR software. Length for up- and down-regulated introns was determined using the Ensembl gene annotation for rat (rno5.0).

Repeat feature analyses. A list of repeat family and class, and their coordinates in the rat genome were obtained using the repeatMasker track of the UCSC table browser. Subsequently, the number of repeat families present in each circRNA flanking intron and the frequency of repeat family occurrence per 1kb intron length was determined. To calculated the average number of repeats per intron, we normalized the total number of each detected repeat family to the number of introns for up-regulated, un-changed and all circRNAs.

Reverse complementary motif (RCM) analysis. Identification of reverse complementary motifs was conducted as previously described (Ivanov et al., 2015; Zhang et al., 2014). In short, intron alignments using BLAST (Altschul et al., 1990) were carried with the parameters "blastn -word_size 7 -gapopen 5 -gapextend 2 -penalty -3 -reward 2" and only local alignment with a BLAST score >25 were kept.

Alu repeat analysis. Orientation of Alu repeats in flanking intron pairs was determined as described previously (Jeck et al., 2013). In brief, paired flanking introns were analyzed for repeatMasker Alu elements using the BEDTools software suit. When at least one plus and one minus stranded Alu family element were detected on either side of the flanking introns a complementary Alu pair was identified. Alu repeats that showed the same orientation were termed non-complementary. Single Alu repeats were designated as those that were found in one of the flanking introns but not both.

Minimum free energy calculation. RNA-fold was used to calculate the minimum free energy per intron sequence (Lorenz et al., 2011). Only sequences with a size ranging from 100-10000nt were included into the analysis. Obtained MFE values were normalized to intron length.

RNase R treatment. Total RNA (3µg) was incubated for 45 min at 37°C with 10 U of RNase R (Epicentre). RNA was subsequently purified using the RNA Clean and Concentrator kit (Zymo Research) according to manufacturer's protocol. Reverse transcription was performed using the QuantiTect Reverse Transcription kit (QIAGEN).

cDNA synthesis and Quantitative real-time PCR. RNA was isolated as described above. Reverse transcription was performed using the QuantiTect Reverse Transcription kit (QIAGEN). For circRNA detection, the primers were designed to amplify the back-splice junction. To quantify the parent mRNA, the primers were designed to amplify the exon-exon junction upstream or downstream of the back-splicing exons.

High-resolution *in situ* hybridization in primary hippocampal neurons. Cultured neurons were fixed for 25min in 4% paraformaldehyde/PBS. The *in situ* hybridization was performed using the ViewRNA miRNA ISH Cell Assay kit (Thermo Fisher Scientific) following the manufacturer's protocol, omitting the dehydration/rehydration step as well as the protease treatment. Dendrites were then stained with an anti-MAP2 antibody (Millipore AB5622, 1:1000 dilution).

Image acquisition and processing. Confocal microscopy was performed using a Zeiss LSM880 confocal laser fluorescence microscope system. Maximum intensity projections of image series of 20-32 confocal planes taken at 0.485µm intervals using a 40x oil immersion objective were used for image analysis. The circRNA *in situ* signal in the cell body was quantified using a custom MATLAB script and normalized to cell body area.

Statistical analysis. After pretesting the normality of each dataset using the D'Agostino-Pearson-test and Kolmogorov-Smirnov-test, the following statistical analyses were conducted: two-tailed unpaired t-test (**Figs.1B,C, E** and **2A,B,C**), Mann-Whitney U-test (**Figs.2F,G,H**, **3D and suppl.Fig.3I, J**), Kruskal-Wallis test followed by Dunn's multiple comparisons test (**Figs.1G**, **2E**, **3A,B,C** and **suppl. Figs. 2C, 3A**), one-way ANOVA followed by Bonferroni multiple comparisons test (**suppl.Fig.1A**) and Wilcoxon-matched pairs signed rank test (**suppl.Fig3B,F**). In **Figs.1G, 2A,E,F,G,H, 3A,B,C,D** and **suppl.Figs.2C, 3A,B,F,I,J** box-whiskers plots show minimum, first quartile, median, third quartile, and maximum.

Data and Software availability.
The accession number for the raw sequencing data reported in this paper is NCBI BioProject: PRJNA561523. Processed data used for analyses in this manuscript are included in Data S1. All scripts used are briefly described in methods, and will be available upon request.

Supplemental References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.

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

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) *26*, 139-140.

Rosonina, E., and Blencowe, B.J. (2004). Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. RNA (New York, NY) *10*, 581-589.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) *26*, 841-842.