Short intronic repeat sequences facilitate circular RNA production

Dongming Liang and Jeremy E. Wilusz

Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA

Recent deep sequencing studies have revealed thousands of circular noncoding RNAs generated from protein-coding genes. These RNAs are produced when the precursor messenger RNA (pre-mRNA) splicing machinery "backsplices" and covalently joins, for example, the two ends of a single exon. However, the mechanism by which the spliceosome selects only certain exons to circularize is largely unknown. Using extensive mutagenesis of expression plasmids, we show that miniature introns containing the splice sites along with short (~30- to 40-nucleotide) inverted repeats, such as Alu elements, are sufficient to allow the intervening exons to circularize in cells. The intronic repeats must base-pair to one another, thereby bringing the splice sites into close proximity to each other. More than simple thermodynamics is clearly at play, however, as not all repeats support circularization, and increasing the stability of the hairpin between the repeats can sometimes inhibit circular RNA biogenesis. The intronic repeats and exonic sequences must collaborate with one another, and a functional 3′ end processing signal is required, suggesting that circularization may occur post-transcriptionally. These results suggest detailed and generalizable models that explain how the splicing machinery determines whether to produce a circular noncoding RNA or a linear mRNA.

Most of the eukaryotic genome is transcribed, yielding a complex repertoire of transcripts that includes tens of thousands of individual noncoding RNAs with little or no predicted protein-coding capacity (for review, see Wilusz et al. 2009; Cech and Steitz 2014; Yang et al. 2014). Although relatively few noncoding RNAs have been assigned a function, their existence is a challenge to the long-held assumption that most genetic information is expressed as and carried out by proteins. Unexpectedly, even protein-coding regions of the genome—including genes with highly evolutionarily conserved ORFs—generate noncoding RNAs. Unlike linear messenger RNAs (mRNAs), these noncoding RNAs often have covalently linked ends, making them circles (for review, see Wilusz and Sharp 2013; Jeck and Sharpless 2014). A handful of such circular RNAs were serendipitously found over the past 25 years (Nigro et al. 1991; Capel et al. 1993; Cocquerelle et al. 1993; Zaphiropoulos 1997; Burd et al. 2010), but it is becoming increasingly clear that this class of noncoding RNA is much larger than previously appreciated (Salzman et al. 2012). Over 25,000 circular RNAs, derived from ~15% of actively tran-

scribed genes, were recently identified in a single human cell type (Jeck et al. 2013). These highly stable transcripts contain almost exclusively exonic sequences, localize to the cytoplasm, and, at least in two cases, function as efficient modulators of microRNA activity (Hansen et al. 2013; Memczak et al. 2013). Thousands of circular RNAs have additionally been found in numerous other eukaryotes (Salzman et al. 2012; Memczak et al. 2013; Guo et al. 2014; Wang et al. 2014) as well as in archaea (Danan et al. 2012). For some genes, the abundance of the circular RNA exceeds that of the associated linear mRNA by a factor of 10, raising the interesting possibility that the function of some protein-coding genes may actually be to produce circular noncoding RNAs, not proteins.

Because eukaryotes contain split, intron-containing genes, their precursor mRNAs (pre-mRNAs) must be modified such that introns are removed and exons are joined together. If a pre-mRNA is spliced in the standard way (meaning exon 1 is joined to exon 2, which is joined to exon 3, etc.), a linear mRNA is generated that can be

subsequently translated to produce a protein. Alternatively, the splicing machinery can "backsplice" and join a splice donor to an upstream splice acceptor, thereby yielding a circular RNA that has covalently linked ends. Supporting this biogenesis model, canonical splicing signals generally immediately flank regions that circularize, and exon circularization can be observed when splicing substrates are incubated in mammalian nuclear extracts (Braun et al. 1996; Pasman et al. 1996; Schindewolf et al. 1996). Once generated, circular RNAs appear to almost always be noncoding, as the genic start and/or stop codons have been removed. Although artificial circular RNAs containing an internal ribosome entry site (IRES) can be translated in cellular extracts (Chen and Sarnow 1995), endogenous circular RNAs have not been found to associate with ribosomes in cells (Jeck et al. 2013; Guo et al. 2014). Therefore, the choice between production of a linear mRNA and a circular RNA must be tightly regulated, as it fundamentally changes the functional output of a gene.

With the exception of the first and last exons of genes, every other exon in the genome has splicing signals at its 5′ and 3′ ends and theoretically can circularize. However, every exon does not circularize, and, in some cases, multiple exons are present in a circular RNA. So how does the spliceosome determine whether a region of a pre-mRNA yields a circle? The kinetics of splicing likely provides a partial answer because introns are generally removed as the gene is still being transcribed (i.e., cotranscriptionally) (for review, see Brugiolo et al. 2013). To allow sufficient time for a backsplice to occur, the intron immediately preceding the circularizing exons likely must be spliced slower than the average intron. Sequence-specific elements may additionally be required for circularization, as introns flanking exons that circularize commonly contain repeat sequences (particularly Alu elements) in an inverted orientation (Jeck et al. 2013). These repeats may base-pair, bringing the splice site sequences into close proximity to each other and facilitating catalysis. Inverted repeats of >15 kb surround the mouse *Sry* exon that circularizes, and deletion analyses have suggested that at least 400 complementary nucleotides are necessary for *Sry* circularization in vivo (Dubin et al. 1995). Artificially surrounding an exon with introns containing 800 nucleotides (nt) of perfectly complementary repeats likewise is sufficient to allow circularization (Hansen et al. 2013). However, it is unclear whether the repeat length rules defined for *Sry* or these artificial expression vectors apply to most genes that circularize. This is because Alu elements, which are only ~300 nt in length, flank many exons that naturally generate circles. Detailed mechanistic models for how circularization generally occurs are therefore currently lacking, limiting our understanding of how circular RNAs are regulated and function.

To reveal the mechanism by which circular RNAs are produced, the pre-mRNAs of the human ZKSCAN1 (zinc finger protein with KRAB and SCAN domains 1), HIPK3 (homeodomain-interacting protein kinase 3), and EPHB4 (ephrin type B receptor 4) genes were cloned into easily manipulatable expression vectors. By extensively mutagenizing these vectors, we identified minimal regions that are sufficient for circular RNA production in cells. Surprisingly, miniature introns (<100 nt) containing only the splice site sequences along with short inverted repeats (~30–40 nt) were sufficient to allow the intervening exons to efficiently circularize. The repeats must base-pair to one another; however, more than simple thermodynamics is clearly at play, as not all intronic repeat sequences support circularization. In fact, for all three genes tested, strengthening the hairpin between the repeats sometimes strongly inhibited circularization. Within the repeat sequences, we found that G-U wobble base pairs as well as poly(A) stretches can limit circular RNA production. We further show that circularization requires a functional 3′ end processing signal as well as collaboration between the intronic repeats and the exonic sequences. These results suggest a generalizable model for how the splicing machinery determines whether to produce a circular RNA or a linear mRNA.

Results

The human ZKSCAN1 gene produces an abundant circular RNA

As a model to study the mechanism by which circular RNAs are produced from protein-coding genes, we first chose to focus on the human ZKSCAN1 gene. The ZKSCAN1 protein is overexpressed in certain gastro-esophageal cancers (van Dekken et al. 2009) and has been suggested to regulate GABA type A receptor expression in the brain (Mulligan et al. 2012). Deep sequencing of the transcriptomes of numerous human cell lines, including HEK293, HeLa, and H1 human embryonic stem cells (hESCs), revealed exon–exon junction reads connecting the 3′ end of exon 3 to the 5′ end of exon 2 (Salzman et al. 2012; Jeck et al. 2013; Memczak et al. 2013). This suggests that exons 2 and 3 of ZKSCAN1 may be spliced together to form a covalently linked 668-nt circular RNA (Fig. 1A). Using Northern blots and an oligo probe complementary to exon 2 of ZKSCAN1, we detected an abundant transcript of the appropriate size for the circular RNA in normal human brain and liver tissues (Fig. 1B) as well as in the Huh7 human hepatoma cell line (Fig. 1C, lane 1). In Huh7 cells, the ZKSCAN1 mRNA is approximately threefold more abundant than the circular RNA. However, the circular RNA is considerably more stable (half-life of >6 h compared with ~2 h for the linear mRNA) as determined by inhibition of transcription via flavopiridol, an elongation inhibitor (Fig. 1C; Supplemental Fig. 1A).

ZKSCAN1 circular RNA expression was further measured using quantitative PCR (qPCR) after reverse transcription on RNAs derived from an independent set of tissue donors (Supplemental Fig. 1B). The highest circular RNA levels were present in the brain, liver, skeletal muscle, and salivary gland, although the transcript was detectable in all tissues examined. Consistent with previous reports (Salzman et al. 2013; Guo et al. 2014), there was no clear correlation between ZKSCAN1 mRNA and circular RNA levels, suggesting that the higher circular
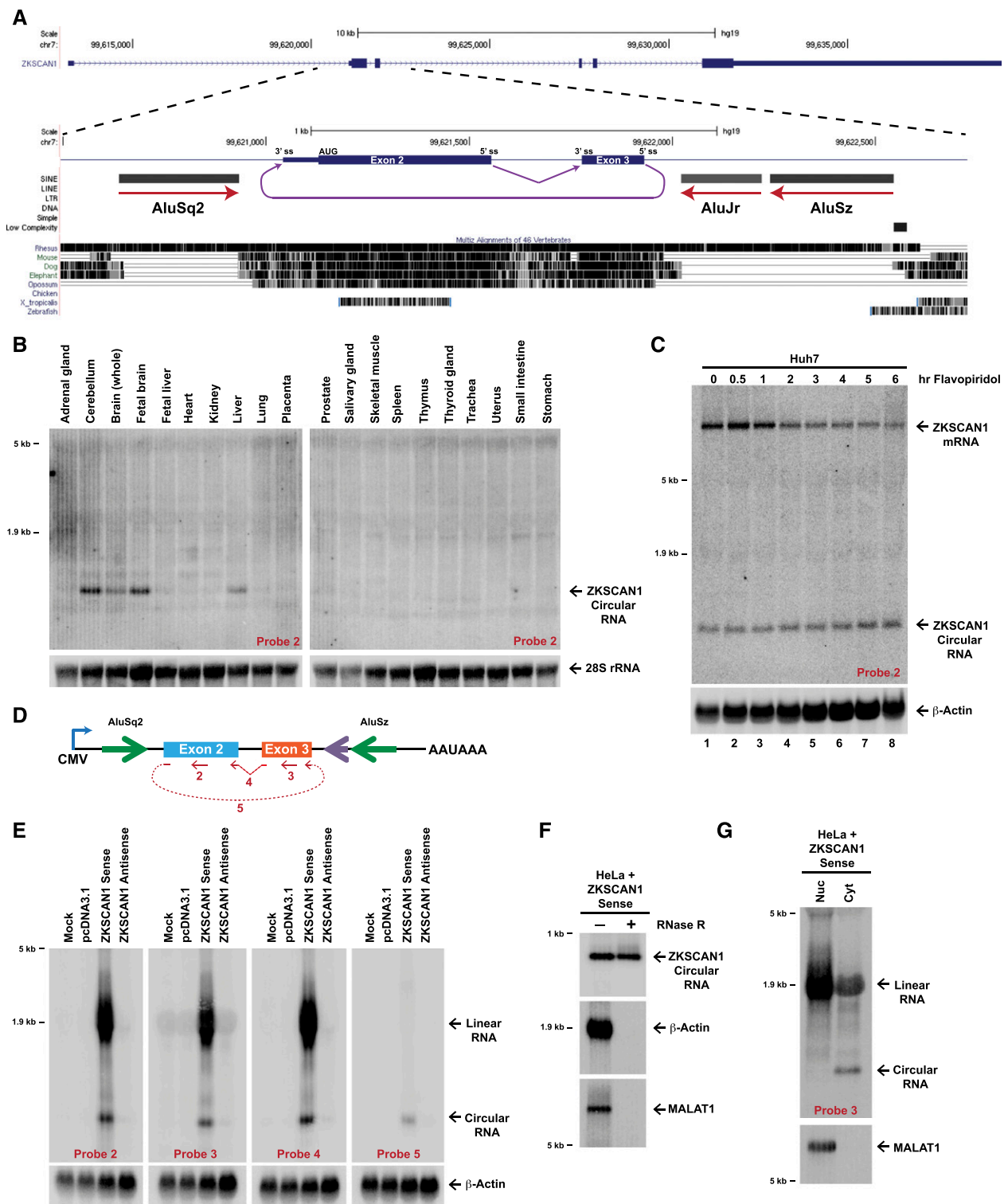
Figure 1. The human ZKSCAN1 gene generates a circular RNA. (*A*) Exon/intron structure of the human ZKSCAN1 locus, highlighting a 2232-nt region that includes exons 2 and 3. A circular RNA is formed when the 5′ splice site at the end of exon 3 is joined to the 3′ splice site at the beginning of exon 2 (purple). Repetitive elements in the designated orientations and evolutionary conservation patterns are shown. (*B*) Ten micrograms of total RNA from 20 normal human tissues was probed for ZKSCAN1 circular RNA expression. 28S ribosomal RNA was used as a loading control. (*C*) Total RNA from Huh7 cells treated with flavopiridol, a transcriptional elongation inhibitor, was subjected to Northern blot analysis. (*D*) The 2232-nt region of the ZKSCAN1 pre-mRNA was cloned into pcDNA3.1(+). The regions targeted by Northern oligonucleotide probes are denoted in red. (*E*) Plasmids containing the ZKSCAN1 region in the sense or antisense orientations were transfected into HeLa cells, and Northern blots were performed. β-Actin was used as a loading control. (*F*) The ZKSCAN1 circular RNA is resistant to RNase R digestion. (*G*) Transfected HeLa cells were fractionated to isolate nuclear and cytoplasmic total RNA, which was then subjected to Northern blot analysis with a probe to ZKSCAN1 exon 3. A probe to endogenous MALAT1 was used as a control for fractionation efficiency.

RNA expression levels observed in these tissues are not simply by-products of increased transcription.

Generation of an expression plasmid that accurately produces the ZKSCAN1 circular RNA

Repeat sequences are commonplace in introns, and, indeed, there are 15 short interspersed elements (SINEs) in the intron upstream of ZKSCAN1 exon 2, and 13 downstream from exon 3. Among these repeats, there are three primate-specific Alu elements in an inverted orientation that immediately flank the ZKSCAN1 circularizing exons (Fig. 1A). The upstream AluSq2 element is highly complementary to the downstream AluSz element (82% identity over 287 nt) (Supplemental Fig. 1C) but not to the downstream AluJr element, which is a member of a separate subfamily of Alu elements. As these inverted Alu elements are located close to the exons that circularize, we hypothesized that relatively short introns may be sufficient to support circularization. A 2232-nt region of the ZKSCAN1 pre-mRNA, spanning from 547 nt upstream of exon 2 to 796 nt downstream from exon 3, was thus cloned into the pcDNA3.1(+) expression vector (Fig. 1D). It should be noted that this vector includes only a portion of the surrounding introns and does not include either of the surrounding exons (exons 1 or 4).

Expression vectors containing the 2232-nt insert in the sense or antisense orientation were transiently transfected into HeLa cells, and total RNA was isolated 24 h later. Circular RNA expression from the ZKSCAN1 sense plasmid was detected using Northern blots and four different oligonucleotide probes, including one (probe 5) complementary to the junction between the end of exon 3 and the beginning of exon 2 (Fig. 1D,E). Confirming that a true circular RNA was produced, the vector-derived transcript was resistant to digestion by the 3′–5′ exonuclease RNase R, unlike linear β-actin mRNA or the long noncoding RNA MALAT1 (Fig. 1F). In addition, RT–PCR followed by Sanger sequencing confirmed that the 5′ splice site of exon 3 was properly joined to the 3′ splice site of exon 2 (Supplemental Fig. 1D). Finally, like the handful of other characterized circular RNAs (Hansen et al. 2013; Jeck et al. 2013; Memczak et al. 2013), the ZKSCAN1 circular RNA generated from our expression plasmid localized exclusively to the cytoplasm (Fig. 1G). We thus conclude that our plasmid generates bona fide ZKSCAN1 circular RNA. Circularization was additionally observed when the plasmid was transfected into mouse EpH4 cells (Supplemental Fig. 1E), indicating that the factors required to produce the ZKSCAN1 circular RNA are conserved between humans and mice.
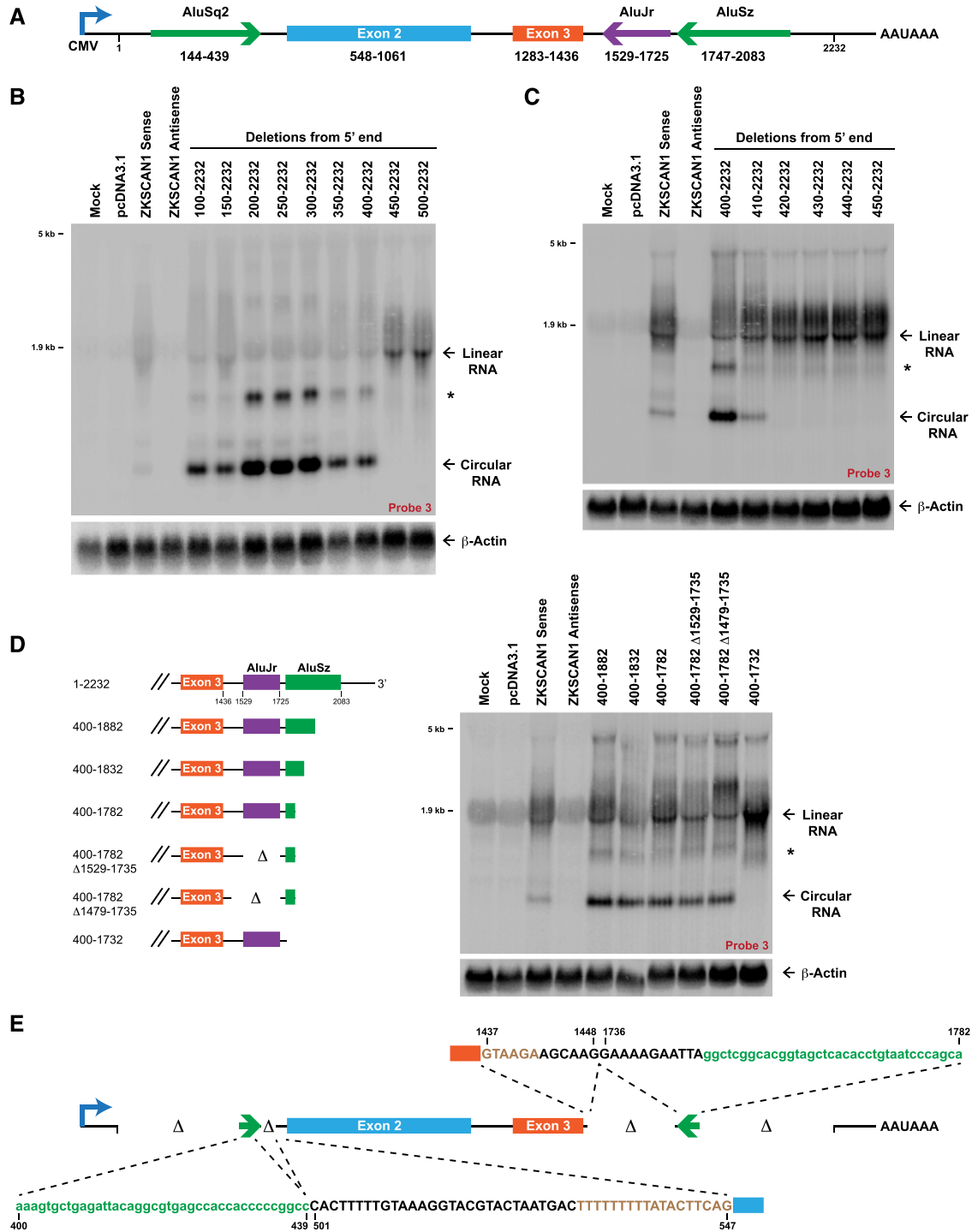
Short repeat sequences are sufficient for ZKSCAN1 circular RNA production

Having shown that the intronic regions immediately surrounding exons 2 and 3 are sufficient for ZKSCAN1 circular RNA production, we next aimed to identify the minimal sequence elements. Focusing first on the upstream intron, we progressively deleted nucleotides from the 5′ end, expecting to lose circular RNA production

once a portion of the AluSq2 element (nucleotides 144–439) was removed (Fig. 2A,B). Surprisingly, deleting the first 100 nt of this intron resulted in a more than eightfold increase in ZKSCAN1 circular RNA levels and the production of almost no linear RNA, suggesting the presence of an element that strongly inhibits circularization (100–2232 construct) (Fig. 2B). Cotransfection of a GFP expression plasmid confirmed equal transfection efficiency across samples (Supplemental Fig. 2A). Additional deletion constructs revealed that a cryptic 5′ splice site, located between nucleotides 14 and 23, was likely responsible for the inhibition (Supplemental Fig. 2B–D). Mutations that maintain the strength of the cryptic 5′ splice site had no effect, whereas single point mutations that disrupt the site were sufficient to completely relieve the repression on circularization (Supplemental Fig. 2E). This suggests that the intronic cryptic splice site may compete with the true splice sites for recognition by the spliceosome.

Upon continuing to delete from the 5′ end, we determined that the first 400 nt were dispensable for ZKSCAN1 circular RNA production (Fig. 2B). Only 40 nt of the AluSq2 element (nucleotides 400–439) remain in the 400–2232 construct, indicating that a short repeat is able to support circularization as efficiently as the full-length element (compare 400–2232 with 100–2232) (Fig. 2B). Circularization was only weakly observed upon deleting 410 nt and was completely eliminated once 420 nt or greater were removed (Fig. 2C). We therefore conclude that a small (∼30- 40-nt) region of the AluSq2 element is sufficient for ZKSCAN1 circular RNA production. Even higher circular RNA expression levels were obtained with constructs containing moderate-sized repeats (200–2232, 250–2232, and 300–2232), although an additional transcript of ∼1.3 kb was observed (Fig. 2B). This RNA was resistant to digestion by RNase R and contains only exons 2 and 3, as determined by Northern blots and RT–PCR (data not shown). This transcript may be a concatenated circular RNA corresponding to exons 2-3-2-3 that is generated by trans-splicing between two different primary transcripts or readthrough of the poly(A) signal (analogous to rolling circle replication). Alternatively, it may represent two independent 668-nt circles that are intertwined.

To then identify the minimal sequence elements present in the downstream intron, we generated an analogous set of constructs containing deletions from the 3′ end (Fig. 2D; Supplemental Fig. 3B). Circularization was not observed when the complete AluSz element was deleted even if the AluJr element was present (400–1732 construct) (Fig. 2D). The AluJr element was, in fact, dispensable for circularization, whereas a 36-nt region of the AluSz element was sufficient to support ZKSCAN1 circular RNA production (400–1782 Δ1479–1735 construct) (Fig. 2D). Interestingly, the location of the minimal AluSz element in the downstream intron does not appear to be critical, as moving it >250 nt from its endogenous location (400–1782 construct) to immediately next to the 5′ splice site (400–1782 Δ1479–1735 construct) had no effect on circularization efficiency. Several additional deletion constructs (Supplemental Fig. 3C,D) allowed

Figure 2. Short repeat sequences are sufficient to support ZKSCAN1 circularization. (*A*) Numbering scheme for the ZKSCAN1 expression plasmid. (*B,C*) ZKSCAN1 plasmids containing deletions at their 5′ ends were transfected into HeLa cells, and Northern blots were performed. The asterisk indicates an additional circular RNA species (see the text). (*D*) Using ZKSCAN1 expression plasmids starting at nucleotide 400, deletions were similarly introduced into the downstream intron. (*E*) Sequences of the minimal upstream and downstream introns that support circularization (400–1782 Δ440–500 Δ1449–1735 plasmid) are shown at the *bottom* and *top*, respectively. Repeat sequences (green) and splice sites (brown) are highlighted.

the identification of a final set of minimal introns that are sufficient to support efficient ZKSCAN1 circularization in cells (Fig. 2E). For the upstream intron (originally 547 nt),

87 nt are sufficient, which include 40 nt of the AluSq2 element as well as the 3′ splice site and branch point sequences. For the downstream intron (originally 796 nt),

59 nt are sufficient, which include the 5′ splice site and 36 nt of the AluSz element. Consistent with a role for the splicing machinery in ZKSCAN1 circularization, mutating the 5′ splice site at the end of exon 3 eliminated ZKSCAN1 circular RNA production (Supplemental Fig. 3E). From these results, it is clear that <40 nt of intronic inverted repeats are sufficient to trigger spliceosome-mediated circularization at the ZKSCAN1 locus. Therefore, the repeat length requirements for ZKSCAN1 circularization are likely much simpler than those previously suggested to be required at the *Sry* locus (Dubin et al. 1995).

Base-pairing between the repeat elements is necessary for ZKSCAN1 circularization

As the minimal sufficient regions of the upstream AluSq2 and downstream AluSz elements are highly complementary, these sequences may base-pair and bring the splice sites into close proximity to one another (Fig. 3A). Consistent with this model, circularization was not observed when three of the base pairs between the repeats were disrupted (Mut 5′ SINE and Mut 3′ SINE) (Fig. 3A,B). When base-pairing was re-established by introduction of mutations in both repeats, a significant rescue in the level of the ZKSCAN1 circular RNA was detected (Mut 5′ + 3′ SINE) (Fig. 3A,B). This indicates that base-pairing between AluSq2 and AluSz is necessary for circularization. Although the repeats are highly complementary, there are seven mismatches over the minimal 36-nt region (Fig. 3A). Based on simple thermodynamics, we hypothesized that strengthening the hairpin between the repeats may increase the efficiency of circularization. However, two different perfectly complementary hairpins [Perfect (Mut 5′) and Perfect (Mut 3′)] (Fig. 3A) produced the circular RNA at levels similar to that obtained with the wild-type vector (Fig. 3C). This suggests that strengthening the hairpin over a certain thermodynamic threshold does not increase circularization efficiency. To test this threshold model further, the minimal repeats were converted to their reverse complement (Rev Comp) sequences (Fig. 3A). The Rev Comp hairpin is predicted to be more stable than the wild-type structure; surprisingly, almost no circular RNA expression was detected (Fig. 3C). We thus conclude that not every pair of inverted repeat elements is able to support circularization.

It is possible that a sequence-specific element critical for circularization may have been lost when the hairpin was converted to its Rev Comp. Arguing against this model, ZKSCAN1 circular RNA production was maintained when the perfect repeat sequences were scrambled (while maintaining base-pairing) (Supplemental Fig. 4). Furthermore, although nucleotides 400–439 of AluSq2 are sufficient for circularization, this minimal region is not required for circularization if other portions of the AluSq2/AluSz elements are present (Fig. 3D). The minimal sufficient repeats can, in fact, be replaced with unrelated 40-nt regions from the AluSq2/AluSz elements (Figs. 3E,F). However, analogous to above, not all 40-nt regions can support circularization (360–1826) (Fig. 3F), even when they are predicted to form hairpins stronger than other sequences that do support circularization.

Upon searching for features that distinguish between hairpins that do and do not support circularization, we noticed that both the Rev Comp (Fig. 3A) and 360–1826 (Fig. 3E) hairpins contain an increased number of G-U wobble base pairs. Replacing all of the wobble pairs in the Rev Comp hairpin with canonical Watson-Crick base pairs was sufficient to allow efficient circular RNA production (Supplemental Fig. 5, Rev Comp.5 construct). Interestingly, converting a single wobble pair to a canonical U-A base pair (Supplemental Fig. 5, Rev Comp.1 and Rev Comp.3) or even a U-C mismatch (Supplemental Fig. 5, Rev Comp.9) was also sufficient, suggesting that subtle distortions in the hairpin specifically caused by G-U wobbles may significantly alter circularization efficiency. We thus conclude that more than simple thermodynamics is at play and that circularization is highly sensitive to how the repeat sequences base-pair to each other.
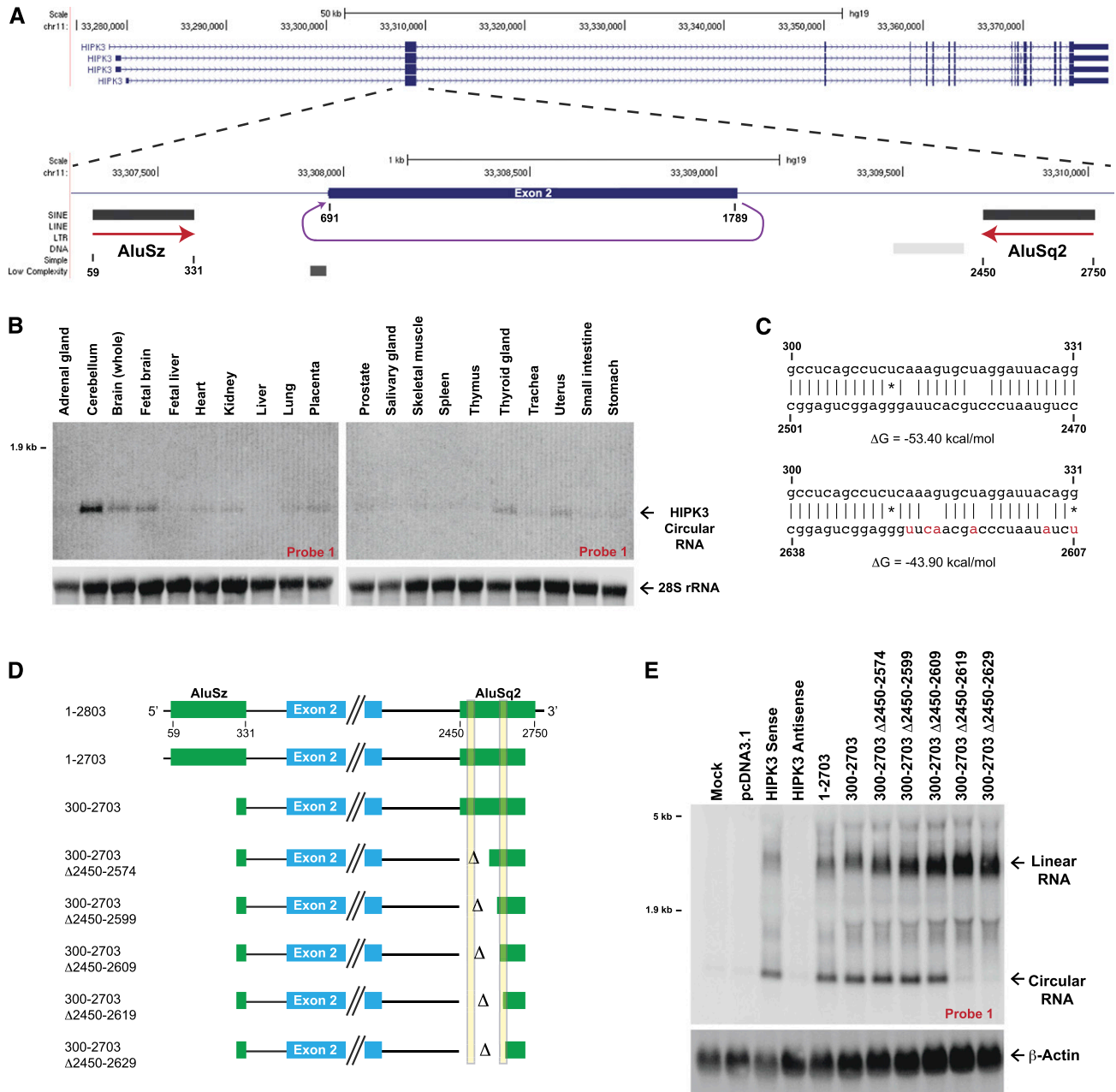
Short repeats are sufficient for production of the HIPK3 circular RNA

We next aimed to determine whether the rules for ZKSCAN1 circular RNA biogenesis can be applied more generally to explain how other circular RNAs are produced. The human HIPK3 gene has been implicated in multidrug resistance of cancer cells and encodes a serine–threonine kinase that regulates transcription and apoptosis (Begley et al. 1997; Curtin and Cotter 2004). Deep sequencing of several human cell lines previously suggested that HIPK3 also yields an abundant 1098-nt circular RNA comprised solely of exon 2 (Fig. 4A; Salzman et al. 2012; Jeck et al. 2013; Memczak et al. 2013). Using Northern blots and a probe complementary to the middle of exon 2, we determined that the HIPK3 circular RNA is most abundantly expressed in the brain, particularly in the cerebellum, but is also detectable in several other tissues (Fig. 4B). Like the ZKSCAN1 locus, highly complementary Alu repeats in an inverted orientation flank HIPK3 exon 2 (Fig. 4A; Supplemental Fig. 6A), suggesting that the biogenesis mechanisms for both RNAs may be similar.

We therefore cloned a 2803-nt region of the HIPK3 pre-mRNA into pcDNA3.1(+) and confirmed that this expression vector efficiently generates a circular RNA when transfected into HeLa cells (Supplemental Fig. 6B,C). Consistent with the results from the ZKSCAN1 locus (Fig. 2), a 32-nt region of the upstream AluSz element (nucleotides 300–331) was sufficient for HIPK3 circular RNA production (Figs. 4D,E; Supplemental Fig. 6D). This region of AluSz is highly complementary to nucleotides 2470–2501 of the downstream AluSq2 element (Fig. 4C). Surprisingly, a HIPK3 plasmid ending at nucleotide 2503, thus containing this putative AluSq2-pairing site, failed to circularize (Supplemental Fig. 6E). In fact, deletions from the 3′ end of the plasmid suggested that >200-nt of the downstream AluSq2 element may be required for circularization (Supplemental Fig, 6E). However, this apparent discrepancy between minimal repeat lengths in the two introns was resolved upon the realization that the minimal AluSz region is also highly complementary
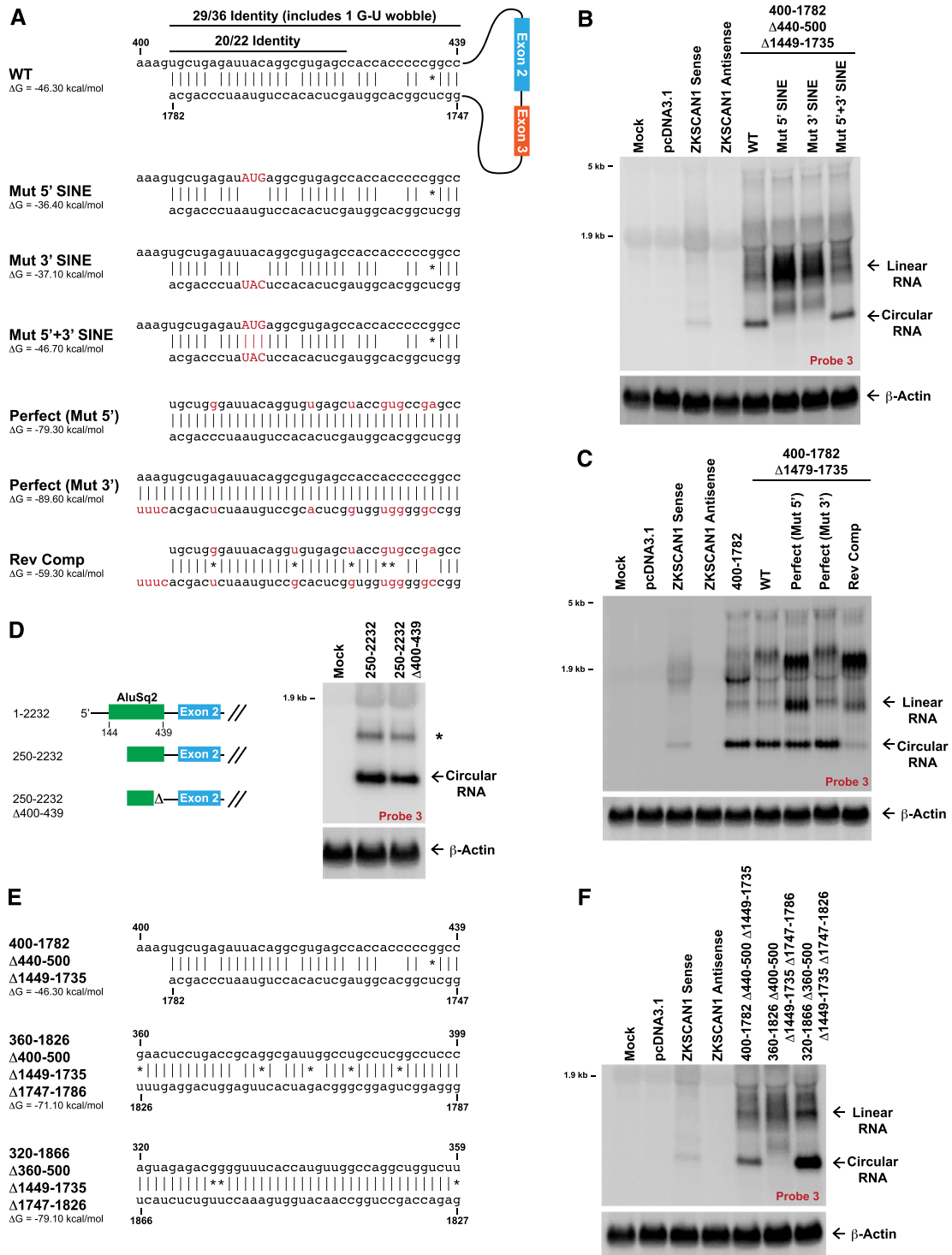
Figure 3. Only specific short repeat sequences are able to support circular RNA production. (*A*) Mutations (denoted in red) in the minimal sufficient repeat sequences were introduced into the ZKSCAN1 expression plasmid. mFold was used to calculate hairpin stabilities, assuming a 7-nt linker (AGAAUUA) between the two repeat sequences. (*B,C*) ZKSCAN1 plasmids containing wild-type (WT) or mutant repeats were transfected into HeLa cells, and Northern blots were performed. The minimal sufficient introns (as depicted in Fig. 2E) were used in *B*, while slightly longer introns were used in *C*. (*D*) Although sufficient for circular RNA production, the minimally sufficient AluSq2 region (nucleotides 400–439) was not required for circularization. (*E*) The minimal AluSq2/AluSz repeats (nucleotides 400–439 and 1747–1782) were replaced with other 40-nt regions of the Alu elements. The remainder of the plasmid was unchanged, allowing the effect of altering only the repeat sequences to be measured. mFold was used to calculate hairpin stabilities as above. (*F*) Northern blots revealed that the thermodynamic stability of the hairpins is not an adequate predictor of circularization efficiency.

Figure 4. Specific short repeats are sufficient for HIPK3 circularization. (*A*) Exon/intron structure of the human HIPK3 locus, highlighting a 2803-nt region that includes exon 2. A circular RNA is formed when the 5′ splice site at the end of exon 2 is joined to the 3′ splice site at the beginning of exon 2 (purple). Repetitive elements in the designated orientations are shown. (*B*) Ten micrograms of total RNA from 20 normal human tissues was probed for HIPK3 circular RNA expression. 28S ribosomal RNA was used as a loading control. (*C*) Nucleotides 300–331 of the upstream AluSz element are highly complementary to two different regions of the downstream AluSq2 element. Sequence differences between the two downstream elements are shown in red. (*D*) Deletions were introduced into the HIPK3 expression plasmid. The two AluSq2 complementary regions are highlighted in yellow. (*E*) HIPK3 plasmids were transfected into HeLa cells, and Northern blots were performed. Deleting portions of the 2607–2638 complementary region eliminated circular RNA production.

to nucleotides 2607–2638 of the downstream AluSq2 element (Fig. 4C). Whereas nucleotides 2470–2501 were dispensable for HIPK3 circularization, no circular RNA was produced when the 2607–2638 region was disrupted; e.g., by removing 13 nt to generate the 300–2703 Δ2450–2619 plasmid (Figs. 4D,E). We therefore conclude that

base-pairing between nucleotides 300–331 of the upstream AluSz element and nucleotides 2607–2638 of the downstream AluSq2 element is sufficient to support HIPK3 circularization. Interestingly, the hairpin formed between the AluSz element and nucleotides 2607–2638 is predicted to be less stable than that formed with nucle-

otides 2470–2501 (Fig. 4C), yet it is only the 2607–2638 region that supports circularization. In total, these results suggest that circularization at the HIPK3 and ZKSCAN1 loci have similarly short intronic repeat requirements.

Sequences within longer repeats can inhibit circularization

Our results from the ZKSCAN1 and HIPK3 loci clearly suggest that some, but not all, short repeats are able to support circular RNA production. As longer repeats contain more sequence and thus more potential for base-pairing, it seems likely that longer repeats might more uniformly support circularization. The EPHB4 locus, however, revealed that a longer repeat sequence can sometimes strongly inhibit circularization. EPHB4 encodes a receptor tyrosine kinase that is commonly overexpressed in human cancers (for review, see Noren and Pasquale 2007). In addition, a 360-nt circular RNA is generated from exons 11 and 12 (Jeck et al. 2013; Memczak et al. 2013; Salzman et al. 2013). Inverted Alu repeats flank these exons (Fig. 5A; Supplemental Fig. 7A), and we determined that a 1428-nt region of the EPHB4 pre-mRNA was sufficient for EPHB4 circular RNA production (Supplemental Fig. 7B–D). Initial deletions from the 5′ end of the EPHB4 expression plasmid suggested that >100 nt of the AluSx1 element may be required for circularization (Figs. 5B,C). However, analogous to the HIPK3 locus, only the middle of the AluSx1 element was critical for circularization, and a minimal sufficient region of <50 nt (nucleotides 200–250) was identified (200–1428 Δ251–343 construct) (Fig. 5C).

Surprisingly, the 200–1428 EPHB4 expression plasmid, which contains 144 nt of the AluSx1 element, produced approximately fivefold less circular RNA than the 200–1428 Δ301–343 plasmid, which contains only 100 nt of the repeat (Fig. 5C). Similarly, the 250–1428 EPHB4 plasmid failed to generate the circular RNA but did upon deletion of nucleotides 301–343 (Fig. 5D). These results indicate that nucleotides 301–343 of the repeat sequence can inhibit circularization. A poly(A) tract is present from nucleotide 321 to 343, and deletion of this region was sufficient to allow the 250–1428 plasmid to produce the circular RNA (250–1428 Δ321–343 construct) (Fig. 5D). Binding of poly(A)-binding protein (PABP) to this region may sterically interfere with the ability of the hairpin to form. Alternatively, as a poly(A) tract is a low-complexity sequence, this portion of the repeat may bind more promiscuously to other transcripts in *trans*, thereby inhibiting circularization in *cis*.

## Pre-mRNA 3′ end formation is critical for circularization

As a circular RNA is covalently closed, the mature transcript contains no 5′ cap or 3′ poly(A) tail. Nevertheless, we found that disrupting the downstream polyadenylation signal in our expression plasmids (by placing it in the antisense orientation) caused the ZKSCAN1 circular RNA to no longer be produced (Fig. 6B). This suggested a potential role for 3′ end formation in circular

RNA biogenesis. ZKSCAN1 terminating in the noncanonical MALAT1 3′ end processing mechanism, which involves recognition and cleavage of a tRNA-like structure by the endonuclease RNase P (Wilusz et al. 2008), was able to circularize (Fig. 6A,B). Therefore, circularization is dependent not on polyadenylation but instead on the act of forming a 3′ end. After RNase P cleavage, the 3′ end of MALAT1 is stabilized by a triple helical structure (Wilusz et al. 2012; Brown et al. 2014). We previously showed that when the triple helix is mutated, RNase P cleavage occurs, but the mature RNA is subsequently degraded (Wilusz et al. 2012), thereby uncoupling 3′ end cleavage from RNA stabilization. ZKSCAN1 pre-mRNA ending in the mutant triple helix produced a circular RNA, although at lower levels than the wild-type triple helix (Fig. 6B). We thus conclude that for a circular RNA to be produced from our expression plasmids, the pre-mRNA must be cleaved at its 3′ end. When the linear transcript has a stable 3′ end, greater levels of the circular RNA are produced. This suggests that circular RNA formation may at least in part occur post-transcriptionally.

Circularization requires the intronic repeats to collaborate with the exons

Having defined roles for intronic repeats and 3′ end formation signals in circular RNA biogenesis, we next addressed whether the exonic sequences play a role in circularization. A 51-nt artificial exon composed of restriction enzyme sites was inserted between the ZKSCAN1 minimal sufficient introns (as defined in Fig. 2E), thereby generating the "CircRNA Mini Vector" (Fig. 7A). No circular RNA was produced when this plasmid was transfected into HeLa cells (data not shown), indicating that additional exonic elements are required. Inserting most of exons 2 and 3 from ZKSCAN1 (nucleotides 548–1417) between the EcoRV and SacII sites restored efficient circular RNA production (Figs. 7A,B). As the internal intron between exons 2 and 3 (nucleotides 1062–1282) was not included in this vector, it is clear that the ZKSCAN1 minimal introns can support the production of circular RNAs comprised of a single or multiple exons. Inserting smaller ZKSCAN1 exonic regions into the CircRNA Mini Vector revealed that exon 3 (nucleotides 1283–1436) was dispensable for circularization (548–1047 construct) (Fig. 7B). Further decreasing the length of the inserted exon, however, eliminated circular RNA production. Whereas ZKSCAN1 exons ≥500 nt in length circularized uniformly well, ZKSCAN1 exons ≤400 nt circularized poorly or not at all (Fig. 7B; Supplemental Fig. 8). This suggests a correlation between exon length and circularization efficiency, a result that is consistent with previous computational analysis that indicated that longer than average exons are generally present in circular RNAs (Jeck et al. 2013).

We next determined whether the ZKSCAN1 minimal introns are able to support the circularization of exons derived from other genes. The ZKSCAN1 introns supported efficient production of the 360-nt EPHB4 exon 11/12 circular RNA (Fig. 7C), indicating that some smaller exons are able to circularize. Surrounding GAPDH exon
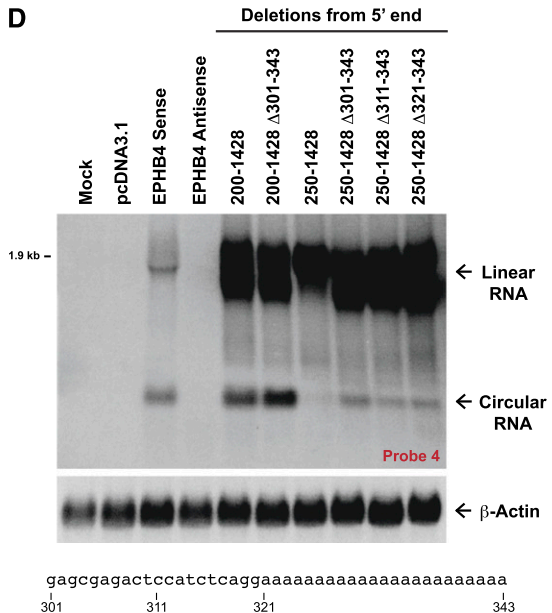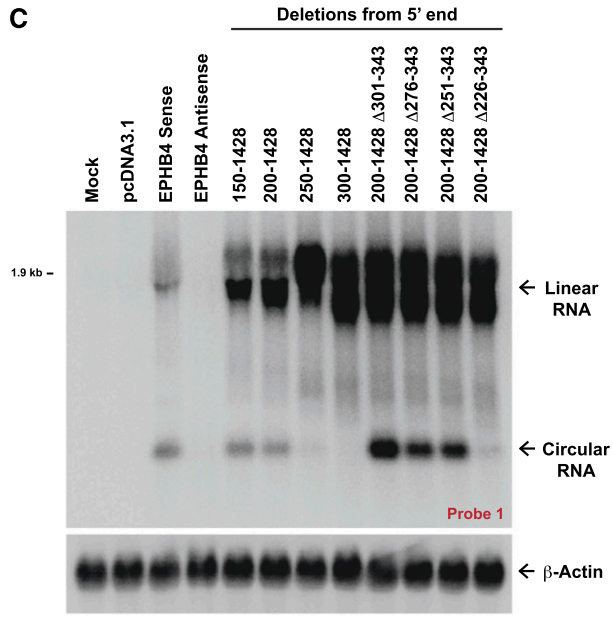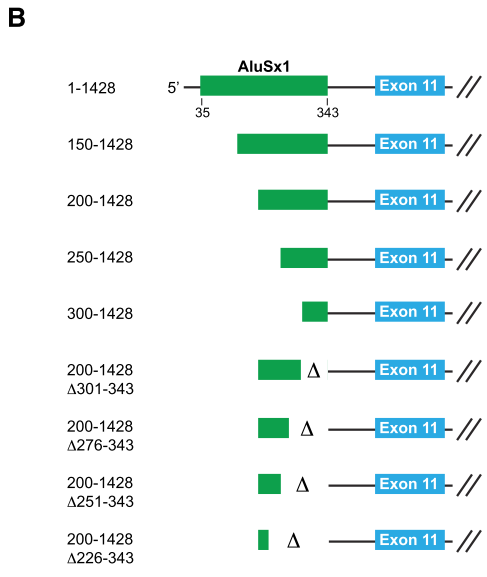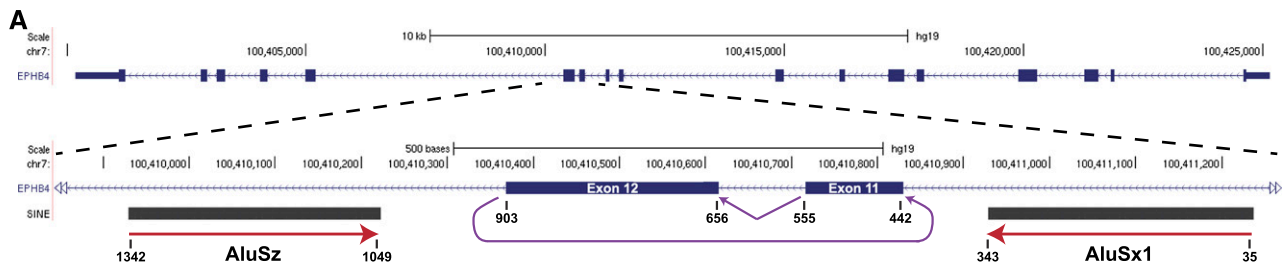
Figure 5. EPHB4 circularization is inhibited by a portion of the flanking Alu repeats. (*A*) Exon/intron structure of the human EPHB4 locus, highlighting a 1428-nt region that includes exons 11 and 12. A circular RNA is formed when the 5′ splice site at the end of exon 12 is joined to the 3′ splice site at the beginning of exon 11 (purple). Repetitive elements in the designated orientations are shown. (*B*) Schematics of EPHB4 expression plasmids. Exon 12 and the downstream intron are not shown for simplicity. (*C*) EPHB4 plasmids were transfected into HeLa cells, and Northern blots were performed. Probe-binding sites are shown in Supplemental Figure 7C. The 250–1428 plasmid contains 94 nt of the AluSx1 element, but no circularization was observed. In contrast, the 200–1428 Δ251–343 plasmid contains only 50 nt of the AluSx1 element and efficiently generated the EPHB4 circular RNA. (*D*) The poly(A) tract at the 3′ end of the AluSx1 element (nucleotides 321–343) inhibits circularization.
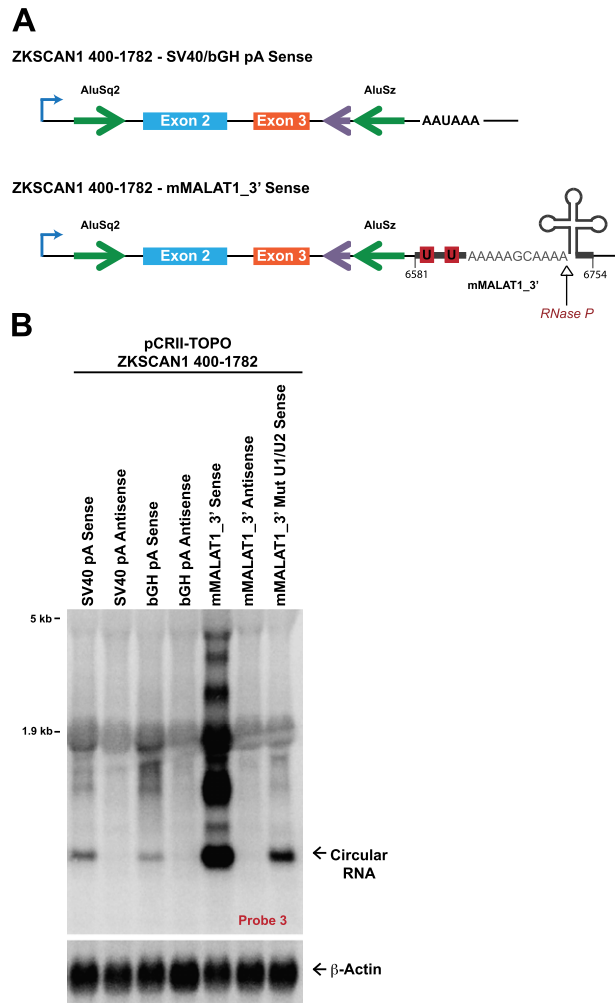
Figure 6. 3′ end formation is required for circularization. (*A*) Schematics of ZKSCAN1 expression plasmids. The complete polyadenylation (pA) signals, which include the AAUAAA sequence, were placed in the antisense orientation or replaced by nucleotides 6581–6754 of mouse MALAT1. This 174-nt region, denoted as mMALAT1_3′, includes a tRNA-like structure that is recognized and cleaved by RNase P (Wilusz et al. 2008). U-rich (denoted in red) and A-rich motifs then form a triple helical structure, thereby protecting the mature 3′ end from exonucleolytic degradation (Wilusz et al. 2012). To generate the mMALAT1_3′ Mut U1/U2 plasmid, U → A mutations were introduced into both U-rich motifs, thereby disrupting the triple helix. (*B*) HeLa cells were transfected with pCRII-TOPO plasmids expressing ZKSCAN1 followed by differing 3′-terminal sequences. Northern blots were then performed using probe 3 from Figure 1D.

7, which does not naturally circularize, with the ZKSCAN1 minimal introns similarly resulted in circular RNA production (Fig. 7D). However, circularization was not observed when exon 2 of HIPK3, which is 1098 nt, was inserted (Fig. 7E). Short intronic repeats in their endogenous context were sufficient to support HIKP3 circularization (Fig. 4E), although the introns used in that plasmid were not as minimalized as those used here. From these various exonic sequences tested, we conclude

that circularization clearly requires collaboration between the exons and their flanking introns.

Discussion

For many genes, pre-mRNA splicing can result in the production of a linear mRNA or a circular noncoding RNA (Fig. 7F). However, the mechanism by which the spliceosome selects only certain exons to circularize is largely unknown. In the present study, we demonstrated that short repeat elements in the flanking introns are critical for circularization of some intervening exons. The repeats must base-pair, which suggests that the splice sites must be brought into close proximity to one another to promote backsplicing. Although mechanistically simple, this step occurs in a highly selective manner, as the sequence of the repeats can drastically alter the efficiency of circular RNA production. Exons of longer length appear to circularize more efficiently, and 3′ end processing is required for circularization, suggesting that circular RNA production may occur post-transcriptionally. Our data thus provide new insights into how to produce a circular RNA as well as how the critical choice between linear versus circular RNA is regulated.

A proposed general model for how circular RNAs are produced

Although recent deep sequencing studies have clearly revealed that many genes produce circular RNAs, our current understanding of the circularization mechanism is rather limited and based largely on a single gene: mouse *Sry*. The *Sry* exon is flanked by consensus splice sites as well as >15-kb inverted repeats (Capel et al. 1993). This rather unusual genomic organization led Capel et al. (1993) to propose that the repeats form an intramolecular stem–loop structure, resulting in the positioning of the splice donor close to the splice acceptor. Consistent with this model, complete deletion of either of the inverted repeats eliminated *Sry* exon circularization in cells (Dubin et al. 1995). Although 400 complementary nucleotides were proposed to be necessary to produce the *Sry* circular RNA (Dubin et al. 1995), smaller repeat sequences were not tested. Based on these results from the *Sry* locus, a ciRS-7 (also known as CDR1as) circular RNA expression plasmid was recently generated by artificially flanking the exon with ∼800 nt of perfectly complementary repeats (Hansen et al. 2013). However, such long repeats do not naturally flank the ciRS-7 exon or the large majority of other exons. Smaller inverted repeats, particularly Alu elements that are ∼300 nt in length, are statistically enriched in introns flanking exons that circularize (Jeck et al. 2013), although these repeats have yet to be experimentally shown to facilitate the production of circular RNAs.

Based on our results from the ZKSCAN1, HIPK3, and EPHB4 genes, we built on these prior studies to propose the following model for how the spliceosome determines whether to produce a linear mRNA or a circular noncoding RNA (Fig. 7F). Although the most common outcome of
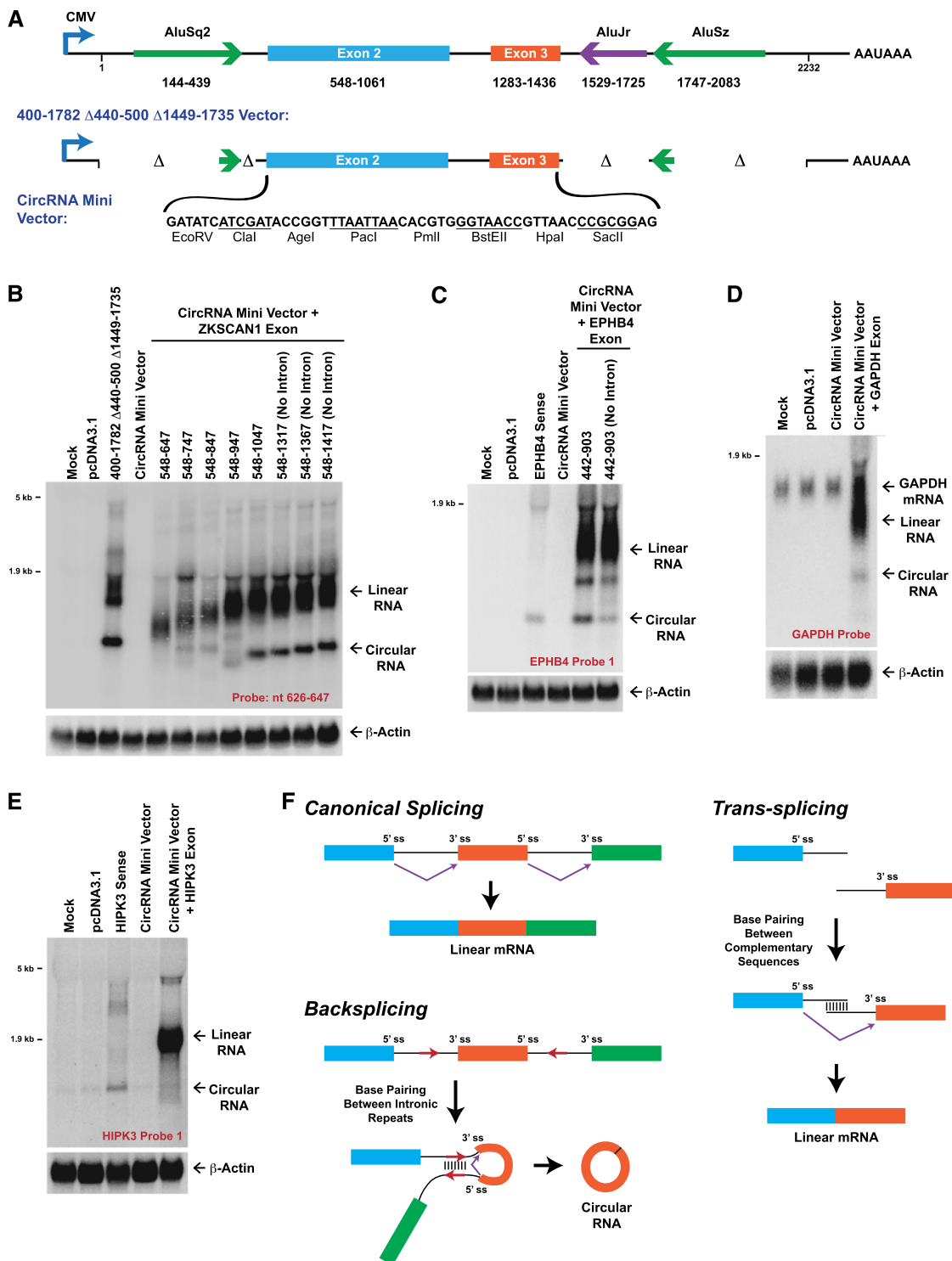
Figure 7. Collaboration between the exon and introns allows circularization. (*A*) To facilitate the identification of exon sequences that can be circularized, exons 2 and 3 of the minimal ZKSCAN1 400–1782 Δ440–500 Δ1449–1735 expression plasmid were replaced with an artificial 51-nt exon composed of restriction enzyme sites. (*B*) Segments of the ZKSCAN1 exons were then inserted between the EcoRV and SacII sites. The numbering scheme is as in *A* and Figure 2A. The intron between exons 2 and 3 (nucleotides 1062–1282) was not included in the vectors designated "No Intron." HeLa cells were then transfected, and Northern blots were performed. (*C*) Exons 11 and 12 of EPHB4 with or without the internal intron were inserted into the CircRNA Mini Vector. The numbering scheme is as per Figure 5A. (*D*) Exon 7 of GAPDH, which is 336 nt, was inserted into the CircRNA Mini Vector. In addition to the circular RNA, the endogenous GAPDH transcript was detected on the Northern blot. (*E*) The minimal ZKSCAN1 introns were unable to support circularization of HIKP3 exon 2. (*F*) Comparison of different pre-mRNA splicing mechanisms. Whereas canonical splicing produces a linear mRNA, base-pairing between intronic repeat sequences can trigger backsplicing and the formation of a circular RNA. When base-pairing occurs between complementary sequences on two independent primary transcripts, *trans*-splicing can generate a chimeric linear mRNA.

splicing appears to be linear mRNA production, circularization is initiated when repeats in the introns immediately flanking the exons base-pair to one another (Fig. 3). As was proposed at the *Sry* locus, this brings the intervening splice sites into close proximity, facilitating catalysis. Short inverted repeats (~30–40 nt) are sufficient, although more than simple thermodynamics regulates this critical event. First, the sequence content of the repeats can significantly alter circularization efficiency (Figs. 3–5). Low-complexity sequences [such as poly(A) tracts] or subtle distortions in the hairpin between the repeats (e.g., due to G-U wobble pairs) can inhibit circularization. Additional RNA secondary structures and RNA-binding proteins likely also play significant roles in determining whether a given repeat is accessible. Second, given that 3′ end processing was required for circularization from our plasmids (Fig. 6), the process probably occurs post-transcriptionally. Backsplicing thus appears to be a relatively slow process, and the flanking introns likely must be spliced slower than average. If there is not sufficient time for the repeats to base-pair and/or recruit the spliceosome, a linear mRNA is produced. Third, the intronic repeats must collaborate with the exonic sequences (Fig. 7). In general, longer exons appear to circularize more efficiently than shorter exons, possibly because their increased length provides flexibility to the transcript and allows the intronic repeats to base-pair more easily. Due to these multiple regulatory steps, circularization is thus restricted to only specific exons despite the fact that repeat sequences comprise >45% of the human genome and are commonplace in introns.

Repeats play a critical role in shaping exon circularization patterns

Although Alu elements flank the ZKSCAN1, HIPK3, and EPHB4 exons that circularize, other classes of repeat sequences are likely able to support circularization. For example, we found that inverted DNA transposon repeats are sufficient to support circularization at the human GCN1L1 locus (Supplemental Fig. 9). Furthermore, considering that the ZKSCAN1 circular RNA was still produced after scrambling the flanking repeat sequences (Supplemental Fig. 4), it is highly likely that complementary sequences that are not annotated repeats may be able to support circularization. However, it should be noted that the simple presence of a pair of inverted repeats does not mean that a circular RNA will be produced at a similar efficiency in all cell types. For example, we observed that the ratio of the ZKSCAN1 mRNA versus the circular RNA present in various tissues varied significantly (Supplemental Fig. 1B). This may be because the transcripts have different half-lives across cell types or because intron pairing between the repeats efficiently occurs only under certain conditions. RNA-binding proteins, especially those that bind dsRNA, may modulate the accessibility of the repeat sequences to base-pair. For example, ADAR (adenosine deaminase acting on RNA) enzymes may generally inhibit circular RNA production, as they convert adenosines in double-stranded regions to inosines, thereby unwinding hairpins (for review, see Nishikura 2010). Supporting this model, A-to-I editing is most commonly observed in intronic repeats that are engaged in intramolecular base-pairing (Athanasiadis et al. 2004), exactly the sort of structure that is needed for circularization. Further studies are needed to understand how base-pairing between the introns may affect spliceosome assembly as well as the recruitment of well-characterized splicing regulatory factors, including hnRNP and SR proteins.

Interplay between 3′ end formation and circularization

In addition to the intronic inverted repeat sequences, a functional 3′ end processing signal was surprisingly found to be required for circularization from our expression plasmids, suggesting that circularization may occur post-transcriptionally (Fig. 6). Although most introns are thought to be spliced cotranscriptionally (for review, see Brugiolo et al. 2013), there is increasing evidence that some introns are not spliced until after polyadenylation occurs (Vargas et al. 2011; Bhatt et al. 2012). These incompletely spliced, polyadenylated transcripts accumulate in the chromatin faction and, after a significant lag period, are eventually spliced to produce functional mRNAs. Although some of these post-transcriptional splicing events are due to functionally impaired splicing signals, there is evidence that alternatively spliced exons can also be slowly spliced (Vargas et al. 2011). This suggests that regulatory events, such as the binding of alternative splicing regulatory proteins, can cause transcription and splicing to become uncoupled. We propose that the kinetics of splicing are likely slowed during the circularization process, possibly because pairing of the intronic repeats interferes with exon definition, thereby preventing the cotranscriptional formation of a fully spliced linear mRNA. By forming a stable 3′ end, the transcript is able to avoid degradation by nuclear RNA surveillance mechanisms (for review, see Schmidt and Butler 2013) and allow the spliceosome and other regulatory factors to determine whether to produce a linear mRNA or a circular RNA. Alternatively, as polyadenylation and splicing can be functionally coupled (Niwa and Berget 1991), the 3′ end processing machinery may recruit or otherwise directly interact with circular RNA biogenesis factors, thereby promoting circularization.

In summary, our findings provide key insights into the mechanism by which the spliceosome generates circular RNAs. Considering that thousands of circular RNAs are produced in eukaryotic cells, base-pairing between intronic repeats appears to be a widespread phenomenon by which the transcriptome is expanded. Interestingly, although the exonic sequences of protein-coding genes may be highly evolutionarily conserved, the genomic repeat landscape varies significantly among species (for review, see Shedlock and Okada 2000). Therefore, the production of a circular RNA from a given exon can be highly species-specific, and very different functional

RNAs, some of which are circular, may be produced from a given gene. Exon circularization thus likely represents an underappreciated way by which the coding capacity of eukaryotic genes is expanded and the ultimate functions of genes are modulated.

Materials and methods

Expression plasmid construction

To generate the ZKSCAN1, HIPK3, and EPHB4 expression plasmids, the indicated sequences were inserted into pcDNA3.1(+) (Life Technologies) as further described in the Supplemental Material. Genomic coordinates in human genome version 19 (hg19) for the full-length inserts were as follows: ZKSCAN1 (chr7: 99,620,495–99,622,726), HIPK3 (chr11: 33,307,269–33,310,071), and EPHB4 (chr7: 100,409,844–100,411,271). For the CircRNA Mini Vector depicted in Figure 7A, exonic sequences were inserted between the minimal ZKSCAN1 introns using the EcoRV and SacII sites. To generate ZKSCAN1 plasmids terminating in various 3′ end sequences (bGH polyadenylation signal, SV40 polyadenylation signal, or the mMALAT1_3′ region), the cGFP ORF in our previously described pCRII-TOPO CMV-cGFP expression plasmids (Wilusz et al. 2012) was replaced with the designated region of ZKSCAN1. Additional details and the sequences of the inserts for all plasmids are provided in the Supplemental Material. mFold was used to calculate hairpin stabilities, assuming a 7-nt linker (AGAAUUA) between the two repeat sequences.

Transfections and RNA analysis

HeLa and Huh7 cells were grown at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) containing high glucose (Life Technologies), supplemented with penicillin–streptomycin and 10% fetal bovine serum. One microgram of each expression plasmid was transfected using Lipofectamine 2000 (Life Technologies), and total RNA was isolated using Trizol (Life Technologies) as per the manufacturer's instructions. Nuclear and cytoplasmic fractionation was performed as previously described (Wilusz et al. 2008). To inhibit RNA polymerase II transcription, Huh7 cells were treated with flavopiridol (1 μM final concentration; Sigma) for 0–6 h at 37°C. For RNase R treatments, 10 μg of total RNA was treated with 20 U of RNase R (Epicentre) for 1 h at 37°C. Northern blots using oligonucleotide probes were performed as previously described (Wilusz et al. 2008). All oligonucleotide probe sequences are provided in Supplemental Table 1.

The Human Total RNA Master Panel II (Clontech) was used to examine the expression of endogenous circular RNAs in normal human tissues. To measure ZKSCAN1 circular RNA and mRNA levels, reverse transcription using random hexamers and qPCR was performed as previously described (Sunwoo et al. 2009). qPCR primer sequences are provided in Supplemental Table 1.

Acknowledgments

References

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

Begley DA, Berkenpas MB, Sampson KE, Abraham I. 1997. Identification and sequence of human PKY, a putative kinase with increased expression in multidrug-resistant cells, with homology to yeast protein kinase Yak1. *Gene* **200:** 35–43.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150:** 279–290.

Braun S, Domdey H, Wiebauer K. 1996. Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. *Nucleic Acids Res* **24:** 4152–4157.

Brown JA, Bulkley D, Wang J, Valenstein ML, Yario TA, Steitz TA, Steitz JA. 2014. Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat Struct Mol Biol* **21:** 633–640.

Brugiolo M, Herzel L, Neugebauer KM. 2013. Counting on cotranscriptional splicing. *F1000Prime Rep* **5:** 9.

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. 2010. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* **6:** e1001233.

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R. 1993. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73:** 1019–1030.

Cech TR, Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157:** 77–94.

Chen CY, Sarnow P. 1995. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268:** 415–417.

Cocquerelle C, Mascrez B, Hetuin D, Bailleul B. 1993. Missplicing yields circular RNA molecules. *FASEB J* **7:** 155–160.

Curtin JF, Cotter TG. 2004. JNK regulates HIPK3 expression and promotes resistance to Fas-mediated apoptosis in DU 145 prostate carcinoma cells. *J Biol Chem* **279:** 17090–17100.

Danan M, Schwartz S, Edelheit S, Sorek R. 2012. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res* **40:** 3131–3142.

Dubin RA, Kazmi MA, Ostrer H. 1995. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167:** 245–248.

Guo JU, Agarwal V, Guo H, Bartel DP. 2014. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* **15:** 409.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* **495:** 384–388.

Jeck WR, Sharpless NE. 2014. Detecting and characterizing circular RNAs. *Nat Biotechnol* **32:** 453–461.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. 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 SD, Gregersen LH, Munschauer M, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495:** 333–338.

Mulligan MK, Wang X, Adler AL, Mozhui K, Lu L, Williams RW. 2012. Complex control of GABA(A) receptor subunit mRNA expression: variation, covariation, and genetic regulation. *PLoS ONE* **7:** e34586.

Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B. 1991. Scrambled exons. *Cell* **64:** 607–613.

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

Niwa M, Berget SM. 1991. Mutation of the AAUAAA poly-adenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* **5:** 2086–2095.

Noren NK, Pasquale EB. 2007. Paradoxes of the EphB4 receptor in cancer. *Cancer Res* **67:** 3994–3997.

Pasman Z, Been MD, Garcia-Blanco MA. 1996. Exon circulari-zation in mammalian nuclear extracts. *RNA* **2:** 603–610.

Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7:** e30733.

Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. 2013. Cell-type specific features of circular RNA expression. *PLoS Genet* **9:** e1003777.

Schindewolf C, Braun S, Domdey H. 1996. In vitro generation of a circular exon from a linear pre-mRNA transcript. *Nucleic Acids Res* **24:** 1260–1266.

Schmidt K, Butler JS. 2013. Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. *Wiley Interdis-cip Rev RNA* **4:** 217–231.

Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. *BioEssays* **22:** 148–160.

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. 2009. MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* **19:** 347–359.

van Dekken H, Tilanus HW, Hop WC, Dinjens WN, Wink JC, Vissers KJ, van Marion R. 2009. Array comparative genomic hybridization, expression array, and protein analysis of critical regions on chromosome arms 1q, 7q, and 8p in adenocarcinomas of the gastroesophageal junction. *Cancer Genet Cytogenet* **189:** 37–42.

Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, Schedl P, Tyagi S. 2011. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147:** 1054–1065.

Wang PL, Bao Y, Yee MC, Barrett SP, Hogan GJ, Olsen MN, Dinneny JR, Brown PO, Salzman J. 2014. Circular RNA is expressed across the eukaryotic tree of life. *PLoS ONE* **9:** e90859.

Wilusz JE, Sharp PA. 2013. Molecular biology. A circuitous route to noncoding RNA. *Science* **340:** 440–441.

Wilusz JE, Freier SM, Spector DL. 2008. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135:** 919–932.

Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23:** 1494–1504.

Wilusz JE, Jnbaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, Sharp PA. 2012. A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* **26:** 2392–2407.

Yang L, Froberg JE, Lee JT. 2014. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci* **39:** 35–43.

Zaphiropoulos PG. 1997. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol Cell Biol* **17:** 2985–2993.