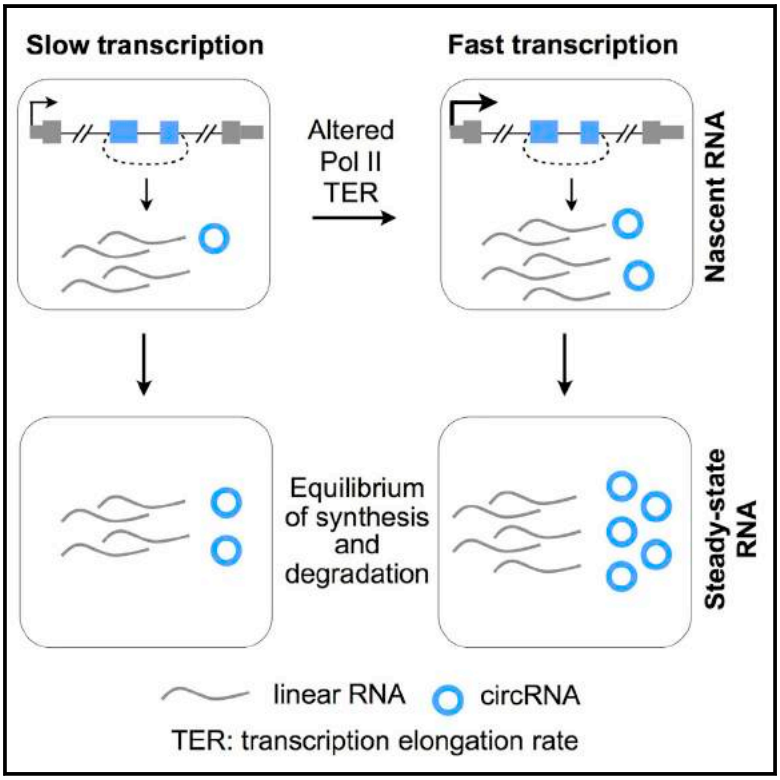
Article

Cell Reports

The Biogenesis of Nascent Circular RNAs

Graphical Abstract



Authors

Yang Zhang, Wei Xue, Xiang Li, ...,
Jia-Lin Zhang, Li Yang, Ling-Ling Chen

Correspondence

liyang@picb.ac.cn (L.Y.),
linglingchen@sibcb.ac.cn (L.-L.C.)

In Brief

Zhang et al. study the link between circRNA processing and transcription using 4sUDRB-seq. They find that circRNA production from pre-mRNA back-splicing is slow and largely occurs post-transcriptionally. The authors argue that circRNAs that are abundant at a steady-state level tend to be transcribed quickly and accumulate.

Highlights

- The efficiency of circRNA processing from pre-mRNA is very low in cells
- CircRNA processing correlates positively with Pol II elongation rate
- CircRNA processing largely occurs post-transcriptionally
- Fast transcription and accumulation lead to circRNA upregulation in neurons

Accession Numbers

GSE73325

Zhang et al., 2016, Cell Reports *15*, 611–624
April 19, 2016
http://dx.doi.org/10.1016/j.celrep.2016.03.058

Cell Reports
Article

The Biogenesis of Nascent Circular RNAs

Yang Zhang,[1,4] Wei Xue,[2,4] Xiang Li,[1] Jun Zhang,[1] Siye Chen,[1] Jia-Lin Zhang,[1,2] Li Yang,[2,3,*] and Ling-Ling Chen[1,3,*]
[1]State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[2]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[3]School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China
[4]Co-first author
*Correspondence: liyang@picb.ac.cn (L.Y.), linglingchen@sibcb.ac.cn (L.-L.C.)
http://dx.doi.org/10.1016/j.celrep.2016.03.058

SUMMARY

Steady-state circular RNAs (circRNAs) have been mapped to thousands of genomic loci in mammals. We studied circRNA processing using metabolic tagging of nascent RNAs with 4-thiouridine (4sU). Strikingly, the efficiency of circRNA processing from pre-mRNA is extremely low endogenously. Additional studies revealed that back-splicing outcomes correlate with fast RNA Polymerase II elongation rate and are tightly controlled by *cis*-elements in vivo. Additionally, prolonged 4sU labeling in cells shows that circRNAs are largely processed post-transcriptionally and that circRNAs are stable. Circular RNAs that are abundant at a steady-state level tend to accumulate. This is particularly true in cells, such as neurons, that have slow division rates. This study uncovers features of circRNA biogenesis by investigating the link between nascent circRNA processing and transcription.

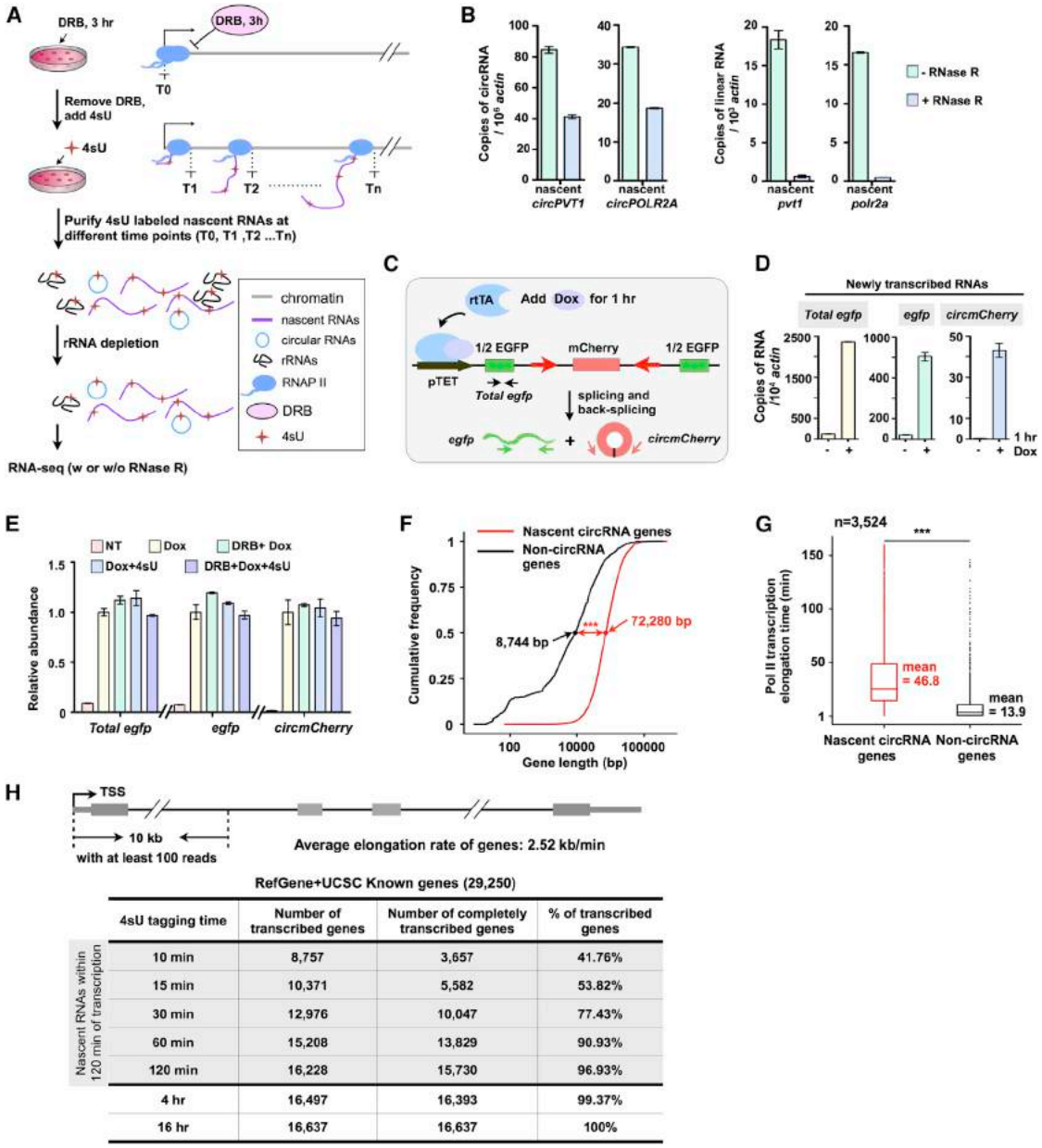
INTRODUCTION

Circular RNAs (circRNAs) are produced from pre-mRNA back-splicing. During back-splicing, a downstream 5′ splice site is joined to an upstream 3′ splice site in a reversed orientation, resulting in a circular RNA molecule with a 3′,5′ phosphodiester bond at the back-splicing junction site. Over 10,000 circRNAs have been identified in metazoans (Salzman et al., 2012, 2013; Jeck et al., 2013; Memczak et al., 2013; Westholm et al., 2014; Zhang et al., 2014; Ivanov et al., 2015).

Although the majority of circRNAs still lack functional annotation, recent observations are beginning to reveal that circRNAs may play potentially important roles in gene regulation (Chen, 2016). For example, a few abundant circRNAs such as *ciRS-7*, which is preferentially expressed in human and mouse brains (Hansen et al., 2011, 2013; Memczak et al., 2013), could act as miRNA sponges. A set of intron-containing circRNAs was shown to regulate RNA polymerase II (Pol II) transcription (Li et al., 2015). In addition, circRNAs may play important physiological roles in different biological processes. Hundreds of circRNAs are regulated during human epithelial-mesenchymal transition (EMT), indicating that certain circRNAs may affect EMT-related cellular functions (Conn et al., 2015). Thousands of circRNAs are expressed at high levels in the brain (Rybak-Wolf et al., 2015; You et al., 2015). Many such circRNAs are upregulated during neurogenesis (Rybak-Wolf et al., 2015) and are more enriched in synaptogenesis than their linear counterparts (You et al., 2015). It is worth noting, however, that how the dynamic expression of circRNAs upon neuronal differentiation is achieved remains unknown.

Understanding how circRNAs themselves are regulated is important to delineate their associated function. CircRNAs are derived from Pol II transcripts, together with their linear mRNA counterparts. The processing of back-splicing requires the canonical spliceosomal machinery (Ashwal-Fluss et al., 2014; Starke et al., 2015; Wang and Wang, 2015) and is modulated by both *cis*- and *trans*-regulators (Ashwal-Fluss et al., 2014; Zhang et al., 2014; Conn et al., 2015; Ivanov et al., 2015; Kramer et al., 2015; Starke et al., 2015). For example, treatment of HeLa cells with a splice inhibitor followed by nascent RNA purification revealed that circRNA biogenesis requires spliceosome assembly (Starke et al., 2015). Most circRNAs in mammals (Jeck et al., 2013; Zhang et al., 2014) and worms (Ivanov et al., 2015) are processed from internal exons with long flanking introns that usually contain reverse complementary sequences. Such sequences are capable of forming paired duplex structures that significantly enhance exon circularization (Liang and Wilusz, 2014; Zhang et al., 2014). RNA binding proteins, such as Muscleblind and Quaking, have also been shown to regulate circRNA formation by bridging splice sites close together to facilitate back-splicing (Ashwal-Fluss et al., 2014; Conn et al., 2015).

Together, these known features of circRNA biogenesis suggest that circRNA processing is, in principle, linked to transcription and pre-mRNA splicing. However, direct evidence of how and to what extent back-splicing is coupled to transcription and splicing is lacking. Furthermore, it is known that the final mRNA levels are balanced between their production and degradation (Rabani et al., 2011). This may be the case for circRNAs as well. However, the currently available circRNA profiling has only examined steady-state levels of circRNAs. In this regard, whether the pervasive detection of the steady-state circRNAs in a cell-/tissue-specific manner (Conn et al., 2015; Salzman



RefGene+UCSC Known genes (29,250)

	4sU tagging time	Number of transcribed genes	Number of completely transcribed genes	% of transcribed genes
Nascent RNAs within 120 min of transcription	10 min	8,757	3,657	41.76%
	15 min	10,371	5,582	53.82%
	30 min	12,976	10,047	77.43%
	60 min	15,208	13,829	90.93%
	120 min	16,228	15,730	96.93%
	4 hr	16,497	16,393	99.37%
	16 hr	16,637	16,637	100%

Figure 1. Detection of Nascent circRNAs by 4sUDRB-Seq

(A) A workflow of 4sUDRB-seq for nascent RNAs. PA1 cells were treated with DRB for 3 hr to block transcription, and 4sU-labeled newly synthesized total RNAs were collected at different time points after DRB removal. 4sU-labeled nascent RNAs were incubated with HPDP-biotin and then purified by streptavidin beads, followed by rRNA depletion and/or additional RNase R treatment to enrich for nascent circRNAs. The resulting purified nascent RNAs were subjected to RNA-seq.

(B) Detection of nascent circRNAs. Nascent circRNAs and their linear mRNA transcripts from the *PVT1* and *POLR2A* gene loci were captured by qRT-PCR from purified nascent RNAs in PA1 cells. Note that nascent circRNAs, but not nascent linear mRNAs, could be detected after RNase R treatment.

(C) A Dox-inducible circRNA mini-gene stable expression system can efficiently produce a circRNA (*circmCherry*) from a split *EGFP* gene. The Dox-inducible circRNA mini-gene contains a pair of complementary sequences (thick red arrows) that have the potential to form RNA pairing across introns (Zhang et al., 2014) flanking the *mCherry* exon.

(D) The production of nascent *circmCherry* is efficient within 1 hr of Dox induction. The primers for qRT-PCR used to detect the different RNAs are labeled in (C).

(E) 4sU incorporation has no measurable effect on transcription of the mini-gene, splicing of *egfp* or back-splicing of *circmCherry* in HeLa cells. 4sU-labeled nascent RNAs were purified by streptavidin pull-down and amplified by qRT-PCR from cells with the mini-gene in the following conditions. NT, non-treated; Dox, add 1 μg/ml Dox for 1 hr; DRB+Dox, add DRB for 3 hr, washout, and then add Dox for 1 hr; Dox+4sU, add Dox and 4sU for 1 hr; Dox+DRB+4sU, add DRB for 3 hr, washout, and then add Dox and 4sU for 1 hr.

(legend continued on next page)

et al., 2013; Starke et al., 2015) truly reflects the endogenous kinetics of circRNA production is unknown.

In the current study, we applied metabolic tagging of newly transcribed RNAs by 4-thiouridine (4sU) (Fuchs et al., 2015) to enrich nascent circRNAs and developed computational algorithms to calculate circRNA processing kinetics globally in human embryonic carcinoma PA1 cells. This comprehensive dataset allowed us to quantitatively measure and compare parental gene transcription elongation, pre-mRNA splicing, and circRNA back-splicing at individual gene loci across a time course lasting 16 hr. We gained a number of previously unknown insights into circRNA biogenesis by analyzing the kinetics of nascent circRNA processing. Moreover, investigation of nascent circRNA processing in undifferentiated human embryonic stem cells (hESCs) and their differentiated forebrain (FB) neuron progenitor cells further revealed how the abundant and dynamic expression of circRNAs is achieved upon neuronal differentiation.

RESULTS

Capturing Newly Synthesized circRNAs by 4sUDRB Sequencing

To study the kinetics of circRNA processing, we have optimized and applied the metabolic tagging of newly transcribed RNAs (including circRNAs) by 4sU (Fuchs et al., 2015; Rädle et al., 2013). 4sUDRB sequencing (4sUDRB-seq) is based on the reversible inhibition of transcription with 5,6-dichloro-1-β-d-ribofuranosylbenzimidazole (DRB) and a pulse labeling with the uridine analog 4sU after DRB removal (Fuchs et al., 2015) (Figure 1A). In our hands, treatment of PA1 cells with DRB for 3 hr efficiently arrested Pol II transcription. After that, DRB removal led to continuous transcription as shown by the observation that the resumption of transcription of examined genes could be captured as soon as 2.5 min after 4sU exposure (Figures S1A–S1C). We also found that 4sU labeling allowed the purification of 4sU-labeled newly transcribed RNAs (Figure S1D). Importantly, newly transcribed RNA signals from the proximal region with respect to the transcription start site could be captured with a short 4sU exposure time, whereas signals from the distal region could be detected only with prolonged 4sU exposure (Figure S1E).

The detected 4sU-labeled nascent circRNAs, for example, POLR2A and *PVT1*, are extremely low (Figure 1B). To exclude the possibility that the 4sU labeling might affect nascent circRNA processing or that the following N-[6-(Biotinamido)hexyl]-3′-(2′-pyridyldithio)propionamide (HPDP-biotin) purification might not be sufficient to capture circRNAs, we designed a Dox-inducible stable circRNA expression system that efficiently produces a circRNA from the *mCherry* exon (*circmCherry*) from a split *EGFP* gene (Figure 1C). The expression of nascent *circmCherry* was at least 100-fold higher than that of the nascent endogenous *circPOLR2A* and *circPVT1* (Figures 1B and 1D). Ten percent of the skipped *mCherry* exon could be back-spliced to *circmCherry* within 1 hr of Dox induction (Figure 1D). The 4sU incorporation had no measurable impact on Pol II transcription, splicing of *egfp*, or back-splicing of *circmCherry* (Figure 1E), suggesting that 4sU labeling does not specifically affect back-splicing on nascent circRNA. In addition, the following HPDP-biotin purification was as efficient as methanethiosulfonate (MTS)-biotin purification in circRNA recovery (Figure S1F). MTS-biotin purification is a recently reported method shown to have high efficiency for 4sU-labeled RNA purification, in particular, for small RNAs, which tend to have fewer uridine residues (Duffy et al., 2015). It should be noted that the spliced linear *egfp* and the back-spliced *circmCherry* are similar in size (~800 nt) and that the average length of nascent circRNAs in PA1 cells is 482 nt (data not shown). Therefore, the 4sUDRB-seq used in this study has at best a limited bias for the collection of nascent circRNAs and their linear RNA counterparts.

4sUDRB-seq has been used to measure transcription elongation rates (TERs) (Fuchs et al., 2014). In this regard, short duration 4sU labeling, usually no more than 15 min, was applied to calculate elongation speeds. In our study, datasets from 4sU labeling for 10 and 15 min (Figure S2A) were used to measure TERs with a newly developed computational pipeline (referred as to TERate; Figure S3A). Non-4sU-labeled pre-mRNAs and circRNAs, which were pre-existing in cells and were non-specifically co-purified with nascent RNAs, were removed as non-specific noise (Figure S3B). Importantly, the TERs calculated by our pipeline were comparable to those calculated by the published method (Fuchs et al., 2014) even though many more genes could be calculated by our method (Figure S3C).

We found that the average length of nascent-circRNA-producing genes was significantly greater than that of non-circRNA-producing genes (Figure 1F). Thus, it required a longer time for such genes to complete their transcription (Figure 1G). We therefore prolonged the 4sU incubation of PA1 cells to 30, 60, and 120 min and even to 4 and 16 hr (Figures 1H and S2A) to identify as many circRNAs as possible and to characterize the kinetics of circRNA processing and decay during transcription. Comparison of read distribution revealed that the great majority of the steady-state RNA sequencing (RNA-seq) mapped reads was located in exons, whereas reads in 4sUDRB-seq within

(F) Length distribution of newly transcribed circRNA-producing genes (red line) and non-circRNA-producing genes (black line) in 4sU-labeled PA1 cells after DRB removal. Gene lengths were extracted from gene annotations. One thousand randomly selected non-circRNA-producing genes were used as a control. ***$p = 5.9 \times 10^{-268}$, Wilcoxon rank-sum test.

(G) Estimated Pol II elongation time required for circRNA-producing (red box) and non-circRNA-producing (black box) genes to complete transcription in PA1 cells. ***$p = 1.3 \times 10^{-256}$, Wilcoxon rank-sum test.

(H) Analysis of transcribed genes by 4sUDRB-seq in PA1 cells. A transcribed gene was suggested by at least 100 4sUDRB-seq reads mapped to its transcription start site (TSS) within the proximal 10-kb region (top). After calculating the TER, the transcribed distance of a given gene was obtained by multiplying TER with the 4sU exposure time at each time point. Complete transcription of a given gene was suggested by the fact that the transcribed distance was no less than the gene length. The statistics of transcribed and completely transcribed genes were summarized from RefGenes/UCSC known genes. Nascent RNAs were defined as 4sU-labeled RNAs detectable within 120 min after DRB removal. Nascent-circRNA-producing genes were defined as genes that could produce nascent circRNAs within 120 min after DRB removal.

120 min of transcription were largely located in introns (Figure S2B).

Together, we have generated rRNA-depleted 4sUDRB-seq datasets from PA1 cells, hESC H9 cells, and H9 differentiated FB cells over a wide time course. These nascent RNA datasets allowed us to capture newly transcribed circRNAs from long genes and to study the coupling of circRNA processing with transcription and splicing.

Back-Splicing Is Far Less Favorable than Canonical Splicing

Steady-state circRNAs have been detected from thousands of gene loci in metazoans. Because they are expressed at low levels from both endogenous genes and expression vectors (Guo et al., 2014; Liang and Wilusz, 2014; Zhang et al., 2014; Starke et al., 2015), the efficiency of back-splicing circularization was predicted to be lower in general than that of canonical splicing. One hypothesis is that spliceosomes are unfavorably assembled at back-splicing sites. However, direct evidence has been lacking and to what degree back-splicing circularization is inefficiently catalyzed by the spliceosome is unclear.

To answer this question, we carried out detailed analyses on the 4sUDRB-seq dataset in PA1 cells (Figure 2). Newly produced circRNAs could be identified by the CIRCexplorer pipeline (Zhang et al., 2014) in different periods of 4sU incubation after DRB removal. Examples of genes that produce nascent circRNAs are shown in Figures 2A and S4. Such nascent circRNAs were resistant to RNase R digestion (Figures 2A and S4), confirming that the nascent circRNAs identified here were in circles. Because circRNA-producing genes are usually long genes whose circle-forming exons are not transcribed within short 4sU labeling time periods (Figures 1F and 1G), only a dozen nascent circRNAs were detected at 10 and 15 min of 4sU labeling time after DRB removal (Figure 2B; Table S1). With increased time points of 4sU labeling, hundreds of nascent circRNAs were identified within 120 min after transcription initiation (Figure 2B). Compared to canonical splicing events at the junction sites of circle-forming exons (Figure 2B), we found that the efficiency of back-splicing was extremely low on a genome-wide scale (Figures 2B and 2C). For example, although 111,865 upstream splicing junction (SJ) reads could be identified at 120 min of 4sU labeling after Pol II transcription continued, a time point at which almost all active genes have completed their transcription in PA1 cells (Figure 1H), only 913 back-splicing junction (BSJ) reads from 641 nascent circRNAs originating from 534 host genes were detected in PA1 cells (Figure 2B). A large number of newly synthesized circRNAs were found at 4 and 16 hr of 4sU labeling; however, the efficiency of back-splicing was still generally low compared with the adjacent canonical splicing events at these time points (Figures 2B and 2C). This was also the case when using other pipelines (MapSplice [Jeck et al., 2013] and find_circ [Memczak et al., 2013]) for circRNA prediction (data not shown).

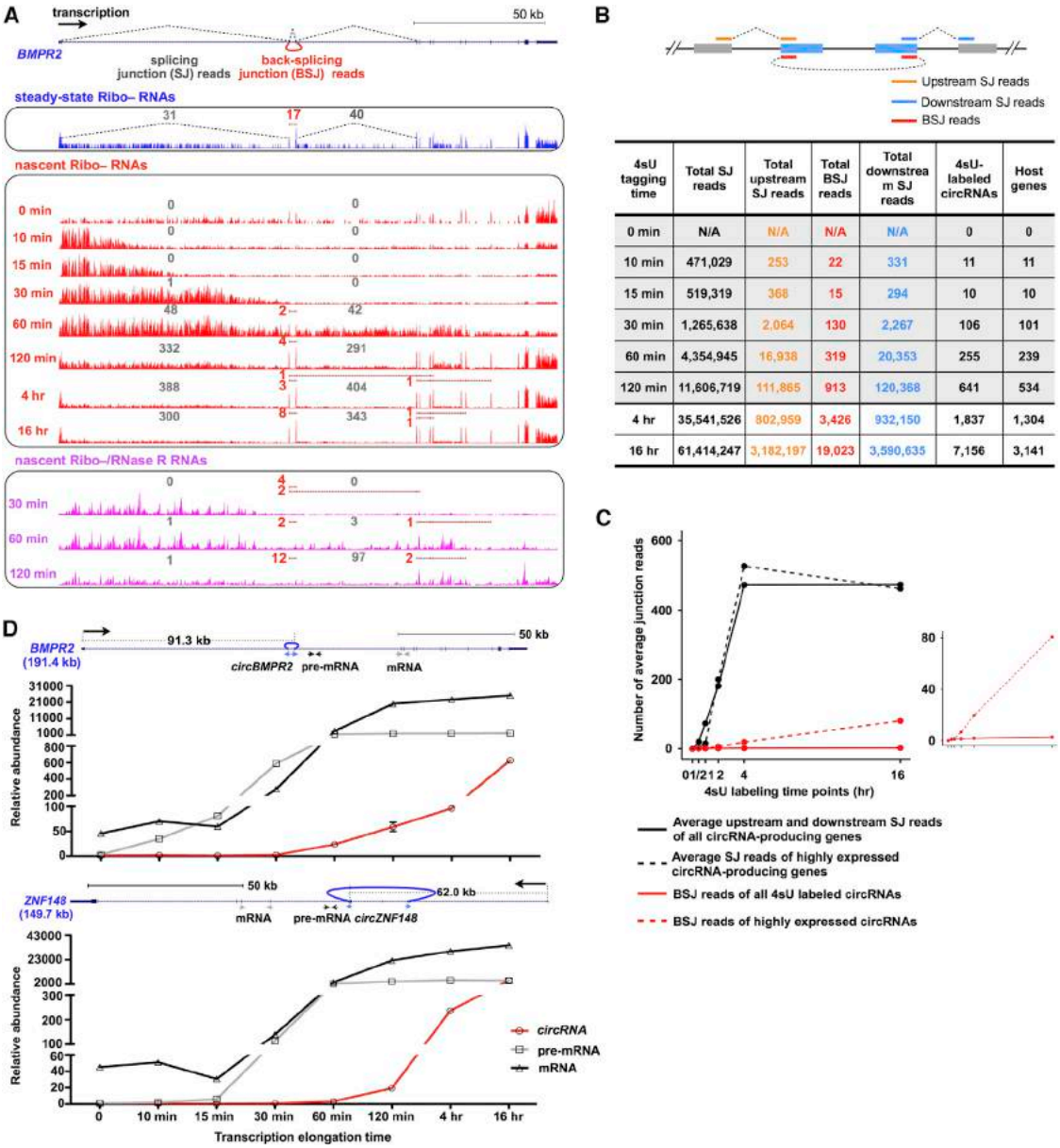
Additionally, we examined nascent circRNA processing by quantitatively analyzing newly synthesized pre-mRNAs, circRNAs, and spliced mRNAs from two circRNA-producing genes, *BMPR2* and *ZNF148*. Both genes are long and their corresponding circRNAs are generated from middle exons (Figures 2A and S4). The relative abundance of pre-mRNA was measured using a primer set located in an intron adjacent to the circle-forming exons, and the spliced mRNA was measured using a primer set recognizing adjacent exons (Figure 2D). We found that the production of pre-mRNAs and spliced mRNAs gradually increased during 60 min of transcription elongation and that the relative abundance of pre-mRNAs stopped increasing after 60 min (Figure 2D). However, nascent circRNAs were only beginning to be detected at this time point, and the detectable nascent circRNAs were extremely rare compared to their linear mRNA counterparts, although more 4sU-labeled *circBMPR2* and *circZNF148* were detected at prolonged labeling time points (Figure 2D). Together, results from 4sUDRB-seq and 4sUDRB qRT-PCR regarding nascent RNAs reveal that back-splicing circularization is far less favorable than canonical splicing in cells (Figure 2).

Back-Splicing Circularization Is Associated with Fast Pol II Elongation Rate

It is known that compatible rates of splicing and transcription permit mRNA processing to occur simultaneously and that the rate of transcription elongation can change the outcome of splicing events (Bentley, 2014). As circRNA biogenesis depends on the spliceosomal machinery (Ashwal-Fluss et al., 2014; Starke et al., 2015; Wang and Wang, 2015), it is possible that circRNA processing could be affected by the Pol II TER. Both the newly developed TERate pipeline (Figures S3A and S3B) and a published method (Fuchs et al., 2014) (Figure S3C) revealed that the average TER of nascent circRNA-producing genes was higher than that of non-circRNA genes (Figures 3A and S3D), for example, 2.90 kb/min versus 2.29 kb/min calculated by the TERate pipeline (Figure 3A). This analysis suggests that circRNA formation correlates with fast Pol II elongation.

To further confirm this observation, we constructed three cell lines that expressed either a wild-type (wt) version of the human Pol II large subunit implicated in elongation control or one of two mutants (Fong et al., 2014). The Pol II mutants carried either R749H or E1126G, which individually decelerates or accelerates transcription, and each mutation was expressed from a vector for α-amanitin-resistance (Am^r) (Figures 3B and S5A). In brief, 293FT cells were transfected with either WT Pol II or a mutant version for 24 hr, followed by the addition of α-amanitin to block endogenous Pol II elongation during the whole experiment. 4sU pulse labeling and nascent RNA collection were then performed at 30 and 120 min after DRB removal to measure the TER (distal versus proximal, D/P) and circularization index (CI, the relative abundance of circRNA [C] versus spliced linear mRNA [L]) of circRNA-producing genes (Figures 3B and 3C).

In agreement with previous reports (de la Mata et al., 2003; Fong et al., 2014), the R749H mutant reduced Pol II TER (Figure 3D, top) and the E1126G mutant promoted Pol II TER (Figure 3E, top). Importantly, the circularization efficiency of the examined circRNAs correlated positively with the altered Pol II elongation rate. Whereas lower levels of nascent circRNAs were detected with reduced TER (Figure 3D, bottom), higher levels of nascent circRNAs were identified with increased TER (Figure 3E, bottom). In addition, it has been shown that the elongation complex tends to travel faster in genes with long first introns (Jonkers et al., 2014). We found that most newly



4sU tagging time	Total SJ reads	Total upstream SJ reads	Total BSJ reads	Total downstream SJ reads	4sU-labeled circRNAs	Host genes
0 min	N/A	N/A	N/A	N/A	0	0
10 min	471,029	253	22	331	11	11
15 min	519,319	368	15	294	10	10
30 min	1,265,638	2,064	130	2,267	106	101
60 min	4,354,945	16,938	319	20,353	255	239
120 min	11,606,719	111,865	913	120,368	641	534
4 hr	35,541,526	802,959	3,426	932,150	1,837	1,304
16 hr	61,414,247	3,182,197	19,023	3,590,635	7,156	3,141

Figure 2. Back-Splicing by the Spliceosome Is Unfavorable

(A) An example of 4sUDRB-seq for nascent circRNA produced from *BMPR2* in PA1 cells. Top: a wiggle-track shows steady-state *bmpr2* mRNA and circRNA (*circBMPR2*) identified from RNA-seq of total RNAs after rRNA depletion (blue). Middle: nascent linear and circular RNAs from *BMPR2* revealed by 4sUDRB-seq with different 4sU labeling time points after DRB removal (red). Bottom: nascent *circBMPR2* was enriched by RNase R treatment before 4sUDRB-seq (pink). Alternative circularization events were also found after RNase R treatment. Splicing junction (SJ) reads (gray), which are circularized exons spliced to their upstream or downstream exons, and back-splicing junction (BSJ) reads (red) for *circBMPR2* were extracted from 4sUDRB-seq, respectively.

(B) The identified BSJ reads are rare compared to splicing events. Top: a schematic drawing of reads mapped to spliced exons, skipped exons, and back-spliced exons. For each indicated transcription elongation time point, the total numbers of detected SJ reads, BSJ reads, reads upstream and downstream of SJ of the exact same circularized exons, 4sU-labeled circRNAs, and originating host genes are listed. Of note, as RNAs transcribed within 120 min of 4sU incubation time generates substantially more unspliced introns (Figure S2B), the depth of each 4sU-labeled sample was sequenced according to the concentration of each purified sample and spiked-in RNA within 120 min of 4sU labeling.

(C) Comparison of the average SJ (black) and BSJ (red) reads from all nascent circRNA-producing genes (lines) and highly expressed nascent circRNAs (dotted lines) across the 4sU labeling time course. The highly expressed nascent circRNAs are type I circRNAs (see Figure 4E for details). y axis, average numbers of SJ and BSJ reads. x axis, 4sU labeling time points.

(D) The back-splicing efficiency of nascent circRNAs is low during transcription as revealed by qRT-PCR. Nascent RNAs purified from PA1 cells at different 4sU labeling time points were subjected to qRT-PCR with primer sets (arrows) that individually recognize circRNAs, pre-mRNAs, and spliced mRNAs at *BMPR2* (top) and *ZNF148* (bottom) gene loci. The relative abundance of circRNAs, pre-mRNAs, and spliced mRNAs at each time point was plotted.

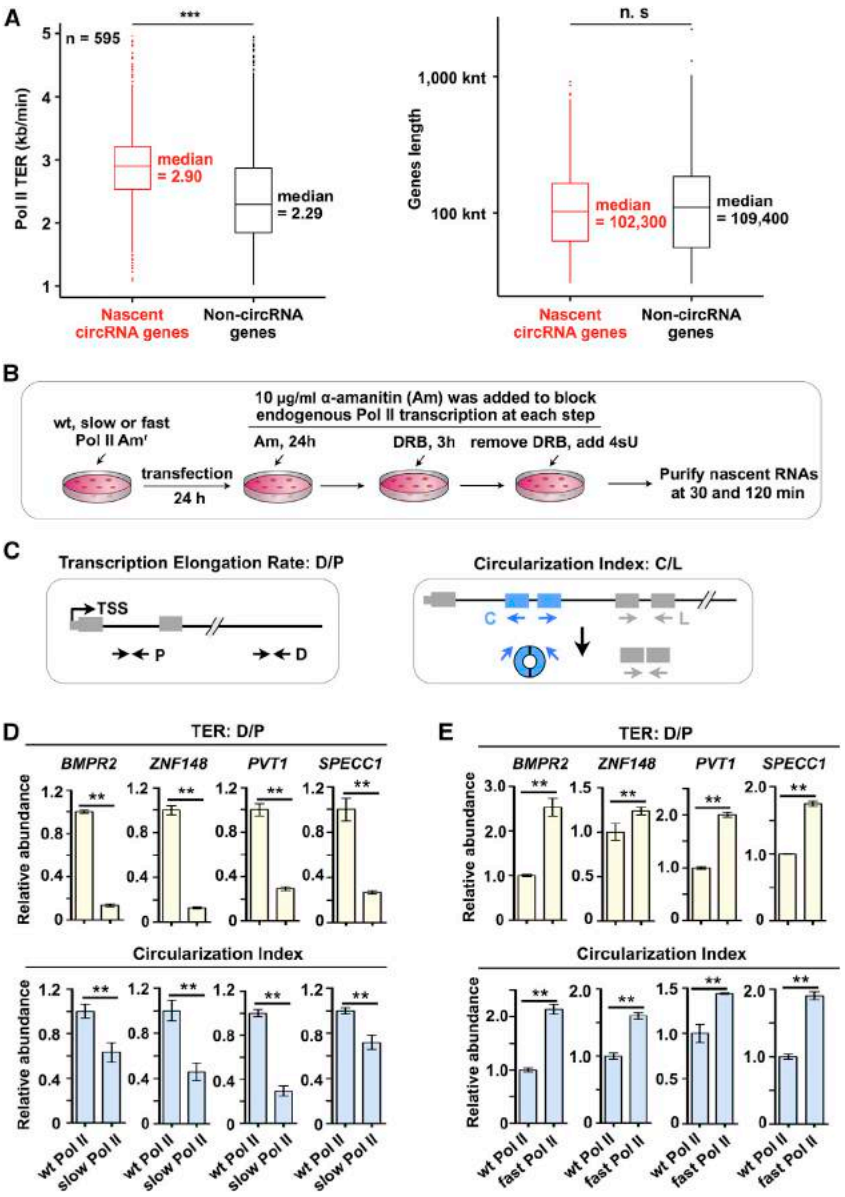


Figure 3. Back-Splicing Is Correlated with Fast Pol II Elongation Rate

(A) The Pol II TER of circRNA-producing genes is higher than that of non-circRNA- producing genes in PA1 cells. Left: TER was calculated as the average value of 10 and 15 min of Pol II transcription after DRB removal. ***$p = 7.2 \times 10^{-39}$, Wilcoxon rank-sum test. Right: the length of nascent circRNA genes and of control non-circRNA genes used in the left panel. Note that gene length is similar between these two groups of genes. n.s., not significant, $p = 0.4$, Wilcoxon rank-sum test.

(B) A schematic drawing showing the experimental procedure carried out in (D) and (E).

(C) Left: a schematic drawing showing how Pol II TER is measured. "D" and "P" primer sets are used to quantify pre-mRNA accumulation at distal and proximal regions with respect to the TSS. Right: a schematic drawing showing how the circularization index is calculated. "C" and "L" primer sets are used to quantify circRNAs and linear RNAs.

(D) A reduced TER (slow Pol II mutant R749H) leads to lower level of nascent circRNA production, as revealed by qRT-PCR at gene loci examined in 293FT cells.

(E) An enhanced TER (fast Pol II mutant E1126G) leads to higher level of nascent circRNA production, as revealed by qRT-PCR at gene loci examined in 293FT cells.

In (D) and (E), error bars represent ±SD of biological repeats. **$p < 0.01$.

synthesized (and steady-state) circRNAs are derived from genes with a significantly long first intron (Figure S5B). These observations further indicate that circRNA formation associates with fast Pol II TER.

Back-Splicing Largely Occurs Post-transcriptionally

Nascent circRNA formation correlates with fast Pol II TER (Figure 3) and a few circRNAs could be produced co-transcriptionally within short 4sU labeling periods (10 and 15 min) after DRB removal (Figure 2). However, we noticed that only a limited number of nascent circRNAs could be detected within 120 min of transcription initiation in the available 4sUDRB-seq datasets (Figures 2B and 2C), whereas thousands of circRNAs were identified with longer 4sU labeling time and at the steady state in PA1 cells (Figure 2B; Table S1). This observation suggested that the majority of circRNAs could be produced after the transcription of their parent genes has completed.

Analysis of the 60- and 120-min 4sUDRB datasets from PA1 cells revealed that most (90%) of the nascent circRNA-producing genes had been transcribed to their termination sites after 60 min (Figures 4A and 4B). Only a small portion of circRNA-producing genes generated nascent circRNAs at 60 min of 4sU labeling. In contrast, a greater number of such genes could generate nascent circRNAs at extended time points (4 and 16 hr) of 4sU labeling (Figure 4C). Indeed, comparison of the number of back-splicing and splicing junction sites in the same exact circRNA-forming exons further revealed that back-splicing was continually increased within the examined 16-hr 4sU labeling period, whereas the metabolism of canonical splicing had mostly achieved an equilibrium at the 4-hr time point (Figure 4D). This was also the case when average SJ and BSJ reads were analyzed for highly expressed nascent circRNAs (Figure 2C). Together, these results show that back-splicing largely occurs post-transcriptionally.

The Steady-State Level of circRNAs Correlates Positively with Levels of Nascent circRNAs

Depending on the time when a nascent circRNA was detected during the 4sU pulse labeling, we classified newly synthesized

circRNAs into two major types for further analysis (Figure 4E). Type I circRNAs could be detected as early as with 120 min of 4sU labeling and also found with 4 and 16 hr of 4sU labeling, whereas type II circRNAs were detected only with 4 and 16 hr of 4sU labeling (Figure 4E). Accordingly, the steady-state expression level of group I circRNAs was remarkably higher than that of group II circRNAs (Figure 4F). Such nascent circRNAs first appeared within 120 min (Figure 4G) and kept accumulating in the later time points. Moreover, the steady-state levels of circRNAs also correlated positively with their nascent levels in examined genes (Figure 4G). These results strongly suggest that the widespread detection of some abundant steady-state circRNAs is largely due to their post-transcriptional accumulation after synthesis.

Here, to compare the expression of circRNAs among different samples, we applied RPM (mapped back-splicing junction reads per million mapped reads) (Zhang et al., 2014) to quantify the relative expression of circRNAs in each sample by normalizing across different sequencing depths (see the Supplemental Experimental Procedures). RPM $\geq$0.1 was used as a cutoff to identify circRNAs with high confidence as previously reported (Zhang et al., 2014).

We measured the CI values of several circRNA-producing genes at nascent and steady-state levels to validate what we observed from the 4sUDRB-seq in PA1 cells (Figures 4E–4G). For all examined circRNAs, their circularization gradually accumulated within 16 hr, and their CI values at the steady state were much higher than those at nascent levels (Figure 4H). Noticeably, all examined circRNAs were stable, whereas their corresponding linear transcripts had an average half-life of 8 hr (Figure 4I). These results indicate that the accumulation of newly synthesized circRNAs significantly contributes to their widespread detection at steady-state levels. By quantitative comparison of six different circRNAs in four commonly used human cell lines, we further confirmed that the steady-state levels of circRNAs correlated positively with their nascent levels (Figure 4J).

Nascent circRNA Processing Is under Tight *cis*-Regulation

It has been shown that complementary sequences (mostly inverted repeated *Alu* elements, IR*Alu*s, in human) across long flanking introns facilitate the back-splicing of exons (Liang and Wilusz, 2014; Zhang et al., 2014; Kramer et al., 2015). If this were the case, we would expect to observe *cis*-complementary sequences embedded within flanking introns of nascent circRNA-forming exons. Indeed, 70% of nascent exon circularization events in PA1 cells were associated with potential RNA pairing across introns (Figure 5A). Importantly, when the intronic RNA pairing across the circle-forming exons was disrupted by CRISPR-Cas9 mediated genome editing at the examined circRNA-producing locus, *GCN1L1* (Figure 5B), no *circGCN1L1* could be detected at this locus at all (Figure 5C). This endogenous analysis demonstrates that the RNA pairing is required for circRNA formation. Notably, the removal of one intronic complementary sequence (ICS) bracketing the circularized exons had no measurable effect on linear mRNA processing (Figure 5C). Thus, genome editing of the intronic RNA pairing that is required for circRNA biogenesis provides a neat way to specifically knock out the expression of a circRNA expression without affecting its residential mRNA.

ICS deletion also had no effect on the Pol II TER at this locus (Figure 5D). Importantly, TER augmentation in PA1 cells by introducing the fast Pol II mutant (E1126G) increased *circGCN1L1* processing, but it could not induce any circRNA formation in ICS knockout cells (Figure 5D). Thus, although circRNA formation is correlated with fast Pol II elongation speed, it also requires RNA pairing in flanking introns. Together, nascent circRNA biogenesis is under tight control by *cis*-complementary sequences in flanking introns, although its efficiency is low.

Neuron-Expressed Genes Do Not Significantly Contribute to Neuronal circRNA Expression

Thousands of steady-state circRNAs are expressed at high levels in mammalian brain (Rybak-Wolf et al., 2015; You et al., 2015). Many such circRNAs appear to be upregulated during neurogenesis independently of their linear isoforms (Rybak-Wolf et al., 2015; You et al., 2015). However, the underlying mechanism that causes this significant circRNA upregulation has remained unclear.

We examined circRNA expression during human neurogenesis by differentiating hESC H9 cells (H9 D0) into FB neuron progenitor cells (Figures 6A and S6A; data not shown) (Chen et al., 2015). At the D26 differentiation time window (FB D26), all FB marker genes were expressed, and such FB neurons could only proliferate slowly (data not shown). We collected total RNAs from these two cell types followed by rRNA depletion and sequenced the steady-state levels of both linear and circular RNAs. In agreement with previous findings (Rybak-Wolf et al., 2015; You et al., 2015), we found that the steady-state levels of circRNAs (Figure 6B), including highly expressed ones with RPM $\geq$0.1, were significantly upregulated upon hESC differentiation into FB neurons. For example, whereas only 162 high-confidence circRNAs were detected in H9 cells, 785 circRNAs with RPM $\geq$0.1 were identified in FB neurons (Figure 6B; Tables S3 and S4).

What contributes to this specific circRNA expression upon neurogenesis? One hypothesis is that the specific expression of neuronal lineage genes could generate significantly more circRNAs, resulting in increased circRNA expression in neurons. To test this hypothesis, we first analyzed lineage-specific expressed genes that are at least 5-fold enriched in either H9 hESCs or FB neurons, and we found that approximately 8% of genes were highly expressed in each of the two cell types (Figure 6C). We then analyzed how many circRNAs could be produced from these cell-type-specifically expressed genes. Whereas 65 of 351 H9-specific genes produced circRNAs, 125 of 347 FB-specific genes generated circRNAs with the current datasets (Figure 6D, left). In addition, only 152 (~3.4%) H9-specific circRNAs and 456 (~4.0%) FB-specific circRNAs were produced due to cell-type-specific gene expression (Figure 6D, right). Thus, neuronal lineage-specific expressed genes did not contribute significantly to the high expression of steady-state circRNAs in hESC-differentiated FB neurons.

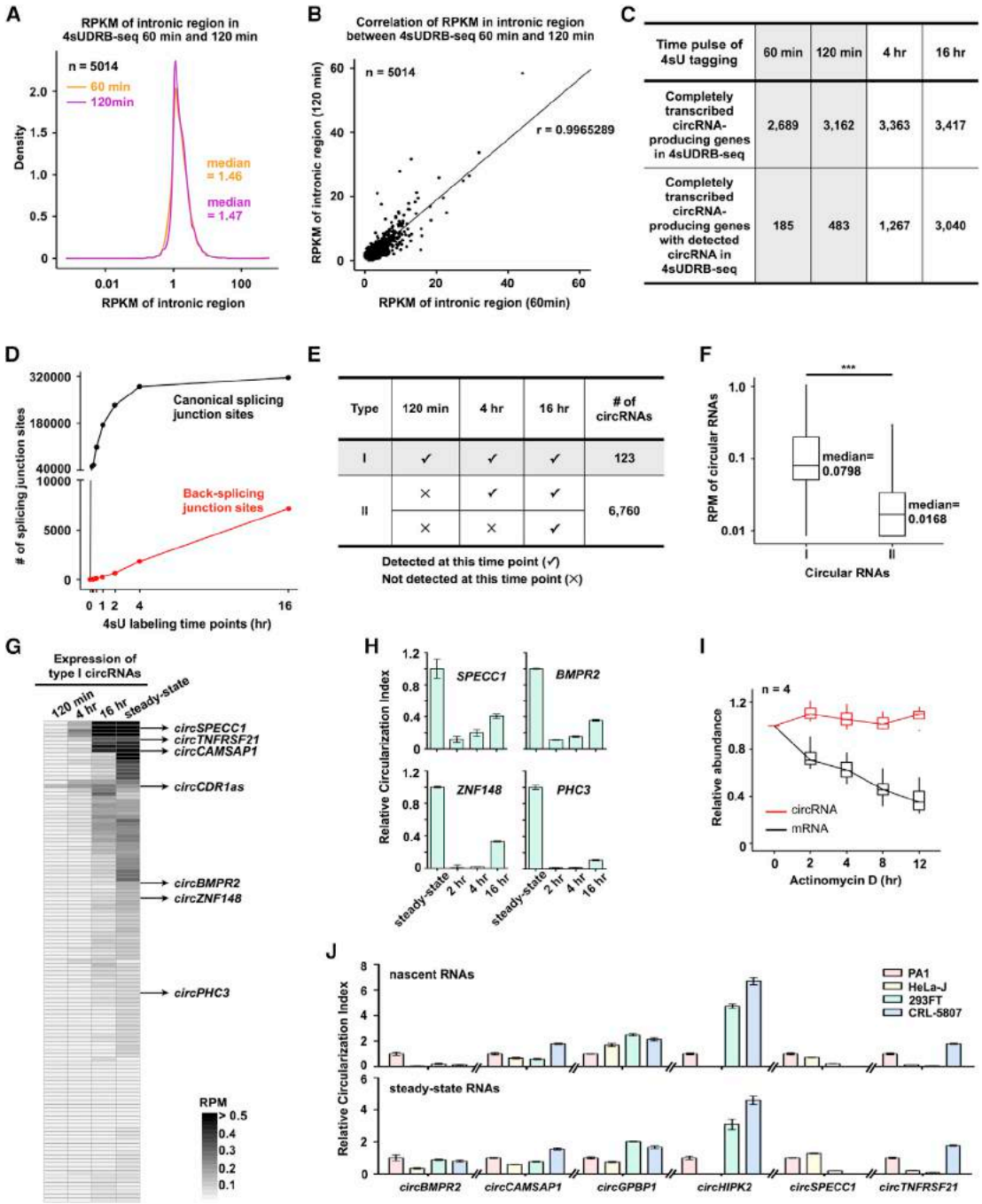


Figure 4. Post-transcriptional Synthesis and Accumulation of circRNA

(A and B) Most nascent pre-mRNAs have completed their transcription within 60 min. Reads per kilobase per million mapped reads (RPKM) distribution (A) and correlation (B) of pre-mRNA expression showed that most pre-mRNAs have similar expression at 60- and 120-min 4sU exposure time points after DRB removal.

(C) CircRNAs are processed post-transcriptionally. Numbers of completely transcribed circRNA-producing genes and such genes that produce nascent circRNAs are indicated at 60- and 120-min and 4- and 16-hr of 4sU labeling time points after DRB removal.

(D) Detection of circRNAs at prolonged 4sU labeling time points. The detected back-splicing sites (red line) and their adjacent canonical splicing sites (black line) were plotted across a 16-hr 4sU labeling period. y axis, numbers of splicing and back-splicing sites. x axis, 4sU labeling time points.

(E) Classification of two types of 4sU-labeled circRNAs in PA1 cells. Type I, circRNAs could be continuously detected at 2, 4, and 16 hr of 4sU labeling. Type II, circRNAs could only be detected at 4- and 16-hr 4sU-labeling time points.

(F) The 4sU-labeled nascent circRNAs early detected (E, type I) accumulated to a higher level at the steady state than those identified only at prolonged 4sU labeling time points (E, type II). ***$p = 2.3 \times 10^{-40}$, Wilcoxon rank-sum test.

(legend continued on next page)

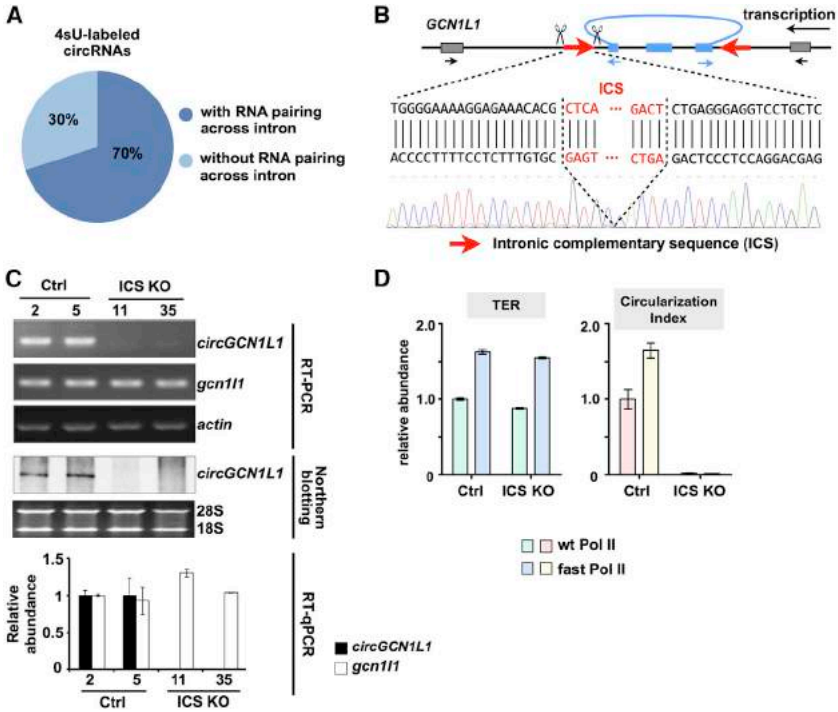


Figure 5. RNA Pairing across Flanking Circle-Forming Exons Is Required for circRNA Biogenesis

(A) Nascent circRNA formation is associated with complementary sequences in flanking introns. All newly synthesized circRNAs detected within 16 hr of 4sU labeling in PA1 cells were analyzed.

(B) A strategy for knocking out circRNA expression at an endogenous locus. The human *GCN1L1* locus is shown. The circle-forming exons in *GCN1L1* are bracketed by one ICS pair with the opposite orientation. One ICS was deleted by CRISPR-Cas9 as indicated. Sanger sequencing of genomic DNA shows the specific ICS deletion in one clone.

(C) Loss of intronic RNA pairing across the circle-forming exons in *GCN1L1* completely blocks *circGCN1L1* formation but has no effect on the linear *gcn1l1* mRNA processing. RT-PCR primers detecting the linear and circular RNAs are indicated in (B) by black and blue arrows. A probe for northern blotting was designed across the BSJ. Two clones of the control (Ctrl) and ICS knockout (KO) were tested.

(D) An enhanced Pol II TER cannot induce *circGCN1L1* formation in ICS KO PA1 cells. Note that introducing the fast Pol II mutant E1126G into the Ctrl PA1 cells led to increased nascent *circGCN1L1* formation. The verification of the Pol II mutant E1126G transfection into Ctrl and ICS KO PA1 cells is not shown.

High Expression of circRNAs in Neurons Is Associated with Enhanced Transcription of circRNA-Producing Genes

To further dissect what causes the differential expression of circRNAs in neurons, we performed 4sUDRB-seq analyses in H9 cells and FB neurons as we had done in PA1 cells. At 120-min 4sU labeling after DRB removal, 1,528 and 2,171 nascent circRNAs were identified in H9 cells and FB neurons (Figure 6E). Although 40% more nascent circRNAs could be detected in FB neurons, back-splicing in general appeared to be much less efficient than canonical splicing in both H9 and H9 differentiated neurons (Figure S6B), similar to what we observed in PA1 cells (Figure 2). For example, only thousands of BSJ reads could be identified in 60- or 120-min 4sU-labeled nascent RNA samples. In contrast, approximately 78,471 to 369,364 upstream or downstream SJ reads could be detected at the same circularized exons (Figure S6B).

Similarly to PA1 cells, it took more than 60 min for greater than 90% of annotated genes to complete transcription in both H9 and FB neurons (Figure S6C). Analysis of the coupling of circRNA processing with transcription revealed that nascent circRNA formation was associated with fast Pol II transcription in both cell types (Figure 6F), consistent with what we observed in PA1 cells (Figure 3). The TERs of circRNA-producing genes expressed in both H9 and FB neurons were higher in FB neurons than in H9 cells (Figure 6G). Importantly, more nascent circRNAs could be produced from these genes in FB neurons than H9 cells (Figure 6G, bottom). We further validated the correlation between transcription speed and nascent circRNA formation from a group of circRNA-producing genes expressed in both H9 cells and FB neurons. These examined genes were transcribed more rapidly in FB neurons than H9 cells (Figure 6H). Correspondingly, the CI values of nascent circRNAs were higher in FB neurons than in H9 cells (Figure 6H). These results together strongly suggest that the processing of nascent circRNA correlates with fast transcription elongation speed and that the high expression of circRNAs in neurons positively associates with the enhanced transcription of circRNA-producing genes.

Accumulation of Steady-State circRNAs in FB Neurons

Next, we asked how the abundant but differential expression of circRNAs is achieved upon neuronal differentiation. To answer this question, we set out to compare the relative abundance of nascent circRNAs with their steady-state levels in H9 cells and

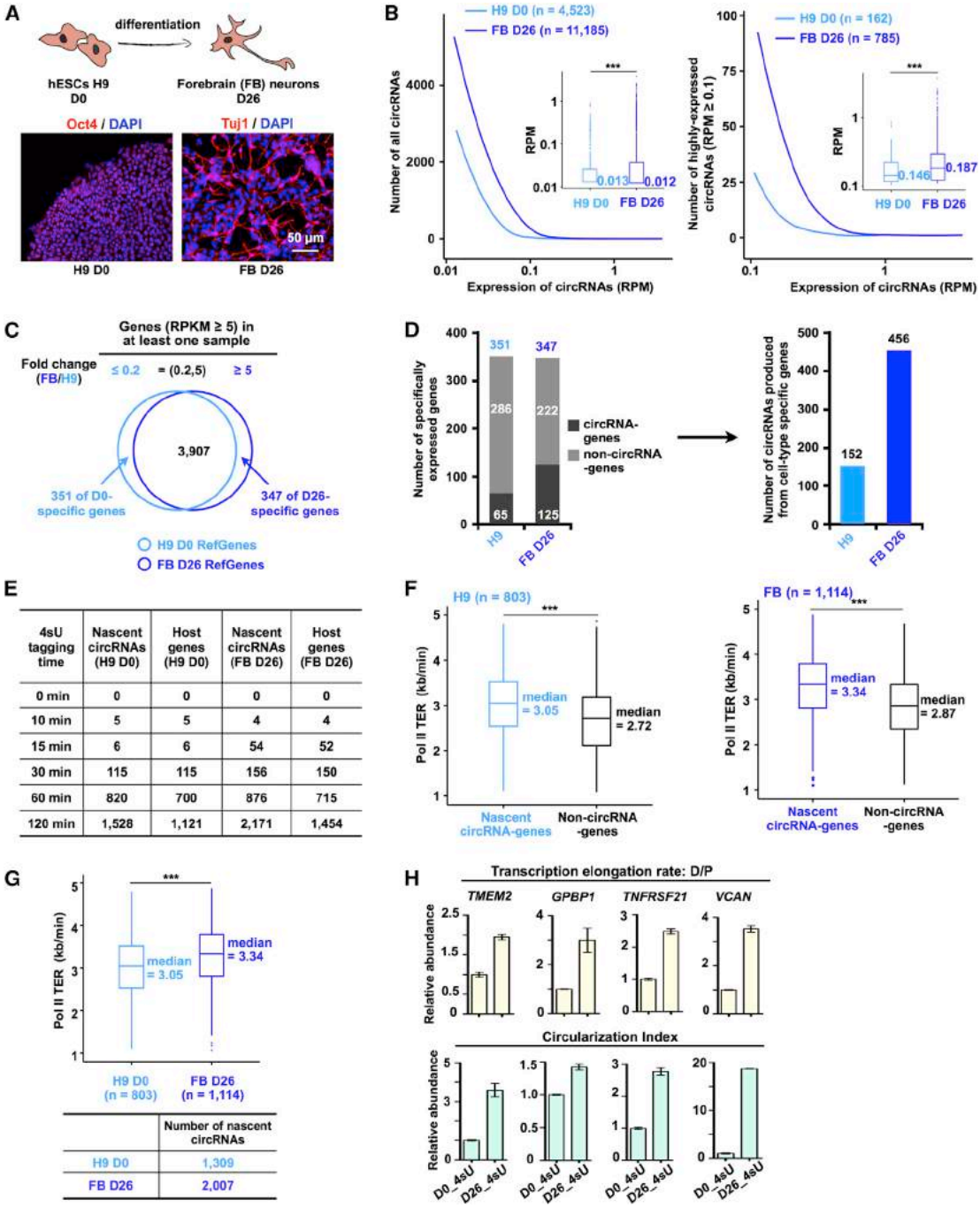
(G) Heatmap illustrating that the steady-state levels of circRNAs correlate positively with their nascent levels for type I circRNAs. Validated type I circRNAs in (H)–(J) are labeled on the right.

(H) The abundance of steady-state circRNAs is largely due to post-transcriptional accumulation. CI values of steady-state and nascent circRNAs at 2, 4, and 16 hr of 4sU labeling were examined by qRT-PCR.

(I) CircRNAs are much more stable than their linear mRNA counterparts. The relative abundance of different circRNAs shown in (H) and their corresponding linear mRNAs were measured by qRT-PCR after actinomycin D treatment in PA1 cells.

(J) The steady-state level of circRNAs correlates positively with their nascent level in different cell lines. Steady-state RNAs and 4sU-labeled RNAs were collected from the indicated cell lines for qRT-PCR. CircRNA expression was normalized to 18S rRNA at the steady-state and nascent levels.

In (I) and (J), error bars represent ±SD of biological repeat experiments.



4sU tagging time	Nascent circRNAs (H9 D0)	Host genes (H9 D0)	Nascent circRNAs (FB D26)	Host genes (FB D26)
0 min	0	0	0	0
10 min	5	5	4	4
15 min	6	6	54	52
30 min	115	115	156	150
60 min	820	700	876	715
120 min	1,528	1,121	2,171	1,454

	Number of nascent circRNAs
H9 D0	1,309
FB D26	2,007

Figure 6. High Expression of circRNAs in Neurons Is Associated with Enhanced Transcription of circRNA-Producing Genes

(A) Differentiation of H9 hESCs into FB neurons, as revealed by immunofluorescence of cell-type-specific markers.

(B) Steady-state circRNAs are upregulated during neuronal differentiation. The number and expression of all circRNAs (left) and high-confidence circRNAs with RPM $\geq$ 0.1 (right) are significantly increased upon neuronal differentiation. ***$p = 2.2 \times 10^{-159}$ (left), ***$p = 2.7 \times 10^{-6}$ (right), Wilcoxon rank-sum test.

(C) Approximately 8% of Refseq genes with high confidence (RPKM $\geq$5) are differentially expressed upon H9 neuronal differentiation.

(D) Neuronal lineage-specific genes do not significantly contribute to neuronal circRNA expression upon differentiation. Left: numbers of circRNA- and non-circRNA-producing genes specifically expressed in H9 or FB cells. Right: numbers of all circRNAs produced from H9 or FB-specific genes.

(E) Numbers of identified nascent circRNAs and their host genes at each indicated 4sU labeling time point after DRB removal in H9 cells and FB neurons are shown.

(F) The Pol II TER of circRNA-producing genes is higher than that of non-circRNA producing genes in H9 cells (left) and FB neurons (right). ***$p = 4.9 \times 10^{-23}$ (left), ***$p = 1.5 \times 10^{-46}$ (right), Wilcoxon rank-sum test.

(legend continued on next page)

FB neurons. We first identified nascent circRNAs with increased expression upon FB neuron differentiation. In total, the expression of 1,772 nascent circRNAs increased ($RPM_{FB} > RPM_{H9}$) and the expression of 1,370 nascent circRNAs decreased ($RPM_{FB} < RPM_{H9}$) upon H9 differentiation to FB neurons (Figure 7A, left) at the 120-min 4sU labeling time point. As expected, the steady-state circRNAs produced from the 1,772 upregulated nascent circRNA-producing genes increased upon differentiation (Figure 7A, right). Strikingly, however, the circRNAs from genes generating lower levels of nascent circRNAs in FB neurons still showed increased expression of steady-state expression levels in these cells, compared to those in undifferentiated H9 cells (Figure 7A, right). Of note, the expression levels of the corresponding linear mRNAs during the nascent and steady-state states were variable during this process (Figure S7).

Indeed, the steady-state expression pattern of linear and circular RNAs was strikingly different between H9 cells and their differentiated FB neurons. Although the total number and steady-state expression level of linear mRNAs remained largely unchanged, the steady-state levels of circRNAs significantly increased in terms of their total number and expression upon neuronal differentiation (Figure 7B). Thus, the synthesis of circRNAs from rapidly transcribed circRNA-producing genes and their accumulation lead to the detection of upregulated steady-state circRNAs in neurons that have slow division rates (Figure 7C).

DISCUSSION

In the current study, we investigated the kinetics of circRNA processing using 4sUDRB-seq with prolonged 4sU incubation times after DRB removal (Figure 1). Our results support the view that back-splicing is extremely inefficient in cells (Figure 2). The low catalytic efficiency could be because the ligation of a downstream 5′ splice site and an upstream 3′ splice site by the spliceosome is sterically unfavorable. It is known that flanking intronic complementary sequences facilitate circularization (Liang and Wilusz, 2014; Zhang et al., 2014; Kramer et al., 2015). Such RNA pairing across circle-forming exons is also highly associated with nascent back-splicing events (Figures 5A and 5B) and critical for circRNA processing (Figures 5C–5E). Thus, nascent circRNA production is under tight *cis*-regulation. Nevertheless, we cannot exclude the possibility that some low expression level circRNAs may be side products of imperfect pre-mRNA splicing (Guo et al., 2014).

It has remained unclear how back-splicing circularization is linked to transcription. A recent study suggested that circularization might occur post-transcriptionally because mutation of the polyadenylation signal of a circRNA-producing linear gene eliminated circRNA production in expression vectors (Liang and Wilusz, 2014). Subsequent work revealed examples where a poly(A) signal was not required for circRNA production from mini-gene vectors (Kramer et al., 2015), indicating that circRNA formation may also occur co-transcriptionally. Indeed, circRNAs could be detected in chromatin-associated RNA from fly heads, and lower amounts of steady-state circRNAs were detected in flies carrying the slow Pol II mutant than wild-type flies (Ashwal-Fluss et al., 2014), implying the co-transcriptional nature of their processing. However, it should be noted that chromatin-bound RNA is not precisely equivalent to nascent RNA as transcripts with mature polyadenylated 3′-ends are found in such chromatin fractions (Bhatt et al., 2012).

Our results suggest that back-splicing of circRNA-forming exons could occur both co- and post-transcriptionally (Figures 2, 3, and 4). On the one hand, similarly to alternative splicing (Braunschweig et al., 2013; Fong et al., 2014), a relatively modest increase or decrease in the elongation rate may have a measurable effect on back-splicing (Figure 3). In addition, it has been proposed that fast elongation favors RNA folding by base-pairing of distal complementary sequences (such as intronic complementary sequences across circle-forming exons), which may result in non-sequential rather than sequential RNA folding during transcription (Bentley, 2014). The positive correlation between fast Pol II elongation speed and nascent circRNA formation with flanking intronic complementary sequences thus indicates that fast elongation may allow non-sequential complementary sequences across introns (rather than within introns) to pair up for back-splicing. On the other hand, although some abundantly expressed nascent circRNA formation events were detected concurrently with Pol II transcription (Figures 2 and 3), significantly more newly synthesized circRNAs were identified after transcriptional completion of their host pre-mRNAs (Figures 2 and 4). This finding is in agreement with the fact that splicing of regulated alternative introns often occurs post-transcriptionally (Braunschweig et al., 2013). As it is well known that the regulation of alternative splicing is complex, and because of recent findings that a couple of proteins could regulate circRNA formation (Ashwal-Fluss et al., 2014; Conn et al., 2015; Ivanov et al., 2015; Kramer et al., 2015), it will be of great interest in the future to identify additional protein factors that are involved in circRNA biogenesis and function.

Although circRNA production is inefficient (Figure 2), their resistance to exonucleolytic degradation allows these RNA circles to accumulate to relatively high levels in cells (Figures 4E–4I). Interestingly, we found that the steady-state levels of circRNAs correlated positively with their nascent levels in all examined cell lines with similar mitotic cycles (Figure 4J). This finding thus suggests that the pervasive detection of steady-state circRNAs in a cell-/tissue-specific manner (Conn et al., 2015; Salzman et al., 2013; Starke et al., 2015) likely reflects the endogenous synthesis of circRNAs: the more nascent circRNAs produced, the higher steady-state levels of circRNAs detected, in particular, when cells have a similar population doubling time. Furthermore, this finding provides additional evidence supporting the notion that circRNA decay is extremely slow.

(G) The Pol II TER of circRNA-producing genes is higher in FB neurons than in H9 cells (top) and more nascent circRNAs are produced from circRNA-producing genes in FB neurons than H9 cells (bottom). ***$p = 2.7 \times 10^{-13}$, Wilcoxon rank-sum test.
(H) Validation of the correlation between Pol II TER and back-splicing efficiency for nascent circRNAs in H9 cells and FB neurons. Error bars represent ±SD of biological repeat experiments.

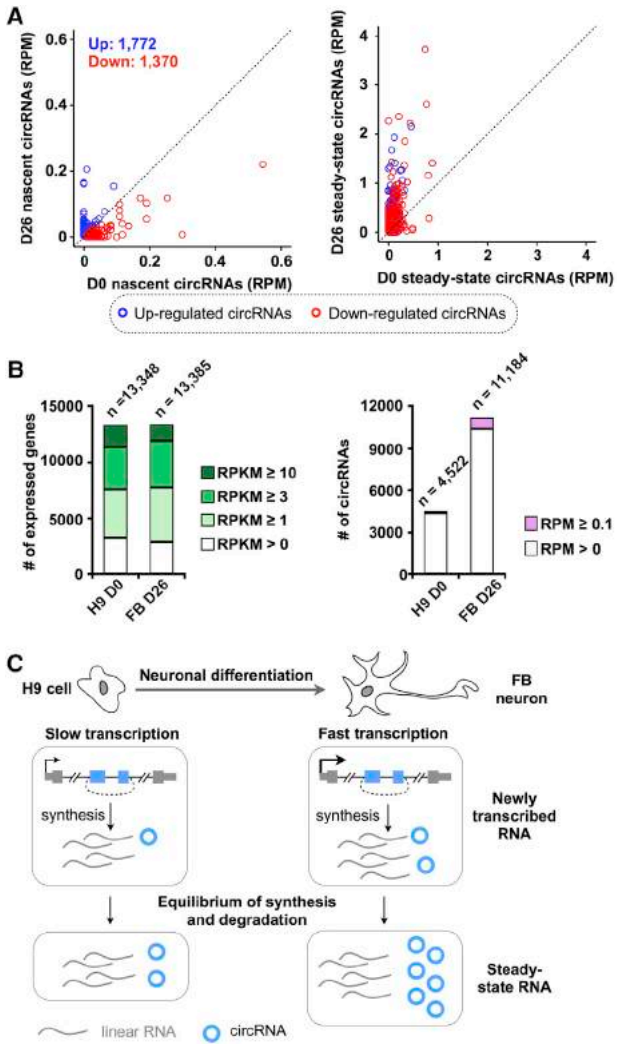


Figure 7. Nascent and Steady-State Levels of circRNAs upon Neuronal Differentiation

(A) The accumulation of steady-state circRNAs in FB neurons is independent of their expression levels at the nascent state. Left: the expression of nascent circRNAs in H9 cells and FB neurons. A total of 1,772 nascent circRNAs increased (blue circles) and 1,370 nascent circRNAs decreased upon H9 cells differentiation into FB neurons (red circles). Right: the expression of steady-state circRNAs in H9 cells and FB neurons. Steady-state circRNAs produced from the nascent circRNA-producing genes shown in the left panel were analyzed. Note that genes generating lower levels of nascent circRNAs in FB neurons still led to increased expression of steady-state circRNAs in these cells compared to those in undifferentiated H9 cells.

(B) Different expression patterns of steady-state linear and circular RNAs between H9 cells and their differentiated FB neurons.

(C) A combined effect results in the abundant but dynamic expression of circRNAs upon neuronal differentiation. On one hand, high expression of circRNAs in neurons is associated with enhanced transcription of their parental genes (top). On the other hand, upon neuronal differentiation, circRNAs accumulate post-transcriptionally to high levels while their linear counterparts do not (bottom).

Many circRNAs are more enriched in mammalian brain than their linear isoforms (Rybak-Wolf et al., 2015; You et al., 2015). We found that the enhanced transcription of circRNA-producing genes upon neuronal differentiation (Figures 6E–6H) and the passive accumulation of circRNAs (Figure 7A) in the slowly dividing neuronal cells, which have a population doubling time of ~5 days (Reubinoff et al., 2001) or longer (data not shown), lead to the observed high expression of steady-state circRNAs in neurons (Figure 7). Interestingly, it appears that the high expression of circRNAs in the brain is independent of their linear transcripts (Rybak-Wolf et al., 2015; You et al., 2015). One possibility is that disparate decay rates between circRNAs and their linear mRNA isoforms may account for this difference. Indeed, circRNAs are very stable, and their corresponding linear mRNAs likely have different decay rates (Figures 7A, 7B, and S7). Thus, the linear-transcript-independent up-expression of circRNAs upon neuronal differentiation could result from a combined effect of augmented transcription of circRNA-producing genes, circRNA accumulation, and diverse decay rates of circRNAs and their linear counterparts upon neuronal differentiation (Figure 7C). However, it is also worth noting that because alternative splicing is prevalent in the brain, back-splicing could also be actively regulated by similar mechanisms.

Taken together, our study on nascent circRNA processing has revealed that circRNA biogenesis occurs largely post-transcriptionally and that nascent circRNA formation is restricted and controlled by *cis*-complementary elements. The functional consequences of these accumulated long-lived circRNAs are waiting to be deciphered.

EXPERIMENTAL PROCEDURES

Additional details can be found in the Supplemental Experimental Procedures.

Metabolic Labeling of Nascent RNAs with 4sU and Nascent RNA Purification

Metabolic labeling of newly transcribed RNAs was performed as described (Fuchs et al., 2014; Rädle et al., 2013) with modifications. PA1, H9, and H9-differentiated FB cells were incubated with 100 μM DRB for 3 hr to block Pol II transcription. Transcription recovered after DRB release and newly transcribed RNAs were labeled with 300 μM 4sU. TRizol was added to stop transcription, and total RNAs were extracted at each indicated transcription elongation time point. Total RNA (100–140 μg) was used for biotinylation and purification of 4sU-labeled nascent RNAs. See the Supplemental Experimental Procedures for details.

rRNA Depletion, RNase R Treatment, qRT-PCR, RNA-Seq of 4sU-Labeled RNA

Prior to nascent RNA-seq library construction, rRNA was depleted from 4sU-labeled RNAs as previously described (Yang et al., 2011; Yin et al., 2015). The ribo$^-$ nascent RNAs were further used for qRT-PCR and RNA-seq. RNase R treatment was performed as described (Zhang et al., 2016).

Construction of Cell Lines with WT, E1126G, and R749H Pol II

293FT cells were transfected individually with vectors expressing Am^r WT human Pol II large subunit (Rpb1) (Rosonina and Blencowe, 2004), or Am^r Pol II carrying a single point mutation, E1126G or R749H (de la Mata et al., 2003; Fong et al., 2014). The experimental procedures measuring TER and back-splicing efficiency are described in Figure 3C.

Tet-on circRNA Expression Vector

An *mCherry* exon flanked by a pair of ICSs (Zhang et al., 2014) was inserted into the intron (between the two 1/2 *EGFP* exons) of *EGFP* to obtain a

construct that expresses the back-spliced circular *mCherry* RNA and spliced linear *egfp* mRNA with high efficiency. The CMV promoter was replaced with a Tet-on promoter to activate transcription in the presence of doxycycline (Dox). The plasmid was stably transfected into HeLa cells with the stable expression of reverse tetracycline-controlled transactivator (rtTA).

Knockout of the *circGCN1L1* Downstream ICS by CRISPR-Cas9

The inverted complementary sequence located in the downstream intron of *circGCN1L1* was annotated as described (Zhang et al., 2014). Two single-guided RNAs (sgRNAs) across the ICS and Cas9 expression vector were transfected into PA1 cells to disrupt the endogenous RNA pairing. The sgRNA sequences are listed in Table S5. Positive double-knockout clones were selected.

Calculation of Transcription Elongation Rate

A stringent computational pipeline (TERate) was developed to measure Pol II TER for all expressed genes. Non-4sU-tagged reads were removed for calculation. A comparison of this method with another published method (Fuchs et al., 2014) was performed. See the Supplemental Experimental Procedures for details.

Back-Splicing Junction and circRNA Prediction

BSJ reads for the steady-state and nascent circRNAs were predicted by CIRCexplorer (Zhang et al., 2014). The BSJ reads of each circRNA were scaled to RPM by normalizing with sequencing depth to represent the relative level of circRNA expression. Wilcoxon rank-sum test was used to compare between different samples. Newly synthesized and steady-state circRNAs in PA1, H9, and H9 differentiated FB cells are listed in Tables S1, S2, S3, and S4.

Statistical Analysis

Statistically significant difference was assessed using Wilcoxon rank-sum test with R platform (R v.3.2.2), and statistical significance was set at $p < 0.05$. To evaluate the relevant correlations between two group datasets, Pearson correlation coefficient (PCC) was also performed with R platform (R v.3.2.2).

ACCESSION NUMBERS

The accession number for the raw sequencing datasets and bigWig track files of 4sUDRB-seq and Ribo$^-$ RNA-seq from PA1, H9, and H9-differentiated FB cells reported in this paper is NCBI GEO: GSE73325.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.03.058.

AUTHOR CONTRIBUTIONS

L.-L.C. and L.Y. conceived and designed the project. Y.Z. designed and performed experiments with the help from X.L., J.Z., S.C., and J.-L.Z.; W.X. performed bioinformatics analyses; L.-L.C., L.Y., Y.Z., and W.X. analyzed the data. L.-L.C. wrote the paper with input from the authors.

ACKNOWLEDGMENTS

We are grateful to G. Carmichael for reading of the manuscript, B. Blencowe for FLAG-Pol II-WT plasmid, X.-O. Zhang and R. Dong for the help in computational analysis, F.-H. Fang for technical support on RNA-seq library preparation, and all lab members for helpful discussion. H9 cells were obtained from the WiCell Research Institute. This work was supported by grants 2014CB964802 and 2014CB910600 from MOST, 91440202, 91540115, and 31271390 from NSFC, and XDA01010206 from CAS.

Received: September 6, 2015
Revised: February 18, 2016
Accepted: March 14, 2016
Published: April 7, 2016

REFERENCES

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

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. Cell *152*, 1252–1269.

Chen, L.-L. (2016). The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol. *17*, 205–211. http://dx.doi.org/10.1038/nrm.2015.32.

Chen, T., Xiang, J.F., Zhu, S., Chen, S., Yin, Q.F., Zhang, X.O., Zhang, J., Feng, H., Dong, R., Li, X.J., et al. (2015). ADAR1 is required for differentiation and neural induction by regulating microRNA processing in a catalytically independent manner. Cell Res. *25*, 459–476.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. Cell *160*, 1125–1134.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525–532.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking distinct RNA populations using efficient and reversible covalent chemistry. Mol. Cell *59*, 858–866.

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. *28*, 2663–2676.

Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol. *15*, R69.

Fuchs, G., Voichek, Y., Rabani, M., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2015). Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with 4sUDRB-seq. Nat. Protoc. *10*, 605–618.

Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. Genome Biol. *15*, 409.

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.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., and Rajewsky, N. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep. *10*, 170–177.

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.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Kramer, M.C., Liang, D., Tatomer, D.C., Gold, B., March, Z.M., Cherry, S., and Wilusz, J.E. (2015). Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. Genes Dev. *29*, 2168–2182.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. Nat. Struct. Mol. Biol. *22*, 256–264.

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.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. *29*, 436–442.

Rädle, B., Rutkowski, A.J., Ruzsics, Z., Friedel, C.C., Koszinowski, U.H., and Dölken, L. (2013). Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. J. Vis. Exp. *78*, e50195.

Reubinoff, B.E., Itsykson, P., Turetsky, T., Pera, M.F., Reinhartz, E., Itzik, A., and Ben-Hur, T. (2001). Neural progenitors from human embryonic stem cells. Nat. Biotechnol. *19*, 1134–1140.

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 *10*, 581–589.

Rybak-Wolf, A., Stottmeister, C., Glažar, 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.

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.

Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-type specific features of circular RNA expression. PLoS Genet. *9*, e1003777.

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.H., and Bindereif, A. (2015). Exon circularization requires canonical splice signals. Cell Rep. *10*, 103–111.

Wang, Y., and Wang, Z. (2015). Efficient backsplicing produces translatable circular mRNAs. RNA *21*, 172–179.

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.

Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., and Chen, L.L. (2011). Genomewide characterization of non-polyadenylated RNAs. Genome Biol. *12*, R16.

Yin, Q.F., Chen, L.L., and Yang, L. (2015). Fractionation of non-polyadenylated and ribosomal-free RNAs from mammalian cells. Methods Mol. Biol. *1206*, 69–80.

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.

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., Yang, L., and Chen, L.L. (2016). Characterization of Circular RNAs. Methods Mol. Biol. *1402*, 215–227.

Cell Reports, Volume *15*

Supplemental Information

The Biogenesis of Nascent Circular RNAs

Yang Zhang, Wei Xue, Xiang Li, Jun Zhang, Siye Chen, Jia-Lin Zhang, Li Yang, and Ling-Ling Chen

Extended Experimental Procedures

HESCs maintenance and differentiation

H9 hESCs were maintained on irradiated-MEF feeder cells and passaged weekly as described previously (Chen et al., 2015). H9 cells were induced to differentiate into forebrain neural cells as described (Chen et al., 2015). In brief, hESCs were detached from MEF cells and suspended in hESCs medium without FGF2 for 4 days in Petri dish. After the aggregates floated in neural induction medium consisting of DMEM/F12, N2 supplement, 1xNEAA and heparin for additional 2-3 days, the aggregates were induced to adhere to Laminin coated substrate, and primitive neuroepithelial (NE) would be observed on days 8-10, followed by the definitive NE cells in the center of colonies occurring at days 14-17 of differentiation. The definitive NE were gently blown off (day 17) and suspended in NIM for one week and then replated on Laminin-coated substrate for terminal differentiation. The nascent RNAs from undifferentiated (H9 D0) and differentiated (FB D26) cells were subjected to qRT-PCR and RNA-seq.

Metabolic labeling of nascent RNA with 4sU and nascent RNA purification

Metabolic labeling of newly transcribed RNA was performed as described (Fuchs et al., 2014; Radle et al., 2013) with modifications. Differentiated or undifferentiated cells were incubated with 100 μM DRB (Sigma, D1916) for 3 hr to block Pol II transcription. Transcription was recovered after DRB release and newly transcribed RNA was labeled with 300 μM 4sU (Sigma, T4509). TRizol (Invitrogen) was added to stop transcription and total RNA was extracted at each indicated transcription elongation time point. Total RNA (100-140 μg) was used for biotinylation and purification of 4sU-labeled nascent RNA. 4sU-labeled RNA was incubated in biotinylation buffer (10 mM Tris pH 7.4, 1 mM EDTA) with 0.2 mg/ml EZ-link biotin-HPDP (Pierce, 21341), which was dissolved in dimethylformamide (DMF, Sigma, D4551) at a concentration of 1 mg/ml at room temperature for 1.5 hr with

rotation. For MTSEA biotin-XX (Biotium) biotinylation, 4sU-labeled RNA was incubated with 0.02 mg/ml MTSEA biotin-XX at room temperature for 1.5 hr with rotation (Duffy et al., 2015). Unbound biotin-HPDP was removed by equal volume chloroform extraction twice, and RNA was precipitated at 13,000 rpm for 15 min at 4 ℃ with 1:10 volume of 5 M NaCl and an equal volume of isopropanol. The RNA pellet was washed twice with 75% ethanol and resuspended in 150 μl DEPC•H_2O. RNA quality was checked by electrophoretical analysis to exclude RNA degradation. 4sU-labeled and unlabeled RNA was separated by using Streptavidin-coated magnetic beads (Invitrogen). Biotinylated RNA was incubated with 150 μl Streptavidin beads at room temperature for 20 min. Beads were washed four times with 0.9 ml of 65 ℃ washing buffer (100 mM Tris pH 7.4, 10 mM EDTA, 1 M NaCl, 0.1% Tween 20), followed by four times with 0.9 ml of room temperature washing buffer. Nascent RNA was eluted with 100 μl 0.1 M dithiotheitol (DTT) twice and was precipitated with 40 μl of 4 M LiCl, 2 μl glycogen and 600 μl ice-cold ethanol.

Ribosomal RNA depletion, RNase R treatment, qRT-PCR, RNA-seq of 4sU labeled RNAs

Prior to construct nascent RNA-seq library, rRNA was depleted from 4sU-labeled RNA as described (Yang et al., 2011; Yin et al., 2015). Ribosomal RNA removal was carried out with RiboMinus™ Human/Mouse Transcriptome Isolation Kit (Invitrogen, K1550-01) according to the manufacturer's protocol. The Ribo– nascent RNA was further applied for qRT-PCR and RNA-seq. RNase R treatment was performed as described (Zhang et al., 2014; Zhang et al., 2016). 4sUDRB-seq and Ribo– RNA-seq libraries were prepared according to the manufacturer's instructions, and then applied for sequencing on Illumina Hiseq 2000 by CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China. Primers for PCR amplification were listed in Table S5.

Construct cell lines with wt, E1126G and R749H Pol II

293FT cells were transiently transfected individually with vectors expressing α-amanitin-resistant (Am^r) human Pol II large subunit (Rpb1) (wt Pol II) (Rosonina and Blencowe, 2004), Am^r fast Pol II carrying a single point mutation (E1126G), or Am^r slow Pol II carrying a single point mutation (R749H) (de la Mata et al., 2003; Fong et al., 2014) by using Mut Express MultiS Fast Mutagenesis Kit (Vazyme C213-01). 24 hr after post-transfection, 10 μg/ml α-amanitin was added to block endogenous Pol II transcription during each the following step (Figure 3B). 24 hr later, DRB was added with a final concentration of 100 μM for 3 hr to block transcription. Newly transcribed RNA was labeled with 300 μM 4sU after DRB removal. Total RNA was extracted at 30 min or 120 min after transcription and followed by 4sU-labeled nascent RNA purification, which was further subjected to cDNA preparation. Nascent RNA from 30 min of 4sU labeling was used to measure the transcription elongation rate by quantifying pre-RNA accumulation at proximal (P) and distal (D) regions with respect to TSSs. Nascent RNA from 120 min of 4sU labeling was used to quantify the back-splicing efficiency.

Half-life analysis of circRNA

PA1 cells were seeded in 3.5 cm dish at a density of 1×10^6 cells and cultured for 24 hr. Fresh medium containing 10 μg/ml of actinomycin D (Sigma) was added to each well and total RNA was isolated at indicated time points. The relative abundance of each circRNA and its linear mRNA counterpart was detected by qRT-PCR.

Calculation of transcription elongation rate with a new computational method

A step-by-step computational pipeline (TERate) was developed to evaluate the transcription elongation rate (TER) of RNA Pol II for all expressed genes by counting normalized average hits. Briefly, each intron was split into 300-bp adjacent bins, and the expressed signal of each 300-bp bin was quantified by Average Hits, which counts average mapped bases within each 300-bp bin by the following equation:

$$Average\ Hits = \frac{\sum_{i=1}^{n}(B_i \times D_i)}{300}$$

where B_i and D_i represent mapped bases (B_i) and their corresponding sequencing depth (D_i). The mean values of Average Hits for 300-bp bins at 10 and 15 min of 4sU labeling after DRB removal is 3.97 and 5.30, respectively. A series of stringent cutoffs were chosen to discriminate signals from backgrounds. About 1,000 of 300-bp intron bins within “TSS + 10 kb” region (+/-1kb) were randomly selected for “expressed bins”. Another 1,000 of 300-bp intron bins within “TES - 5 kb” region (+/-1kb) of very long genes (≥ 60 kb) were randomly selected for “background bins”. As indicated in Figure S3A, 80% of expressed bins were expressed with a cutoff at Average Hits ≥ 5, and at the same time over 90% of background bins were expressed at Average Hits < 5. Thus, Average Hits ≥ 5 was selected to determine whether a region has been transcribed or not, i.e., intron bins with Average Hits ≥ 5 were defined as “transcribed bins”, and other bins with Average Hits < 5 were considered as “noise bins". To eliminate non-specific signals in 4sUDRB-seq and enhance the accuracy of elongation rate estimation, only genes ≥ 30 kb were chosen for calculation. In addition, for genes with multiple isoforms and different TSSs, only the isoform with the nearest distance between TSS and start peak (a high signal in the proximity of the TSS) was selected. In addition, genes with at least 30 of “transcribed bins” were selected for transcription elongation rate calculation. For each selected gene, the position of the last two continuous “transcribed bin” was defined as the Pol II transcription edges, and then the transcription elongation distance for a specific gene was calculated from its TSS to the Pol II transcription edge. Finally, the transcription elongation rate (V) was determined by dividing the transcription elongation distance (L, kb) by transcription time (T, min) with the following equation:

$$V = \frac{L}{T}$$

The final Pol II TER of each gene was calculated as the average value of its Pol II TERs from 10 and 15 min of 4sU-labeling after DRB removal.

In addition, the TER could be also calculated by linearly fitting the 0, 10 and 15 min edges, from which the slope of the linear fit was determined as the final TER (Fuchs et al., 2014). Since the correlation of average elongation rate and the slope of the linear fit for the elongation rate is very high (with Pearson Correlation Coefficient in PA1 cells as 0.913, in H9 as 0.919 and in FB as 0.935, respectively), the average value of TER was used for most analyses in this study.

To remove non-4sU-tagged reads, the transcribed region (Figure S3B, red arrows) of a given gene was first estimated by multiplying the TER with 4sU exposure time (10, 15, 30, 60 and 120 min) at each time point. 4sUDRB-seq reads mapped beyond this region were treated as contamination and removed before nascent circRNA calling. Notably, most nascent circRNAs were predicted with longer 4sU exposure time at 30, 60 and 120 min, and circRNA signals identified at shorter 4sU labeling time were largely from steady-state circRNA contamination.

The source codes of TERate pipeline and related documents can be accessed from https://github.com/YangLab/TERate.

Back-splicing junction and circRNA prediction

Back-splicing junction reads for nascent and steady-state circRNAs were predicted by CIRCexplorer (Zhang et al., 2014) and other methods, including find_circ (Memczak et al., 2013) and MapSplice (Jeck et al., 2013) (data not shown), with human annotations (human refFlat.txt updated at 2013/10/13 and hg19 knownGene.txt updated at 2013/6/30). Back-splicing junction reads for each circRNA were scaled to RPM (mapped back-splicing junction Reads Per Million mapped reads) that represents the level of circRNA expression after normalization by the sequencing depths. Highly expressed circRNAs were defined

with RPM ≥ 0.1 as previously reported (Zhang et al., 2014). *CDR1as* (Memczak et al., 2013) was added to the final list of circRNA annotation. Wilcoxon rank-sum test was used for the comparison between different samples. Newly synthesized and steady-state circRNAs in PA1, hESCs H9 and H9-differentiated FB cells are listed in Tables S1, S3 and S4.

Length distribution of circRNA-producing or non-circRNA-producing genes

Lengths of circRNA-producing or non-circRNA-producing genes were extracted from gene annotations (human refFlat.txt updated at 2013/10/13 and hg19 knownGene.txt updated at 2013/6/30). One thousand randomly selected non-circRNA-producing genes were set up as controls. For genes with multiple isoforms, only the length of the longest isoform was used.

Identification of splicing junction reads

Nascent back-splicing junction reads were predicted by CIRCexplorer (Zhang et al., 2014). Canonical splicing (including exon skipping events) junction reads were obtained from TopHat2 mapping results. The distribution of average junction reads of splicing and back-splicing events was plotted.

Analysis of complementary sequences across introns

Upstream and downstream flanking intron sequences of nascent circRNAs were fetched for complementary sequence analysis as described (Zhang et al., 2014). The paired complementary sequences should be at least 50 bp long with 80% match in sequences.

The first intron length

Most circRNAs are produced from internal exons, especially the second exons. Nascent circRNAs produced from the second exons usually have long upstream (first) introns of their host genes (Figure S5B). One thousand randomly selected non-circRNA producing genes were set up as controls. Length distribution of first introns was compared between

nascent circRNA-producing genes and control genes. Wilcoxon rank-sum test was used for the comparison between two sets of samples.

Evaluation of mRNA expression and pre-mRNA expression

Gene expression of mRNAs was determined by RPKM (Mortazavi et al., 2008). Pre-mRNA expression was evaluated by the average RPKM of introns in 4sUDRB-seq datasets. Pre-mRNAs with at least 1 RPKM were selected for comparison. Pearson Correlation Coefficient (Mukaka, 2012) was applied to evaluate relevant correlations.

Supplemental Figures and Figure Legends

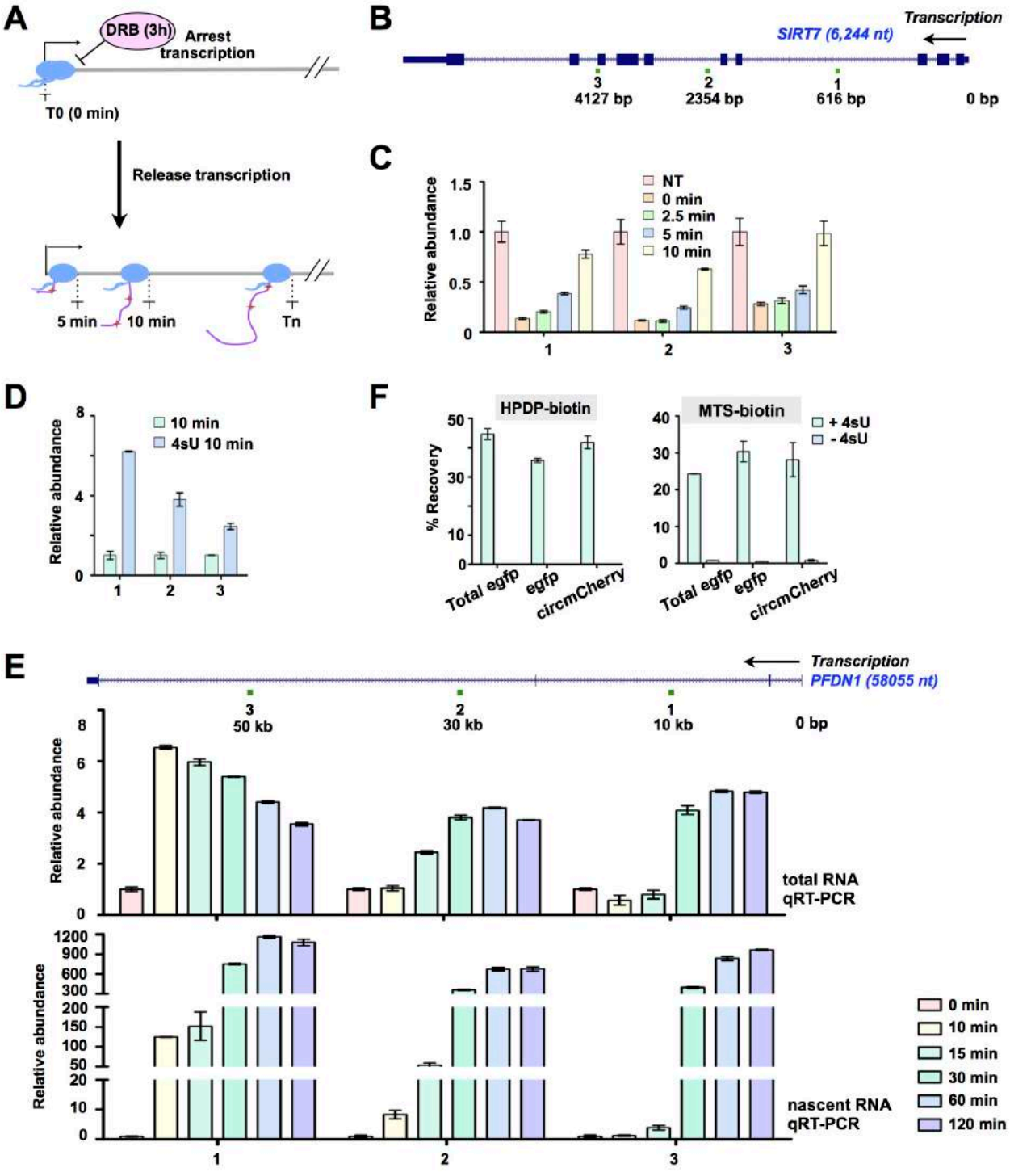


Figure S1. Validation of 4sU metabolic labeling for newly transcribed RNA (related to Figure 1).

(A) A workflow of 4sU tagging to label nascent RNA.

(B) A schematic drawing of *SIRT7* gene and prime sets (green bars) used in (**C**) and (**D**).

(C) DRB reversibly blocks transcription. PA1 cells were treated with DRB for 3 hr, and total RNA was harvested at indicated time points after DRB removal. qRT-PCR of three primer sets of *SIRT7* showed that DRB efficiently blocked transcription and transcription could be recovered after DRB removal. NT, non-treated.

(D) The 4sU-labeled newly transcribed RNA is purified by 4sU-DRB. PA1 cells were treated as described in Figure S1C and then 4sU-labeled newly transcribed RNA was purified as revealed by qRT-PCR of nascent *sirt7* pre-mRNA.

(E) The kinetics of newly transcribed RNA with 4sU labeling. Top, a schematic drawing of *PFDN1* and primer sets used in qRT-PCR. Middle, qRT-PCR of total RNA extracted at indicated time points of 4sU labeling. Bottom, qRT-PCR of newly transcribed RNA purified from indicated time points of 4sU labeling.

(F) Purification of the 4sU-labeled *circmCherry* and linear *egfp* by HPDP-biotin and MTS-biotin. The recovery rates (4sU-purified RNA v.s. total RNA in each case) of the nascent 4sU-labeled *circmCherry* and linear *egfp* by HPDP-biotin and MTS-biotin were measured by qRT-PCR in each purification.

A

Time pulse of 4sU tagging	Total reads	# of mapped reads	# of exonic reads	# of intronic reads	# of intergenic reads
steady-state	125,735,340	117,452,419	83,681,960	20,324,095	13,446,364
0 min	5,708,813	5,335,816	2,937,223	1,115,163	1,283,430
10 min	24,064,138	22,695,430	5,486,329	12,648,216	4,560,885
15 min	53,660,428	50,048,424	7,121,664	33,689,060	9,237,700
30 min	96,083,184	88,441,560	11,314,459	62,439,600	14,687,501
60 min	192,488,809	175,855,492	25,359,921	123,990,885	26,504,686
120 min	260,122,985	239,313,021	51,992,356	157,589,210	29,731,455
4 hr	386,745,248	312,651,920	114,386,464	160,631,197	37,634,259
16 hr	385,816,787	314,047,601	168,256,096	79,350,243	66,441,263

B

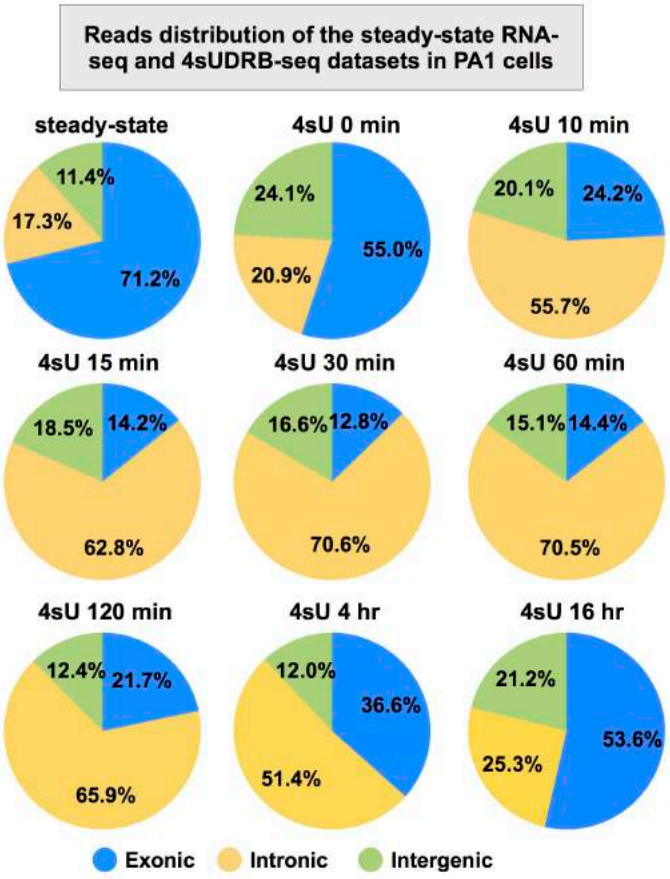


Figure S2. Reads distribution of steady-state RNA-seq and 4sUDRB-seq in PA1 cells (related to Figures 1 and 2).

(A) The numbers of total detected total reads, mapped reads, mapped reads to exons, introns and intergenic regions are listed at each indicated 4sU labeling time point.

(B) The reads distribution of steady-state RNA-seq and 4sUDRB-seq at each time point in PA1 cells. Note that the effective reads mapping to exons and introns are different between steady-state RNA-seq and 4sUDRB-seq at each time point.

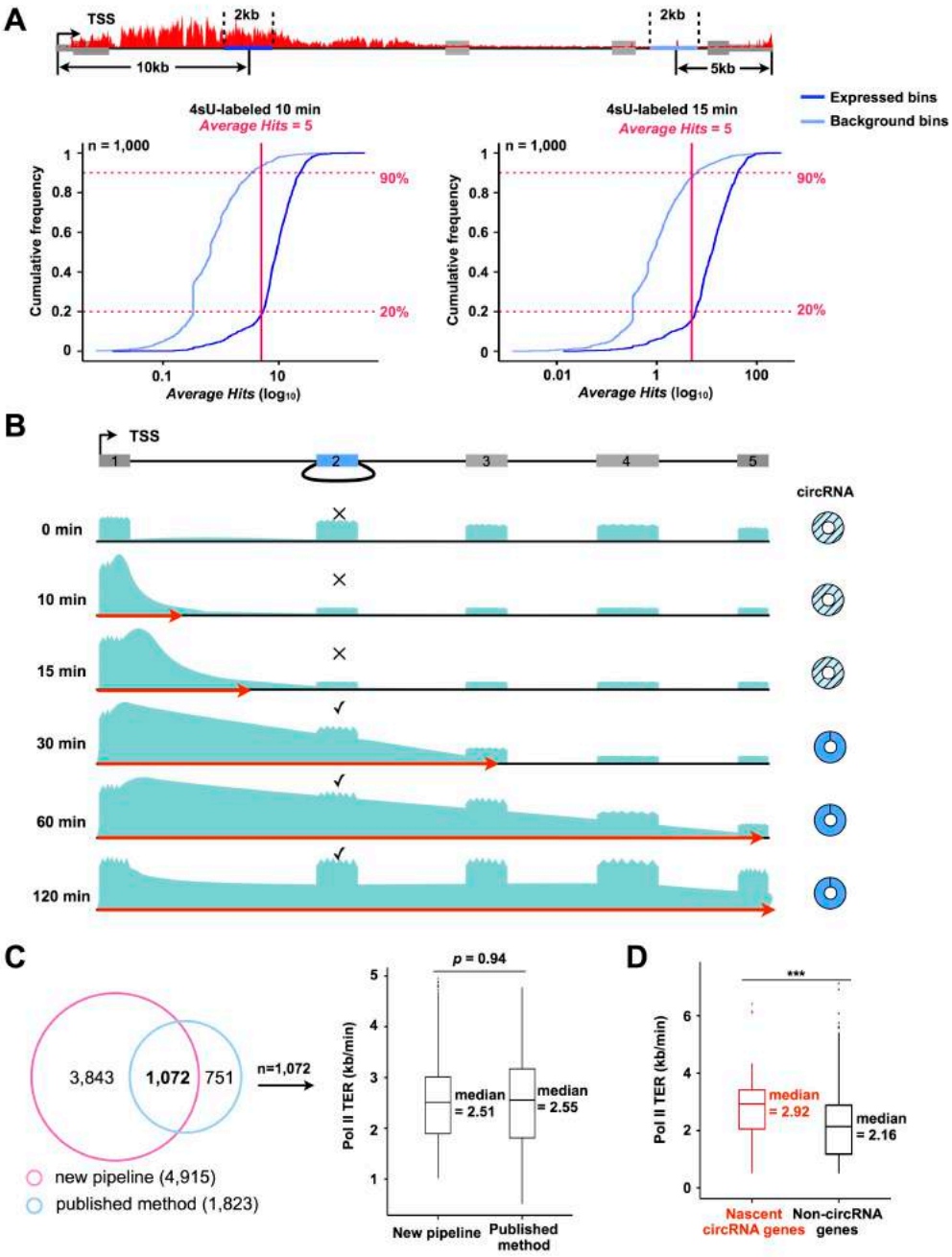


Figure S3. A customized computational pipeline to calculate Pol II TER (related to Figures 1 and 3 and Extended Experimental Procedures).

(A) A schematic drawing of setting up new computational parameters to calculate the Pol II transcription elongation rate (TER). See Extended Experimental Procedures for details.

(B) Removal of 4sUDRB-seq contamination. Due to the non-specific binding during purification, non-4sU-tagged RNAs could also be co-purified with 4sU-labeled nascent RNAs, and thus detected by 4sUDRB-seq, in particular during the short 4sU exposure time. These non-4sU-tagged RNAs include spliced mRNAs and steady-state circRNAs, which are needed to be removed as contamination from the TER calculation. To remove these non-4sU-tagged reads, transcribed region (red arrows) of a given gene was first calculated by multiplying TER with 4sU exposure time (10, 15, 30, 60 and 120 min) at each time point. Mapped 4sUDRB-seq reads beyond such a region were treated as contamination and were removed prior to nascent circRNA calling. Notably, most nascent circRNAs were predicted with longer 4sU exposure time points and circRNA signals identified at shorter exposure time were likely from the steady-state circRNA contamination.

(C) Comparison of different computational methods for calculating Pol II TER. Left, the TERs of 4,915 and 1,823 genes were obtained with our method (magenta circle) and published method (blue circle) (Fuchs, 2014), respectively. Right, among 1,072 overlapping genes, the Pol II TERs calculated by two methods have no apparent difference. *p* value was obtained from Wilcoxon rank-sum test.

(D) The Pol II TER of circRNA-producing genes is higher than that of non-circRNA producing genes in PA1 cells. Pol II TERs were calculated by the published method (Fuchs et al., 2014), from which 167 circRNA-producing genes were calculated. ***$p = 3.2 \times 10^{-7}$, Wilcoxon rank-sum test.

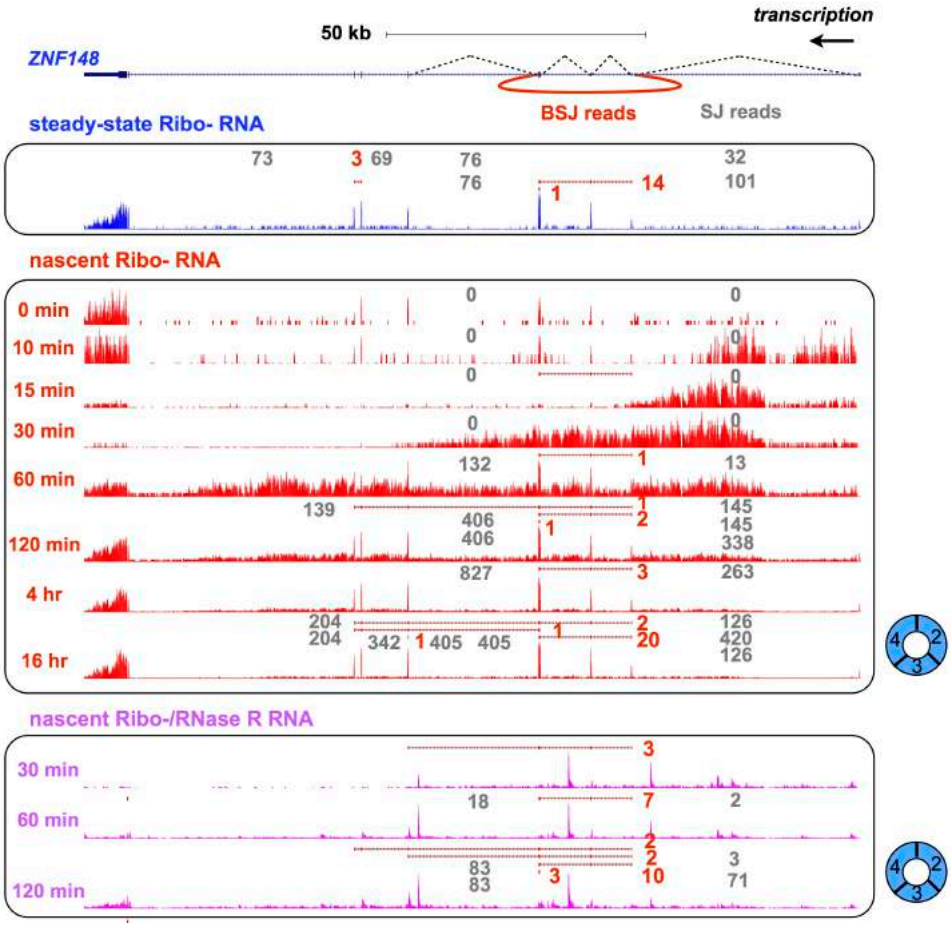


Figure S4. An example of 4sUDRB-seq for nascent circRNAs produced from *ZNF148* in PA1 cells (related to Figure 2). See Figure 2A for details. Note that alternative circularization of newly synthesized circRNAs were also identified at this locus.

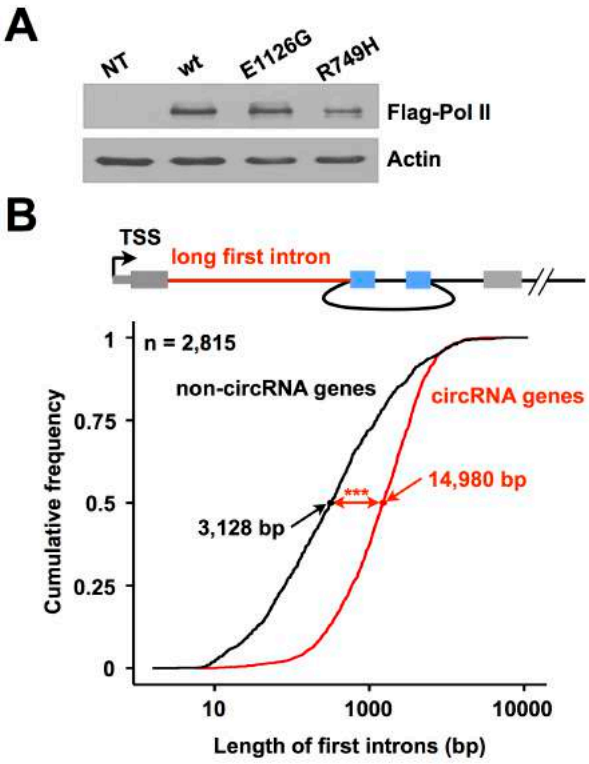
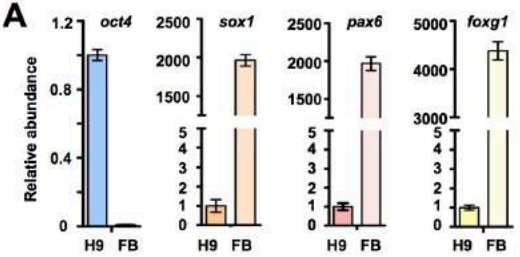


Figure S5. Characterization of nascent circRNA processing (related to Figure 3).

(A) Western blots of 293FT cells expressing wt or Flag-Pol II mutants.

(B) CircRNA-producing genes usually have longer first introns (red line), compared to non-circRNA producing genes (black line). ***p = 1.28 × 10^{-110}, Wilcoxon rank-sum test.

A



B

D0 H9	Total reads	Total SJ reads	Upstream SJ reads	BSJ reads	Downstream SJ reads	Nascent circRNAs	Host genes
0 min	1,630,264	N/A	N/A	N/A	N/A	0	0
10 min	25,523,929	210,259	23	5	35	5	5
15 min	55,747,344	485,261	31	7	60	6	6
30 min	89,718,379	1,610,114	4,642	141	5,601	115	115
60 min	104,382,001	5,651,755	78,471	1197	84,900	820	700
120 min	117,792,209	8,316,677	195,657	2474	241,366	1528	1121

D26 FB	Total reads	Total SJ reads	Upstream SJ reads	BSJ reads	Downstream SJ reads	Nascent circRNAs	Host genes
0 min	4,713,960	N/A	N/A	N/A	N/A	0	0
10 min	20,274,512	48,971	1	9	3	4	4
15 min	59,032,425	348,777	344	99	353	54	52
30 min	130,880,669	1,016,133	2,757	215	2,760	156	150
60 min	286,025,212	5,645,333	74,909	1,420	81,380	876	715
120 min	286,511,613	12,977,304	314,159	4,057	369,364	2,171	1,454

C

D0 H9	# of transcribed genes	# of completely transcribed genes	% of transcribed genes
10 min	10,497	4,383	41.75%
15 min	13,119	8,139	62.04%
30 min	14,775	11,835	80.10%
60 min	15,128	13,816	91.33%
120 min	16,541	16,032	96.92%

D26 FB	# of transcribed genes	# of completely transcribed genes	% of transcribed genes
10 min	8,184	2,716	33.19%
15 min	12,438	7,771	62.48%
30 min	14,796	11,913	80.52%
60 min	17,242	15,800	91.64%
120 min	17,519	16,981	96.93%

Figure S6. Analysis of nascent circRNA processing in H9 cells and H9 differentiated FB neurons (related to Figure 6).

(A) Expression of maker genes of H9 cells and H9 differentiated FB neurons.

(B) The identified BSJ reads are low compared to canonical splicing events at the same exons. The numbers of total detected reads, SJ reads, upstream and downstream SJ Reads, BSJ Reads, nascent circRNAs and their host genes at each indicated transcription elongation time point in H9 and FB cells are listed.

(C) The number and percentage of completely transcribed genes revealed by 4sUDRB-seq in H9 cells (top) and FB neurons (bottom).

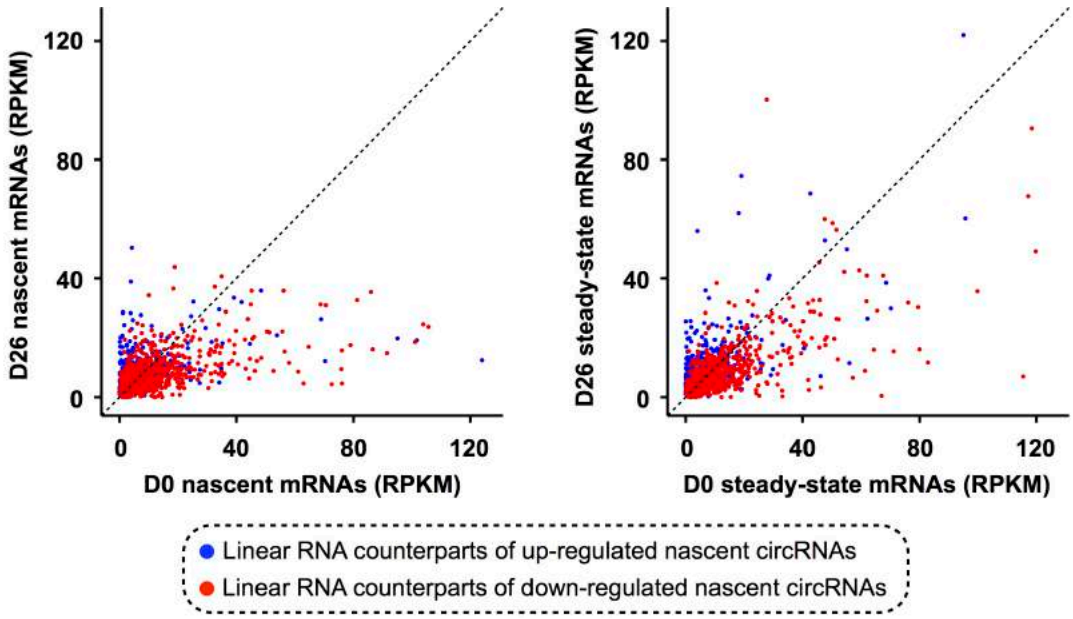


Figure S7. Nascent and steady-state expression of linear RNAs upon neuronal differentiation (related to Figure 7). Linear RNA counterparts corresponding to the up-regulated (blue dots) or down-regulated (red dots) nascent circRNAs in H9 cells and H9 differentiated FB neurons shown in Figure 7A were analyzed. Their nascent (left) and steady-state (right) levels were comparable in H9 cells and H9 differentiated FB neurons.

Supplemental Tables

Supplemental Table 1. List of nascent and steady-state circRNAs in PA1 cells (related to Figures 1-4). Nascent and steady-state circRNAs were determined from 4sUDRB-seq at 10 min, 15 min, 30 min, 60 min,120 min, 4 hr and 16 hr time points or Ribo– RNA-seq in PA1 cells by CIRCexplorer (Zhang et al., 2014), listed with their genomic locations, strand information, exon numbers/sizes/offsets, flanking introns and parent gene symbols.

Supplemental Table 2. List of two types of nascent circRNAs in PA1 cells (related to Figure 4). Two types of nascent circRNAs detected from 4sUDRB-seq in PA1 cells described in Figure 4E. The BSJ reads of newly produced circRNAs and their upstream and downstream SJ reads were listed for comparison.

Supplemental Table 3. List of nascent and steady-state circRNAs in hESC H9 cells (related to Figures 6-7). Nascent and steady-state circRNAs were determined from 4sUDRB-seq at 10, 15, 30, 60 and 120 min time points or Ribo– RNA-seq in H9 cells. See Supplemental Table 1 for details.

Supplemental Table 4. List of nascent and steady-state circRNAs in H9 cells differentiated FB neurons (related to Figures 6-7). Nascent and steady-state circRNAs were determined from 4sUDRB-seq at 10, 15, 30, 60 and 120 min time points or Ribo– RNA-seq in FB neurons. See Supplemental Table 1 for details.

Supplemental Table 5. All primers used in the study (related to Figures 1-6).

References

Chen, T., Xiang, J.F., Zhu, S., Chen, S., Yin, Q.F., Zhang, X.O., Zhang, J., Feng, H., Dong, R., Li, X.J., et al. (2015). ADAR1 is required for differentiation and neural induction by regulating microRNA processing in a catalytically independent manner. Cell Res. *25*, 459-476.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525-532.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Mol. Cell *59*, 858-866.

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. *28*, 2663-2676.

Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol. *15*, R69.

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.

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.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621-628.

Mukaka, M.M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. *24*, 69-71.

Radle, B., Rutkowski, A.J., Ruzsics, Z., Friedel, C.C., Koszinowski, U.H., and Dolken, L. (2013). Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. J. Vis. Exp. *78*, e50195.

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 *10*, 581-589.

Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., and Chen, L.L. (2011). Genomewide characterization of non-polyadenylated RNAs. Genome Biol. *12*, R16.

Yin, Q.F., Chen, L.L., and Yang, L. (2015). Fractionation of non-polyadenylated and ribosomal-free RNAs from mammalian cells. Methods Mol. Biol. *1206*, 69-80.

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., Yang, L., and Chen, L.L. (2016). Characterization of circular RNAs. Methods Mol Biol, *1402*, 215-27.