Molecular Cell
Article

Circular Intronic Long Noncoding RNAs

Yang Zhang,[1,3] Xiao-Ou Zhang,[2,3] Tian Chen,[1] Jian-Feng Xiang,[1] Qing-Fei Yin,[1] Yu-Hang Xing,[1] Shanshan Zhu,[2] Li Yang,[2,*] and Ling-Ling Chen[1,*]
[1]State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology
[2]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology
Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[3]These authors contributed equally to this work
*Correspondence: liyang@picb.ac.cn (L.Y.), linglingchen@sibcb.ac.cn (L.-L.C.)
http://dx.doi.org/10.1016/j.molcel.2013.08.017

SUMMARY

We describe the identification and characterization of circular intronic long noncoding RNAs in human cells, which accumulate owing to a failure in debranching. The formation of such circular intronic RNAs (ciRNAs) can be recapitulated using expression vectors, and their processing depends on a consensus motif containing a 7 nt GU-rich element near the 5′ splice site and an 11 nt C-rich element close to the branchpoint site. In addition, we show that ciRNAs are abundant in the nucleus and have little enrichment for microRNA target sites. Importantly, knockdown of ciRNAs led to the reduced expression of their parent genes. One abundant such RNA, *ci-ankrd52*, largely accumulates to its sites of transcription, associates with elongation Pol II machinery, and acts as a positive regulator of Pol II transcription. This study thus suggests a *cis*-regulatory role of noncoding intronic transcripts on their parent coding genes.

INTRODUCTION

The recent advent of high-throughput approaches has revealed that a large portion of the mammalian genome is transcribed into long noncoding RNAs (lncRNAs, >200 nt in length), including thousands of lincRNAs identified using a "H3K4me3-H3K36me3" chromatin signature of actively transcribed genes (Guttman et al., 2009; Khalil et al., 2009), circular RNAs (circRNAs) generated from back-spliced exons (Hansen et al., 2011; Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2012), intron-derived RNAs (Rearick et al., 2011; Yin et al., 2012), etc. Although detailed functions of many lncRNAs are only beginning to be elucidated, emerging lines of evidence have shown that they regulate multiple biological processes via a variety of mechanisms (for reviews, see Chen and Carmichael, 2010; Guttman and Rinn, 2012; Rinn and Chang, 2012; Wilusz et al., 2009).

While many known lncRNAs are polyadenylated, recent work has revealed that a number of Pol II-transcribed lncRNAs are processed in alternative ways (Burd et al., 2010; Hansen et al., 2011, 2013; Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2012; Sunwoo et al., 2009; Wilusz et al., 2008; Yap et al., 2010; Yin et al., 2012). *MALAT1* (also called *NEAT2*) and *Menβ* (also called *NEAT1_2*), two nuclear retained lncRNAs, are processed at their 3′ ends by RNase P (which is known to process the 5′ ends of tRNAs) (Sunwoo et al., 2009; Wilusz et al., 2008). RNase P cleavage leads to the formation of a mature 3′ end of each lncRNA, which is protected by a highly conserved triple helical structure composed of U●A-U and C●G-C base triples (Brown et al., 2012; Wilusz et al., 2012). Several recent reports identified nonlinearized RNAs that are largely generated from back-spliced exons (Burd et al., 2010; Hansen et al., 2013; Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2012; Yap et al., 2010). The *INK4a/ARF* locus-associated lncRNA *ANRIL* participates directly in epigenetic transcriptional repression (Yap et al., 2010), and this locus also encodes heterogeneous species of RNA transcripts including circular ANRIL isoforms (*cANRIL*) whose expression correlated with *INK4/ARF* transcription and ASVD risk (Burd et al., 2010). By sequencing rRNA-depleted RNAs digested with the RNA exonuclease, RNase R, in human fibroblasts, a large number of circRNAs containing non-colinear exons and *Alu* repeats were identified and proposed to act as competing endogenous RNAs in the cytoplasm (Jeck et al., 2013). In addition, more-recent studies further revealed that some of them can function as efficient microRNA sponges (Hansen et al., 2013; Memczak et al., 2013).

Intronic sequences account for over 20% of human genome and provide yet another source to generate ncRNAs that may lack both 5′ cap structures and 3′ poly(A) tails. Although it is generally believed that most introns or intron fragments are unstable (Rodríguez-Trelles et al., 2006), there are notable exceptions. For instance, some intronic sequences encode small nucleolar RNAs (snoRNAs) (Kiss and Filipowicz, 1995; Petfalski et al., 1998), microRNAs (Ladewig et al., 2012; Okamura et al., 2007), and other ncRNAs (Rearick et al., 2011). We have recently sequenced the long transcripts (>200 nt) of nonpolyadenylated transcriptomes of human cells and have identified that excised introns could accumulate to high levels (Yang et al., 2011). Further studies demonstrated one mechanism for the processing of such excised introns. In this case, introns containing two snoRNAs are processed from their ends by the snoRNA machinery, and the internal sequences between snoRNAs are not removed, leading to the accumulation of lncRNAs with snoRNA caps at both ends (Yin et al., 2012). However, little is known about how the rest of the excised introns without snoRNA ends are stabilized and whether they have regulatory potency in cells.

Here we developed a computational pipeline and identified hundreds of previously undescribed intronic lncRNAs from the poly(A)− RNA-seq in human cells. We further show that a large portion of these RNAs are circular. The formation of such circular intronic RNAs (ciRNAs) can be recapitulated using expression vectors, and their processing depends on consensus RNA motifs near 5′ splice site and branchpoint. In contrast to circRNAs from back-spliced exons (Hansen et al., 2013; Jeck et al., 2013; Memczak et al., 2013), we show that ciRNAs are associated with the nuclear insoluble fractionation and have little enrichment for microRNA target sites. Importantly, knockdown of such RNAs led to reduced expression of their parent genes, and lines of evidence suggest one possible function for ciRNAs as positive regulators of RNA Pol II transcription.

RESULTS

Nonpolyadenylated RNA-Seq with a Customized Computational Pipeline Identified Hundreds of Intronic Long Noncoding RNAs in Human Cells

We have recently identified excised introns that accumulate to high levels in the nonpolyadenylated, "poly(A)−," transcriptomes of HeLa cells and human embryonic stem cells (hESCs) (Yang et al., 2011). By applying known assemblers, such as Scripture (Guttman et al., 2010) and Cufflinks (Trapnell et al., 2010), however, we found many excised introns could not be correctly annotated (Figure 1A). We thus developed a custom computational pipeline by calling sequencing peaks within introns of poly(A)−/ribo− RNA-seq data set followed by a series of stringent criteria to systematically annotate intronic long noncoding RNAs with high confidence (see Figure S1 online). This computational pipeline allowed us to detect hundreds of such RNAs in H9 and HeLa cells, ranging from 200 nt to over 3,000 nt in length (Figure 1B). Coding capacity assessment revealed that the vast majority of them show little evidence for coding potential, and only seven from H9 cells and four from HeLa cells contain potential ORFs (Figure S1C), which were removed from our catalog prior to further analyses. While 450 and 299 such intronic RNAs were detected in H9 and HeLa cells (Table S1 and Table S2), respectively, only 53 are overlapped (Figure S1D), suggesting they are expressed in a cell-specific manner. Notably, the relative abundance of these intronic RNAs is comparable to that of their parent mRNAs (Figure S1E).

We next focused in greater detail on intronic RNAs identified in H9 cells, as these cells have a well-characterized karyotype. By comparing our data with the annotated human RefSeq database and a catalog of ncRNAs, we found that about half of intronic RNAs overlapped with RefSeq annotations, while 217 are undescribed (Figure 1C). Importantly, ten randomly selected and previously unreported intronic RNAs could be successfully detected by northern blot (NB) with expected sizes by the electrophoresis in native agarose gels (Figure 1D).

Many Intronic Long Noncoding RNAs Are Circular

It is generally believed that most introns or intron fragments are unstable (Rodríguez-Trelles et al., 2006). Intronic sequences capped with snoRNP complexes at both ends lead to the formation of stable lncRNAs (Yin et al., 2012); however, only six of such RNAs have been identified from H9 cells so far (Figure 1C). What other mechanism keeps the rest of these intronic sequences from degradation after splicing? One possible way could be the formation of circular RNAs from lariat introns that escape from debranching. It is known that RNase R can degrade linear RNAs and Y-structure RNAs, while preserving the loop portion of a lariat RNA (Suzuki et al., 2006). By sequencing rRNA-depleted, RNase R-digested RNAs, several recent studies identified thousands of circRNAs from back spliced exons (Jeck et al., 2013; Memczak et al., 2013; Salzman et al., 2012). However, due to the lack of poly(A)− enrichment, these studies largely ignored whether an RNA originated from poly(A)$^+$ or poly(A)− fractionation of transcripts and thus failed to identify the majority of intronic RNAs described here.

To investigate how intronic sequences resist exonuclease trimming after splicing, we collected poly(A)−/ribo− RNAs in H9 cells, followed by RNase R treatment and size selection for stranded RNA-seq (Figure 1E). By comparing the relative expression of intronic RNAs in the RNase R treated sample to that in the poly(A) −/ribo− sample (Yang et al., 2011), 103 RNase R-enriched (at least 2-fold) intronic RNAs are selected as circular molecules (ciRNAs) (Table S3), including ones validated by NB (Figure 1F). We further confirmed the existence of ciRNAs by using RT-PCR (Figure S2A) with sets of convergent primers followed by Sanger sequencing (data not shown).

Importantly, while ciRNAs ran very similarly to their linear isoforms on native Agarose gels (Figure 1D and Figure S2B), they migrated much more slowly in denatured PAGE gels and showed only one sharp band at exactly the same position as the RNase R-treated samples (Figures 2A and 2B), suggesting that these lariat-intron derived ciRNAs do not contain linear 3′ tails. Importantly, we could define circular boundary reads (Figure 2C, blue) flanking the 2′,5′-phosphodiester bonds of examined ciRNAs with a computational pipeline (Figure 2C and Figure S3A). Note that predicted circular boundary reads match exactly with Sanger sequencing results of ciRNAs (Figure 2C and data not shown). We further applied the similar analysis to six ENCODE poly(A)−/ribo− RNA-seq data sets from GM12878, HUVEC, HepG2, NHEK, HeLa S3, and K562 cell lines. Strikingly, circular boundary reads could also be successfully defined from these data sets (Figure S2C and Table S3), although these ciRNAs were missed in their annotations.

Analysis of histone modifications of the examined parent genes region using the ENCODE ChIP-seq from human embryonic stem cells (ESCs) revealed that ciRNAs likely do not contain their own promoters but rather derive from their parent transcripts (data not shown). Moreover, ciRNAs are more stable than their parent linear mRNAs, probably due to their circular structures (Figure 2D). Together, our stepwise enrichment for intronic long noncoding RNAs from the nonpolyadenylated fractionation of transcripts clearly revealed a subclass of previously undescribed but highly abundant circular RNAs produced by escape from debranching of intron lariats in human cells.

The Processing of ciRNAs Depends on a Consensus Sequence Motif in an Intron Lariat

Since ciRNAs are derived from introns that have failed to be debranched, we asked whether they are byproducts from aberrant

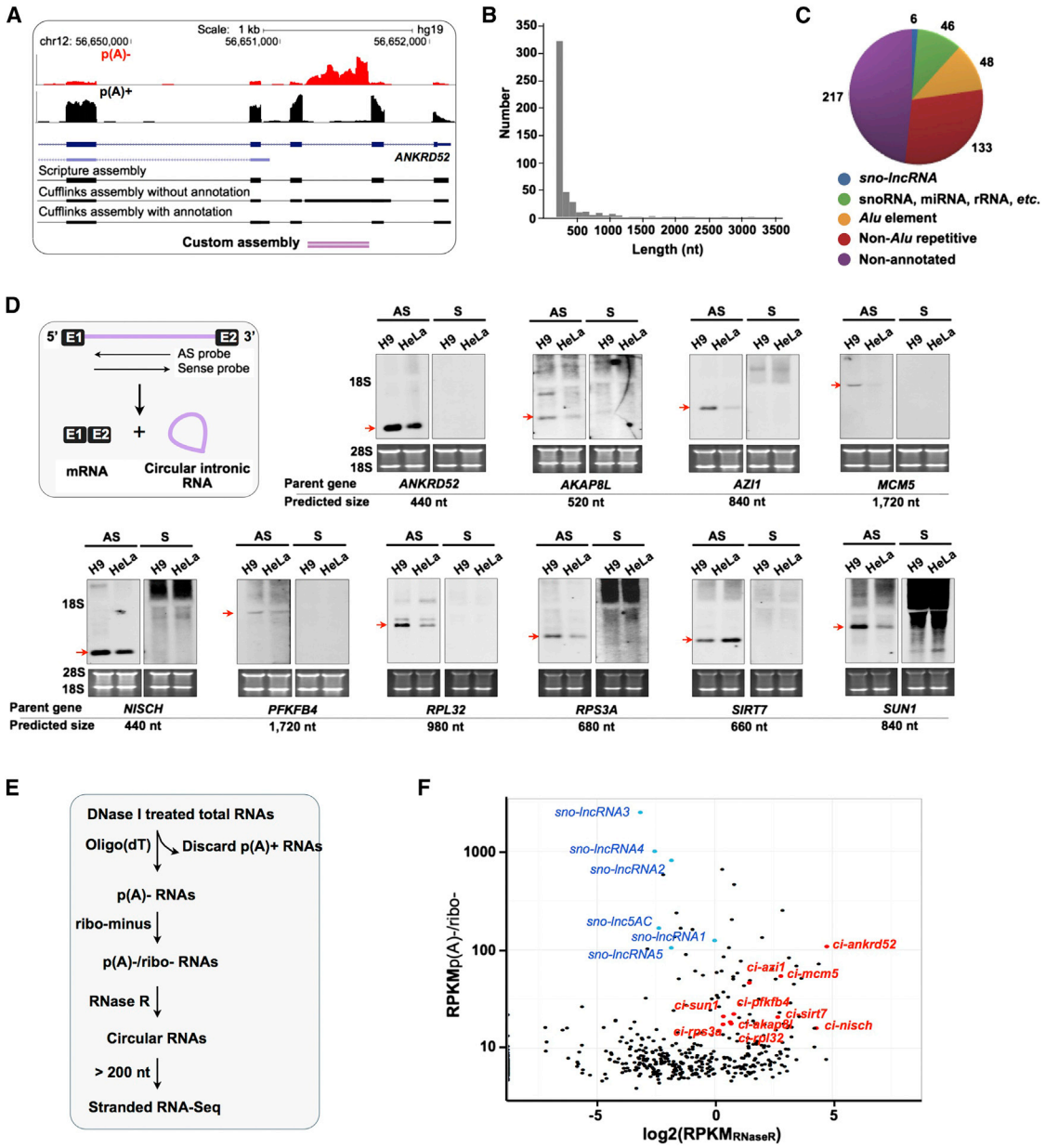


Figure 1. Identification of Circular Intronic RNAs

(A) An example of identified intronic long noncoding RNAs. Intronic RNA signal (red, poly(A)−/ribo− RNA-seq, p(A)− for simple in all figures) of *ANKRD52* is annotated by a custom pipeline but not correctly captured by other assemblers.

(B) Length distribution of identified intronic RNAs in H9 cells (histogram with bin width of 100 nt).

(C) Overlapping of intronic RNAs with RefSeq annotations in H9 cells.

(D) Validation of predicted intronic RNAs. Top left, a schematic drawing shows NB antisense (AS) and sense (S) probes recognize the intron (purple line) between exons (black boxes). NBs with AS probes detect all ten randomly selected intronic RNAs with expected sizes (red arrows). 28S and 18S rRNAs were used as loading controls on native agarose gel.

(E) A schematic flow shows the enrichment of ciRNAs.

(F) Scatterplot analysis of ciRNA enrichment by RNase R in H9 cells. x axis, fold change of the expression level of individual intronic RNAs in the RNase R-enriched sample (see E) versus that in the poly(A)−/ribo− RNA sample (log2). y axis, RPKM of intronic RNAs from the poly(A)−/ribo− RNA sample. Labeled intronic RNAs as examples for linear (blue dots) or circular (red dots) molecules. See also Figures S1, S2, and S7 and Table S1, Table S2, and Table S3.

splicing or whether a specific mechanism is associated with their processing. We reasoned that the processing of ciRNAs may involve specific *cis*-acting intronic sequences. We found this is the case. ciRNA from the gene *ANKRD52* (*ci-ankrd52*) can be recapitulated in expression vectors when the ciRNA-producing full-length intron is inserted along with its natural splice sites

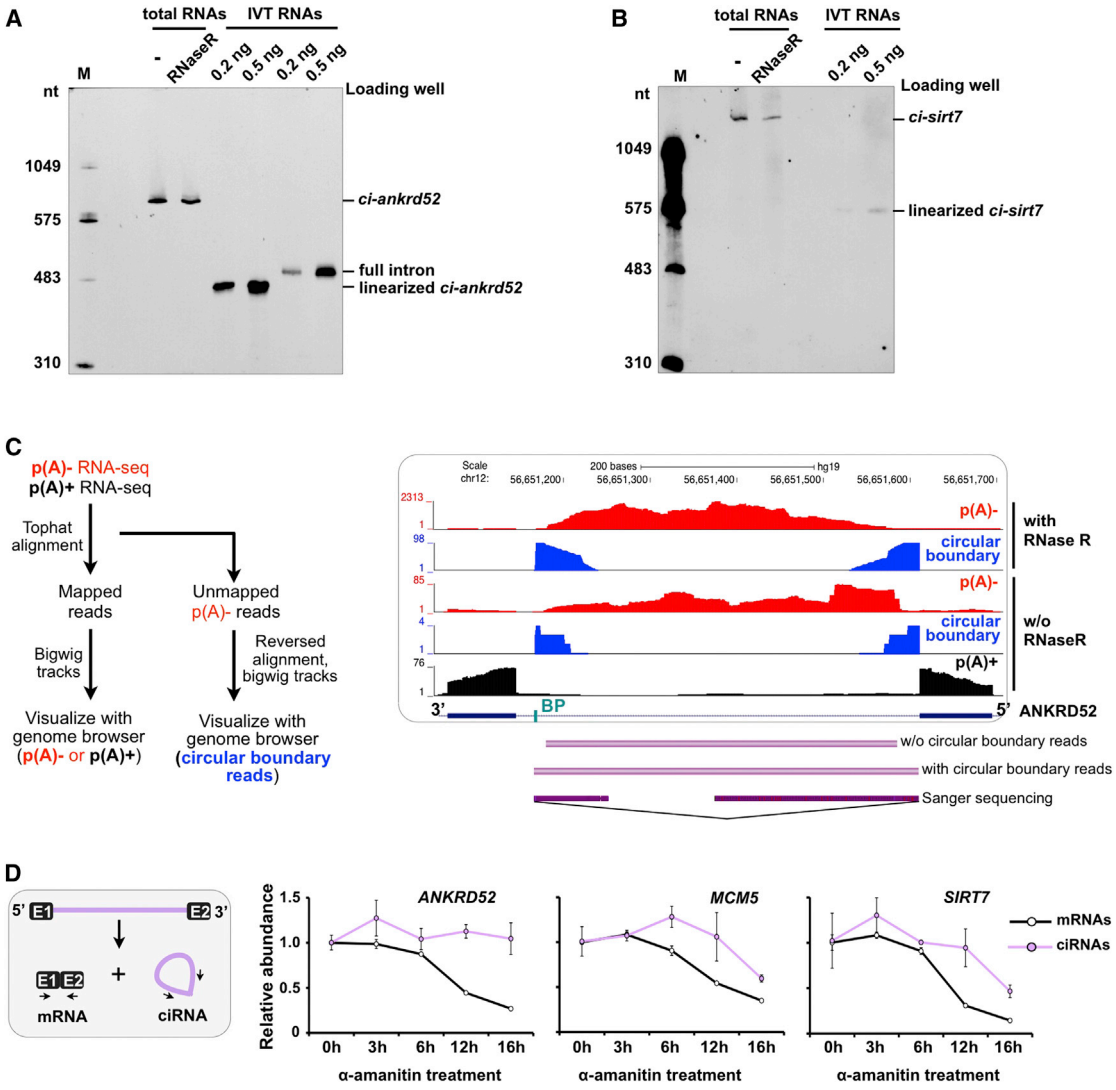


Figure 2. Characterization of ciRNAs

(A) RNase R untreated and treated H9 total RNAs, RNAs of full-length *ci-ankrd52* producing intron (466 nt) or linearized *ci-ankrd52* (444 nt) from IVT, were loaded on 5% denatured PAGE gel. NB was performed with the AS probe for *ci-ankrd52* in Figure 1D.

(B) NB for *ci-sirt7* on denatured PAGE gel.

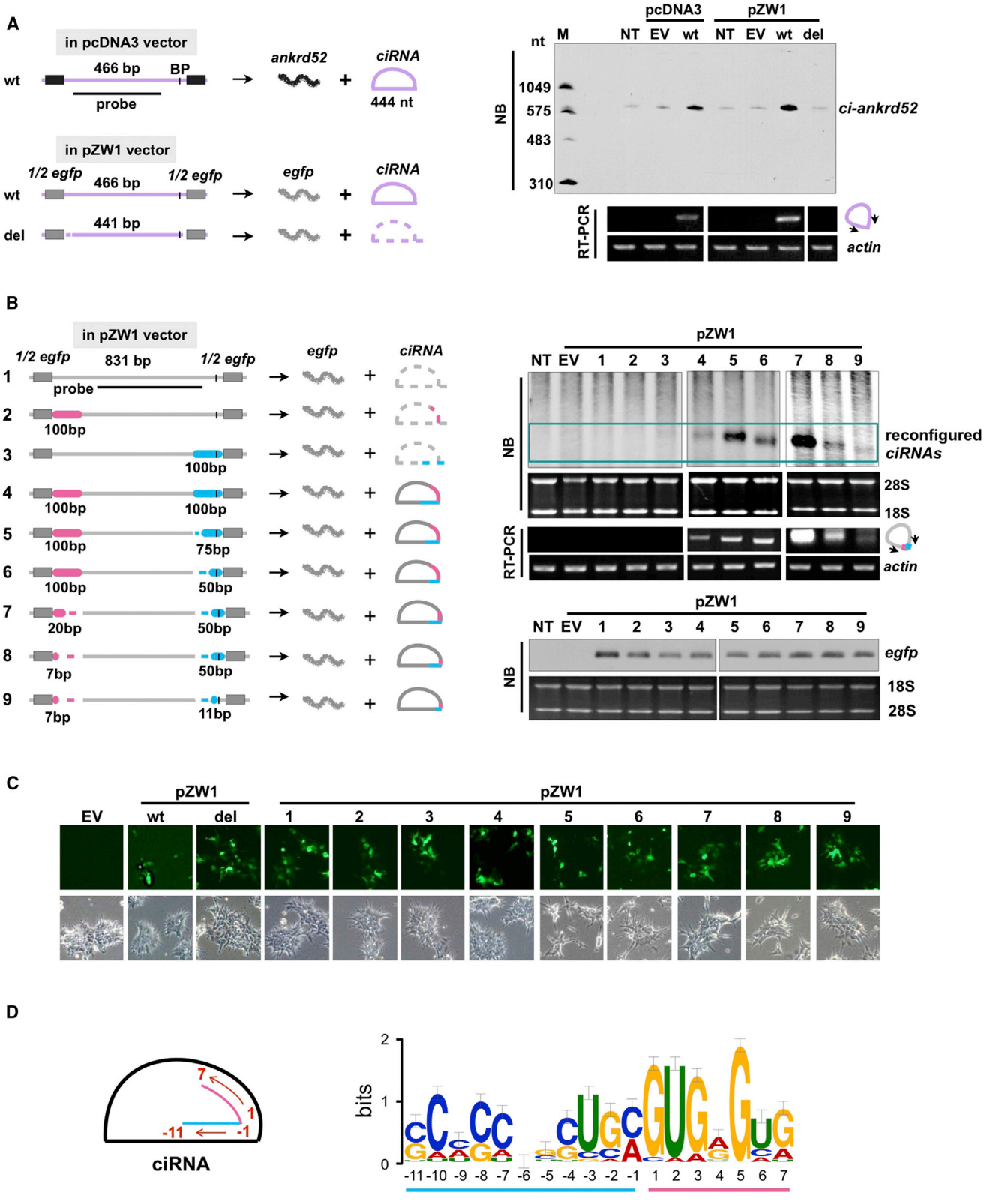
(C) Identification of circular boundary reads. Left, a stepwise analysis to identify circular boundary reads anchoring 2′,5′-phosphodiester bond of ciRNAs from RNA-seq data sets (see Figure S3A). Right, the visualization of *ci-ankrd52* annotation. *ci-ankrd52* in red, *ankrd52* mRNA in black, predicted circular boundary reads in blue, predicted annotation of *ci-ankrd52* in pink, and Sanger sequencing of *ci-ankrd52* in purple, obtained as shown in Figure S2A. Note that the circular boundary reads match to the exact branchpoint nucleotide (green), and no reads were aligned to sequences from branchpoint site to 3′ splice site.

(D) ciRNAs are stable. The relative abundance of each ciRNA and its parent mRNA were measured by RT-qPCR with primer sets shown in left after α-amanitin treatment. Error bars represent standard deviation (± SD) in triplicate experiments. See also Figures S2 and S3 and Table S3.

and exons, or when flanked by other exons (Figure 3A). However, deletion of the intronic sequences close to the 5′ splice site and the branchpoint failed to recapitulate *ci-ankrd52* (Figure 3A), suggesting critical sequences within such an intron are required for proper processing of ciRNAs. We observed that other such RNAs, such as *ci-sirt7*, could also be recapitulated in a similar way (Figure 4F).

To gain further insights into the processing mechanism, we inserted a non-ciRNA-associated full-length intron from *ANKRD52* into the same expression vector, and as expected, no stable intronic transcript could be detected (Figure 3B, lane 1). However, when sequences of ends of this intron were replaced with elements from the 5′ splice site and the branchpoint region of the ciRNA-producing intron, we observed that ciRNAs could now be expressed (Figure 3B, lanes 2–8). Note that the proper splicing occurred in all cases from transfected vectors (Wang et al., 2004), as indicated by mature *egfp* mRNA (Figure 3B) and EGFP expression (Figure 3C).

RNA sequences near the 5′ splice site and branchpoint can promote inefficient debranching, perhaps by generating



(legend on next page)

conformations that are inherently resistant to debranching. By selecting reads that are failed to map genomic regions but map to 2′,5′-phosphodiester sites (Figure S3A), we identified the putative branchpoint region and computed possible consensus motif from ciRNAs. This analysis revealed two types of branchpoint nucleotide, either A or C (Figure S3B), consistent with the notion that human branchpoint sequences are degenerate (Gao et al., 2008). Interestingly, when the circular structure was taken into consideration, we identified a consensus motif containing a 7 nt GU-rich element near the 5′ splice site and an 11 nt C-rich element close to the branchpoint in ciRNA-producing introns (Figure 3D and Figure S3C). This motif is not enriched in regular introns or other reported circRNAs (data not shown), indicating that these key RNA elements might be essential for an intron lariat to escape from debranching. Importantly, a regular intron could be reconfigured into a ciRNA by engineering these two critical elements into both ends (Figure 3B, lane 9), suggesting that these sequences are minimally sufficient for processing of ciRNAs, although the efficiency was low. However, as certain longer sequences containing these key RNA elements substantially increased the processing efficiency of ciRNAs (Figure 3B, lanes 5–8), other *cis*-elements are likely to further facilitate their expression.

ciRNAs Are Abundant in the Nucleus and Are Involved in Local Gene Expression

Since they are processed from introns and are not polyadenylated, we expected ciRNAs to be localized to the nucleus. By cytoplasmic and subnuclear fractionation, we found that ciRNAs derived from introns of *ANKRD52*, *MCM5*, and *SIRT7* are abundantly distributed in the nucleus and are associated with the nuclear insoluble fractionation, while their parent mRNAs are mainly located in the cytoplasm (Figure 4A). Interestingly, the nuclear localization of ciRNAs described here is distinct from circRNAs, which are generated from back splice circularization and act as microRNA sponges in the cytoplasm (Hansen et al., 2013; Jeck et al., 2013; Memczak et al., 2013), suggesting that ciRNAs may function differently from cytoplasmic circRNAs.

To explore their possible roles in the nucleus, we studied the abundant *ci-ankrd52* in greater detail. It is derived from the second intron of *ANKRD52*, which produces a large Ankyrin repeat domain-containing protein with a poorly known function (Stefansson et al., 2008; Zhang et al., 2007). Since we have successfully knocked down the nuclear retained and intron-derived *sno-lncRNA*s by phosphorothioate-modified antisense oligodeoxynucleotides (ASOs) without affecting the expression of their parent gene, *SNURF-SNRPN* (Yin et al., 2012), we reasoned that the similar approach could also be applied to study the function of intron-derived *ci-ankrd52* with appropriate ASOs. We designed different ASOs that can efficiently knock down *ci-ankrd52*, as revealed by both RT-PCR and NB (Figures 4B and 4D). However, surprisingly, we observed that knockdown of *ci-ankrd52* led to a significant reduction of *ankrd52* mRNA expression (Figures 4B and 4D), suggesting *ci-ankrd52* may have a role in local gene expression.

Since ASOa and ASOb that target *ci-ankrd52* are also complementary to the intron-containing pre-mRNA, it is possible that these ASOs directly bound to pre-mRNA of *ANKRD52*, leading to its degradation, which in turn caused reduced *ankrd52* mRNA expression. To exclude this possibility, we designed additional ASOs (ASOd-g) that target several adjacent introns of the ciRNA-producing intron. However, none of them led to a significant downregulation of *ankrd52* mRNA (Figure 4B). To exclude the possibility that some ASOs are more effective than others, we cloned the ciRNA-producing intron and its adjacent introns individually into vectors, which allowed the successful expression of each intronic sequence in the nucleus (Figure 4C). After cotransfection into cells with relevant ASOs, respectively, we found that each ASO can efficiently knock down its targeted intronic sequence (Figure 4C), demonstrating that all designed ASOs are effective. It also suggested that the reduction of parent gene expression was not affected by ASOa or ASOb on pre-mRNA but rather was caused by the knockdown of *ci-ankrd52*. Importantly, ASOc, which targets the non-ciRNA sequence in the ciRNA-producing intron (the sequence between the branchpoint site and the 3′ splice site of the second intron in *ANKRD52*), has no effect at all on *ankrd52* expression (Figure 4B), although it can efficiently knock down the second intron in the expression vector (Figure 4C). Thus, the striking differences of *ankrd52* mRNA after the treatment of ASOc or ASOa/b, which all target the same ciRNA-producing intron, strongly support the notion that the reduced expression of the parent mRNA is due to the knockdown of ciRNA rather than a direct effect of ASOs on target pre-mRNA.

This observed phenomenon could be further illustrated by introducing expression vectors into PA1 cells (Figure 4E), in which the full-length *ci-ankrd52*-producing intron, or the intron containing nucleotide deletions to destroy RNA consensus motifs near to 2′,5′-phosphodiester site, was inserted into the intron region of *egfp* mRNA such that only proper splicing leads

Figure 3. Processing of ciRNAs

(A) Recapitulation of *ci-ankrd52* processing. Left, a schematic drawing of *ci-ankrd52* in expression vectors. The *ci-ankrd52*-producing full intron (466 bp, purple line) was inserted into pcDNA3 along with its natural splice sites (top); the full intron (middle) or such an intron with deletions (bottom, dashed lines represent deletions) was inserted into pZW1 such that only proper splicing leads to EGFP fluorescence(Wang et al., 2004), also see (C). Right, equivalent amounts of RNA from cells transfected with each indicated plasmid were resolved on 5% denatured PAGE gel for NB or for RT-PCR to detect ciRNAs with convergent primer sets. NT, not transfected; EV, empty vector.

(B) A non-ciRNA-associated full intron can be reconfigured to produce a ciRNA. Left, a schematic drawing of the fifth intron of *ANKRD52* (831 bp, gray line) and such an intron containing a series of sequence replacements from *ci-ankrd52*-producing intron in pZW1. Pink and blue boxes indicate the length and position of each sequence replacement. Short vertical bars in black indicate branchpoint site. Right, equivalent amounts of RNA from cells transfected with each indicated plasmid were resolved on agarose gels for NBs to detect reconfigured ciRNAs or *egfp* mRNA.

(C) Each transfection in (A) and (B) achieved a similar efficiency as indicated by EGFP fluorescence intensity.

(D) The consensus RNA elements near the 5′ splice site and branch point in ciRNAs. Error bars (± SD) are generated by GLAM2 to indicate the sample correction. See also Figure S3.

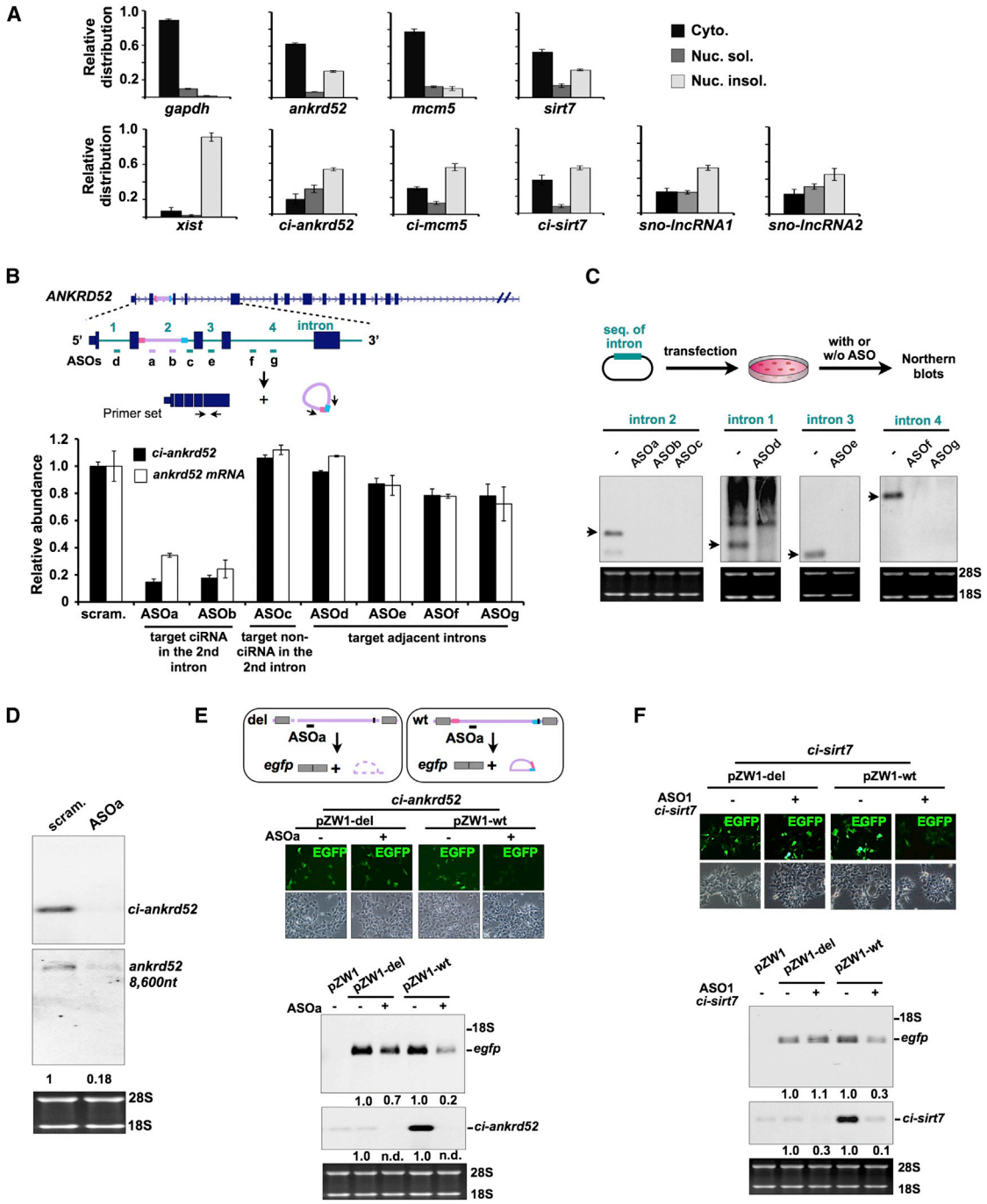


Figure 4. Nuclear Enriched ciRNA Regulates Its Parent Gene Expression

(A) ciRNAs are relatively abundant in nuclear insoluble fractions in PA1 cells.

Bar plots represent relative abundance of RNAs in each fractionation. The relative distribution of *gapdh*, *xist*, and *sno-lncRNAs* revealed a successful cytoplasmic, nuclear-soluble, and nuclear-insoluble fractionation.

(B) Knockdown of *ci-ankrd52* led to reduced expression of its parent gene in PA1 cells. Top, a schematic drawing of the gene organization of *ANKRD52* and *ci-ankrd52* is derived from the second intron. Colored bars represent ASOs, and black arrows represent primer sets. Bottom, the relative expression of *ci-ankrd52* and its parent mRNA under each ASO treatment. Bar plots represent relative expression of genes (normalized to actin).

(C) All designed ASOs are effective for their targeted intronic sequences in *ANKRD52*. Top, a schematic drawing of the experimental flow. Bottom, NBs revealed that all ASOs efficiently knocked down their targeted intronic sequences in PA1 cells. Black arrows indicate individual introns expressed from vectors on NBs.

(legend continued on next page)

to EGFP fluorescence (Wang et al., 2004). In this scheme, *ci-ankrd52* could only be made from the full-length intron (Figures 3A and 4E). The spliced *egfp* was efficiently produced in both vectors, where no pre-mRNA could be detected (Figure 4E). By examining fluorescence intensity, we observed that these two vectors produced comparable EGFP, but the intensity was clearly decreased in cells transfected with the full-length *ci-ankrd52*-producing intron followed by the ASOa treatment (Figure 4E, middle), although introns in both vectors contain the same targeted sites for ASOa. Importantly, *egfp* mRNA was also significantly decreased when *ci-ankrd52* was depleted (Figure 4E, bottom). Moreover, similar experiments revealed that knockdown of *ci-sirt7* led to a reduced expression of its parent gene (*egfp*) from the expression vector (Figure 4F). Together, these lines of evidence strongly suggest that ciRNA plays a role in local gene regulation.

In addition, we found that knockdown of other ciRNAs, such as *ci-mcm5* and *ci-sirt7*, led to a similar reduction on their parent mRNA expression, while treatment with ASOs targeted to the non-ciRNA sequence in ciRNA-producing intron or to an adjacent intron had no detectable effect on mRNAs expression (Figures S4A–S4D). Similar results were observed in H9 cells (Figures S4E–S4G) and HeLa cells (data not shown). Moreover, we found that knockdown of such RNAs had no dramatic effect on expression of genes upstream or downstream of the examined parent genes (Figures S5A–S5C). Finally, transient expression of *ci-ankrd52* has no detectable effect on the mRNA level of *ankrd52*, compared to that in normal cells (Figure S5D), suggesting that ciRNAs do not act in *trans*, probably due to their aberrant localization during overexpression (Figure S5E). Taken together, we conclude that nuclear enriched ciRNAs are involved in *cis* gene expression regulation.

ciRNAs Do Not Likely Modulate Posttranscriptional Local Gene Expression

lncRNAs act on gene express regulation via a variety of mechanisms (for reviews, see Chen and Carmichael, 2010; Guttman and Rinn, 2012; Rinn and Chang, 2012; Wilusz et al., 2009). In our scenario, we reasoned that at least three possibilities existed for the observed ciRNAs mediated in *cis* gene regulation. First, nuclear enriched ciRNAs may carry out a function similar to circRNAs (Hansen et al., 2013; Memczak et al., 2013), acting as microRNA sponges. In this model, knockdown of ciRNAs may result in the release of some microRNAs, which could potentially silence their parent genes. However, careful analyses of several abundant ciRNAs revealed that such RNAs contain only a few microRNA binding sites and that none of these sites are clustered (Figure 5A), suggesting these RNAs do not function as microRNA sponges.

The second possibility is that ciRNAs may be required for the proper processing of their parent pre-mRNAs. In this model, knockdown of such RNAs could affect downstream intron processing and exon definition, which in turn might generate premature stop codon(s), and subsequently caused nonsense-mediated decay (NMD) (Kervestin and Jacobson, 2012; Rebbapragada and Lykke-Andersen, 2009) to reduce the parent gene expression. To test this possibility, we first examined intron processing in *ANKRD52* by analyzing the relative abundance of splicing intermediates. We observed that the ciRNA-processing intron is processed at a similar rate compared to those of its downstream introns (Figure 5B). However, knockdown of *ci-ankrd52* slightly increased the retention of its downstream introns (Figure 5B), leading to the formation of new isoforms of RNAs with retained intron(s) (Figure 5C). Transfection of ASOs which target adjacent introns led to no obvious differences in the relative abundance of intermediates (data not shown), thus ruling out the act of ASOs on this effect. Although the resulting aberrant splicing isoforms only accounted for a small fraction of all detected isoforms (Figure 5C), these results nevertheless suggest that knockdown of *ci-ankrd52* affects downstream splicing events. Further analysis revealed that retained introns in new isoforms of *ankrd52* contain potential premature stop codons (data not shown), suggesting the likelihood of the involvement of NMD in the *ci-ankrd52*-mediated local gene silencing, especially if the altered RNAs were exported to the cytoplasm. However, knockdown of one of key factors in the NMD pathway, UPF1 (Kervestin and Jacobson, 2012; Rebbapragada and Lykke-Andersen, 2009) (Figure 5D), followed by knockdown of *ci-ankrd52*, did not rescue the observed downregulation of *ankrd52* mRNA (Figure 5E).

ci-ankrd52 Accumulates to Its Sites of Transcription

As *ci-ankrd52* does not likely modulate posttranscriptional local gene expression (Figure 5), we reasoned that it may affect the rate or efficiency of transcription. We thus studied the nuclear localization of *ci-ankrd52* in greater detail by RNA in situ hybridization (ISH). Since the antisense RNA probe (about 400 nt) for *ci-ankrd52* also recognizes the pre-mRNA of *ankrd52* during ISH, we first examined the sensitivity of such Digoxigen-labeled probes. Strikingly, the nascent *pre-ankrd52* cannot be detected until the long probe (4,000 nt) has been used (Figure 6A); thus, the 400 nt probe that we used to detect *ci-ankrd52* should only reveal mature ciRNA rather than nascent pre-mRNA.

Interestingly, there are two localization patterns of *ci-ankrd52* with its parent gene loci by RNA/DNA double FISH in the examined cells. About half of the cells showed a strong signal of *ci-ankrd52* accumulation to its sites of transcription, while others

(D) NB revealed the relative abundance of *ci-ankrd52* and *ankrd52* after knockdown of *ci-ankrd52*. The percentage of *ankrd52* under each treatment was determined using ImageJ and labeled underneath.
(E) Knockdown of *ci-ankrd52* led to reduced expression of its parent gene (*egfp*) in expression vectors. Top, a schematic drawing of vectors used in Figure 3A were transfected to PA1 cells followed by the treatment of ASOa which targets *ci-ankrd52*. Middle, EGFP intensity under each indicated treatment. Bottom, NB revealed that *egfp* mRNA was significantly reduced when *ci-ankrd52* was knocked down. The relative abundance of *egfp* and *ci-ankrd52* was determined using ImageJ and labeled underneath. n.d., not detectable.
(F) Knockdown of *ci-sirt7* led to reduced expression of its parent gene (*egfp*) in expression vectors.
In (A) and (B), error bars represent ± SD in triplicate experiments. See also Figures S4 and S5.

A

	CDR1as	*ci-ankrd52*	*ci-sirt7*	*ci-mcm5*
length (nt)	1485	440	680	1680
number of seed matches for top miRNA hits	73 (miR-7) 22 (miR-876-5p) 4 (miR-135ab)	1 (miR-136) 1 (miR-491-5p) 1 (miR-544)	2 (miR-31) 2 (miR-485-5p)	2 (miR-183) 2 (miR-197) 2 (miR-491-5p) 2 (miR-874)

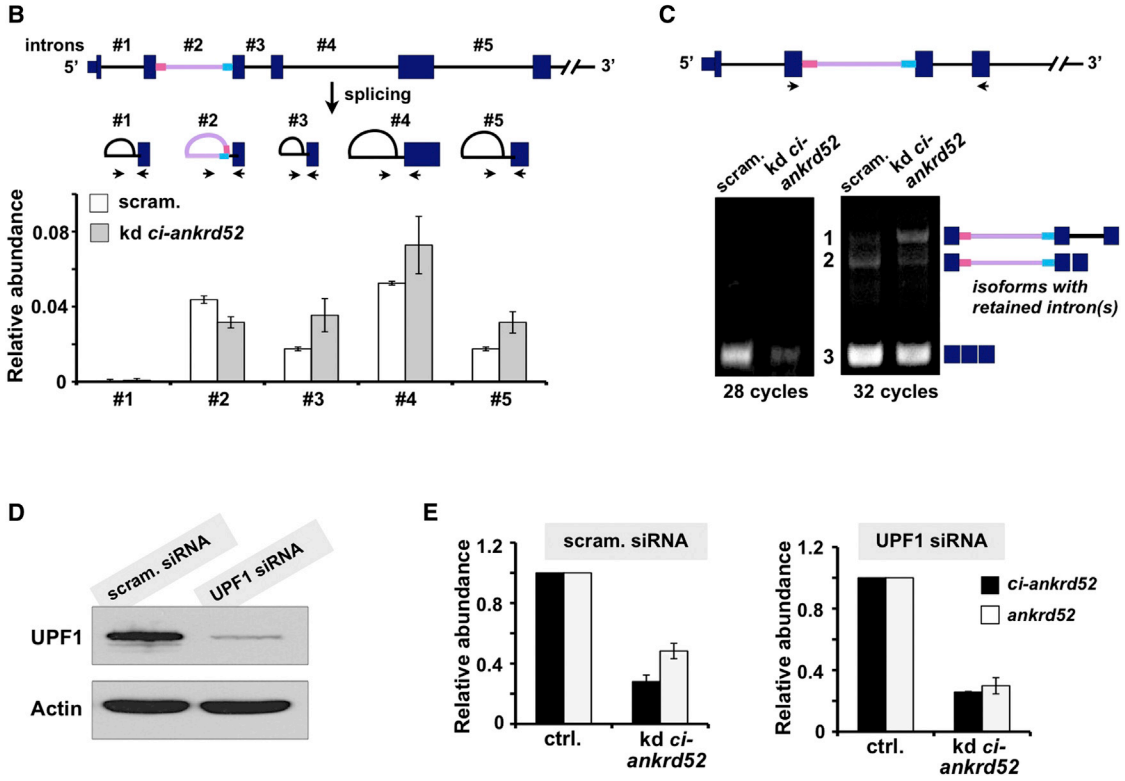


Figure 5. *ci-ankrd52* Does Not Likely Modulate Posttranscriptional Local Gene Expression

(A) ciRNAs show little enrichment for microRNA target sites. The reported *CDR1as* (Memczak et al., 2013) and representative ciRNAs are shown.

(B) The processing of intron lariats in *ANKRD52* before and after knockdown of *ci-ankrd52* in PA1 cells. Top, a schematic drawing of the processing of each indicated intron intermediate, and arrows indicate PCR primer sets. Bottom, bar plots represent relative abundance of intron intermediates under each treatment (normalized to total *ankrd52* RNAs).

(C) Knockdown of *ci-ankrd52* affected the downstream intronic splicing. Semiquantitative RT-PCRs were carried out in control or *ci-ankrd52* knockdown samples. Each PCR band was sequenced, and the relative isoform was drawn.

(D) Knockdown of UPF1 by siRNAs in PA1 cells. Actin was used as a loading control.

(E) Knockdown of UPF1 cannot rescue *ci-ankrd52*-mediated local gene regulation. The *ci-ankrd52* knockdown and scramble treatment were performed in both control and UPF1-depleted PA1 cells, and total RNAs were collected for RT-qPCR analysis.

In (B) and (E), error bars represent ± SD in triplicate experiments.

exhibited weak colocalization (Figure 6B). Importantly, by carefully investigating detectable transcription loci of *ANKRD52* in both H9 and HeLa cells, we found that the majority of these loci are associated with *ci-ankrd52* (Figures 6C and 6D). For example, among 489 transcription loci of *ANKRD52* in H9 cells that we counted, 91% exhibited colocalization with *ci-ankrd52* RNA (Figure 6C). Although we do not yet know what other factors are associated with non-transcription-site-localized *ci-ankrd52* RNA (Figure 6B and data not shown), the specific colocalization with genomic loci of its parent gene in most cells clearly supports one function of such RNAs, that they are involved in the local gene expression regulation (Figure 4).

ci-ankrd52 Interacts with Pol II Machinery and Modulates Its Transcription Activity

Due to the localization to its sites of transcription and function in *cis*, we then examined whether ciRNAs are associated with the transcription machinery. First, we incorporated biotin into the linearized *ci-ankrd52* RNA by in vitro transcription. After incubation with nuclear extracts isolated from PA1 cells, we observed

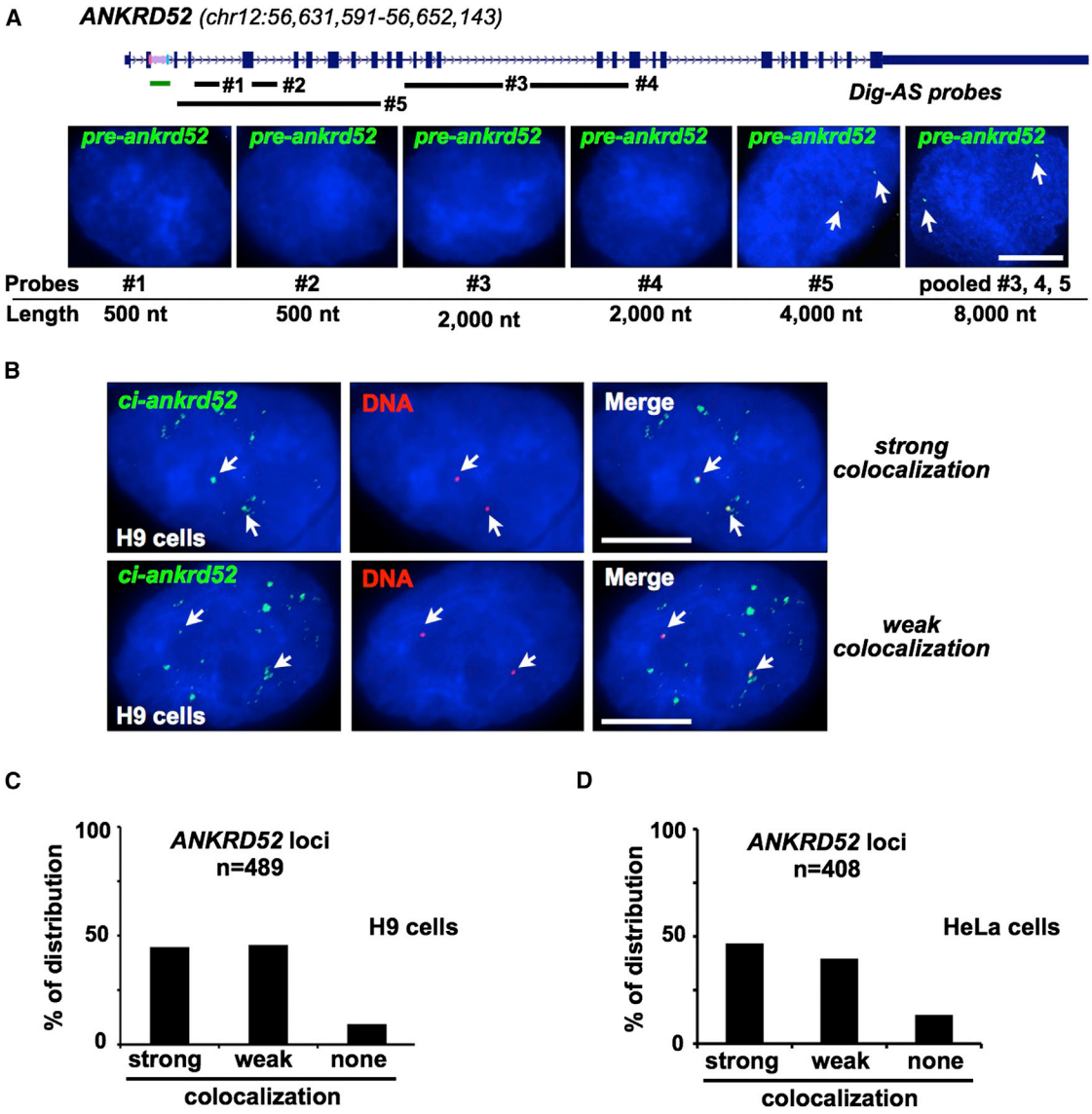


Figure 6. *ci-ankrd52* Accumulates to Its Sites of Transcription

(A) The detection of *ankrd52* pre-mRNA requires long AS probe. Top, a schematic drawing of *ANKRD52* and the length and position of Dig-labeled AS probes. Bottom, representative images of *ankrd52* pre-mRNA ISH with different length of probes in H9 cells are shown. White arrows indicate signals of *ankrd52* pre-mRNA. Note that detectable signals for *ankrd52* pre-mRNA require a Dig-AS probe as long as 4,000 nt. Nuclei were counterstained with DAPI. The white scale bar denotes 5 μm in all images.

(B) Double FISH of *ci-ankrd52* (green) and its adjacent DNA region (*ANKRD52*, red) in H9 cells. Representative images of two colocalization patterns of *ci-ankrd52* with its transcription sites are shown. White arrows indicate the colocalization of *ci-ankrd52* with its parent gene loci.

(C) The majority of *ANKRD52* transcription loci are associated with *ci-ankrd52* in H9 cells. In total, 489 *ANKRD52* transcription loci were counted from 266 cells.

(D) The majority of *ANKRD52* transcription loci are associated with *ci-ankrd52* in HeLa cells. In total, 408 *ANKRD52* transcription loci were counted from 105 cells.

that only the linearized sense *ci-ankrd52*, but not the antisense RNA, could specifically precipitate phosphorylated RNA Pol II (Figure 7A), suggesting the interaction between *ci-ankrd52* and the elongation Pol II can be RNA sequence dependent. Second, costaining of RNA and proteins also revealed that *ci-ankrd52* is at least partially colocalized with phosphorylated Pol II (Figure S6A). Third, by native RNA immunoprecipitation, we found that several ciRNAs were specifically coprecipitated with antisera directed against phosphorylated Pol II, but not with those targeting an abundant nuclear protein, Fibrillarin (Figure 7B). Fourth, we observed an even stronger association of ciRNAs with phosphorylated Pol II, but not with Fibrillarin, under UV or formaldehyde crosslinking conditions (Figures 7C and 7D, Figures S6B and S6C). Meanwhile, another abundant intron-derived lncRNA, *H/ACA sno-lncRNA* (Yin et al., 2012), could not be pulled down by either antibody, demonstrating the specificity of the RNA precipitation assays. Together, these analyses indicate that ciRNAs are associated with the elongation Pol II complex.

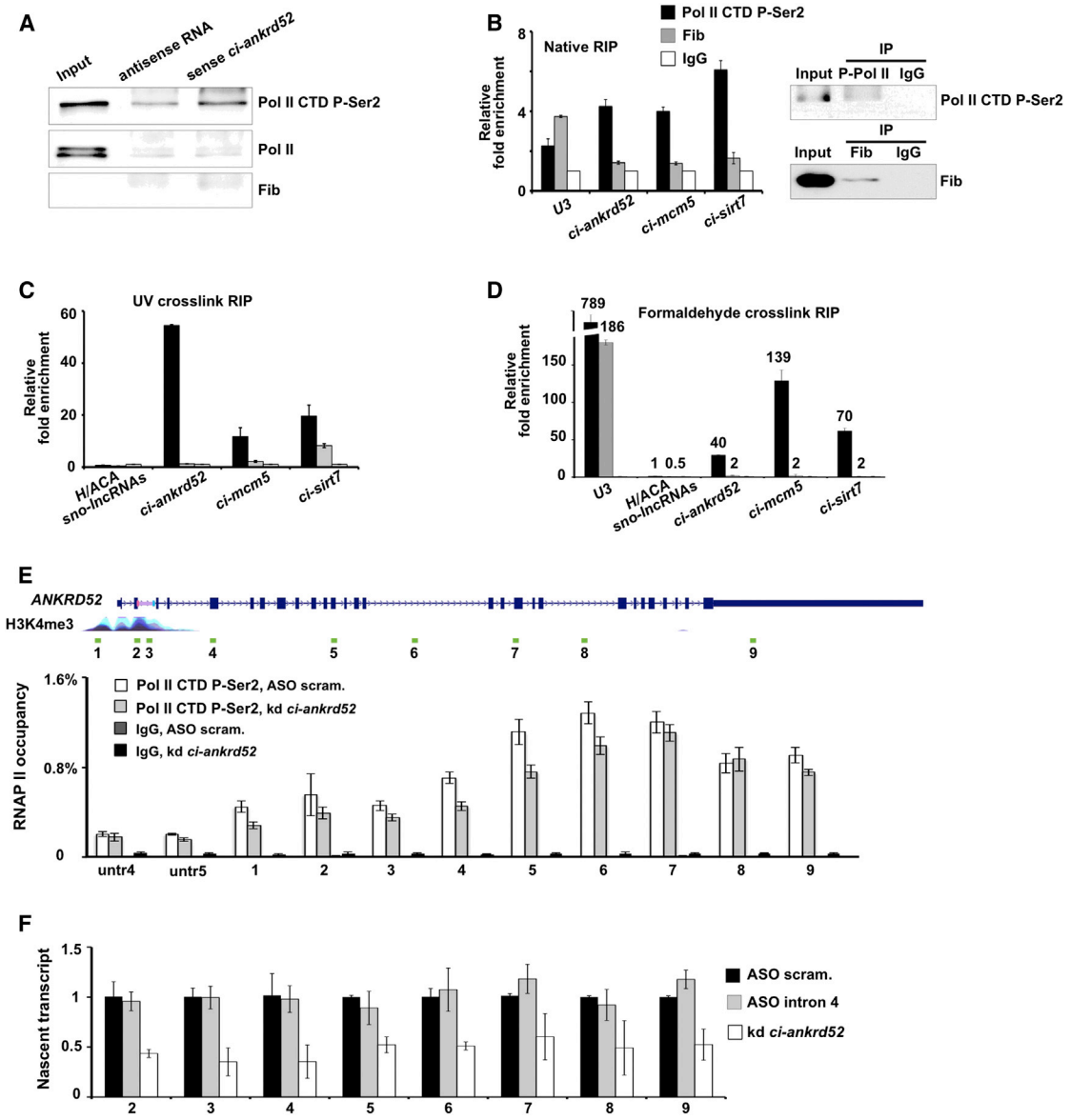


Figure 7. *ci-ankrd52* Modulates RNA Pol II Transcription

(A) Linearized *ci-ankrd52* and phosphorylated RNA Pol II interact in vitro. Biotin-RNA pull-downs using linearized *ci-ankrd52* transcript in PA1 nuclear extracts showed specific binding to phosphorylated Pol II. The nonphosphorylated Pol II and fibrillarin were shown as controls.

(B) Association between endogenous ciRNAs and phosphorylated Pol II by native RNA immunoprecipitation (RIP). Left, native RIP was performed from PA1 cells using anti-Pol II, anti-fibrillarin, and anti-IgG, followed by RT-qPCR. Right, WB shows the precipitation with each indicated antibody.

(C) Association between ciRNAs and phosphorylated Pol II by UV crosslinking RIP from PA1 cells.

(D) Association between ciRNAs and phosphorylated Pol II by formaldehyde crosslinking RIP from PA1 cells.

(E) Chromatin IP with anti-Pol II in scramble and *ci-ankrd52* ASO-treated PA1 cells. Data were expressed as the percentage of Pol II coprecipitating DNAs along *ANKRD52* versus input under each indicated condition.

(F) A crude preparation of nuclei was subjected to nuclear run-on assays under indicated conditions in PA1 cells. Nascent transcription of *ANKRD52* detected from scramble ASO treated nuclei was defined as one.

In (B)–(D), bar plots represent enrichment of RNAs immunoprecipitated by each indicated antibody over IgG for individual RNAs.

In (B)–(F), error bars represent ± SD in triplicate experiments. See also Figure S6.

Since overexpression of *ci-ankrd52* failed to increase endogenous *ANKRD52* expression (Figure S5D), probably due to the aberrant localization of these overexpressed RNAs (Figure S5E), we next sought to knock down ciRNAs and then analyzed phosphorylated Pol II occupancy and nascent RNA transcription. Although knockdown of *ci-ankrd52* has no dramatic effect on Pol II engagement on the gene body (Figure 7E), the depletion of *ci-ankrd52* resulted in a reduced transcription rate of the

nascent transcript of *ankrd52* as revealed by nuclear run-ons (Figure 7F). In addition, we included a control ASO (ASOg) that targets the fourth intron of *ANKRD52* (Figure 4B) in the nuclear run-on experiment. Although we observed a slight decrease (~20%) of *ankrd52* expression after treating PA1 cells with this ASO (Figure 4B), we repeatedly observed that the transcription of nascent *ankrd52* transcripts was not altered under the same treatment (Figure 7F). This observation thus indicated that the reduced transcription efficiency of pre-mRNA of *ankrd52* was mediated by depletion of *ci-ankrd52*. In addition, it was the case for *ci-sirt7*, where depletion of this RNA slowed down its parent pre-mRNA transcription although Pol II engagement on the gene was not altered (Figures S6D and S6E). Together, these lines of evidence suggest that some ciRNAs function as positive regulators of Pol II transcription and play a role in the efficient transcription of their parent genes.

ciRNAs Are Widely Expressed in Human Cells but Are Less Evolutionarily Conserved

Interestingly, we found that ciRNAs such as *ci-ankrd52* (and others, data not shown) are widely expressed in many human tissue samples, upon hESC lineage-specific differentiation (Figure S7A) and in other cultured cell lines (data not shown) but also exhibit a strong tissue-/cell-specific expression pattern. In addition, the expression of ciRNAs and their parent mRNAs is positively correlated in most of examined cell types and tissues (Figure S7A and data not shown), suggesting that such RNAs and their local effects on gene expression regulation are widespread in human cells. However, the tissue-specific expression of ciRNAs also indicates that additional non-RNA elements may be involved in their processing.

Although their parent mRNAs are abundantly expressed, many ciRNAs are species specific to human cells, as they are absent in examined mouse cell and tissue samples (Figures S7B). We then analyzed sequence conservation within intronic RNAs and their adjacent exons by calculating PhastCons scores from multiple alignments of primate genomes. Interestingly, intronic RNAs are much less evolutionarily conserved compared with their adjacent exons (Figure S7C). Further analyses revealed that transfection of human ciRNA-expressing vectors into mouse cells successfully generated these RNAs, suggesting they can be made in mouse (Figure S7D). However, comparison of human ciRNA-producing introns with those in the mouse genome revealed that the murine branchpoint sites do not contain the same consensus RNA elements for ciRNA formation (Figure S7E), suggesting that evolutionarily unconserved intronic sequences could add more complexity to human genomes.

DISCUSSION

It is becoming increasingly apparent that intronic sequences are able to encode functional ncRNAs, including both small and long ncRNAs (Kiss and Filipowicz, 1995; Ladewig et al., 2012; Okamura et al., 2007; Petfalski et al., 1998; Rearick et al., 2011; Yin et al., 2012). Earlier studies have reported that a spliced intron could accumulate as a set of lariat RNA structures with different length tails in the nucleus (Qian et al., 1992) and that stable lariat introns of some viruses were present in infected cells due to atypical splicing (Wu et al., 1998). We report the systematic identification of ciRNAs in human cells (Figures 1 and 2 and Table S3). These RNAs are derived from introns, escape debranching, and depend on consensus RNA elements near the 5′ splice site and the branchpoint for proper processing (Figure 3). However, we do not yet know how these consensus RNA elements work to resist debranching and what protein partners are involved in this process. It is known that a significant fraction of long Pol II transcripts may lack a canonical poly(A) tail (Cheng et al., 2005; Wu et al., 2008; Yang et al., 2011). While linear RNAs with noncanonical 3′ ends (Sunwoo et al., 2009; Wilusz et al., 2008; Yin et al., 2012) and circular RNAs produced from back-spliced exons (Hansen et al., 2013; Jeck et al., 2013; Memczak et al., 2013) account for many RNAs that lack a poly(A) tail, we report here that ciRNAs represent yet another type of circular RNA molecules derived from intronic sequences.

ciRNAs are abundant in the nucleus, associated with the nuclear insoluble fraction (Figure 4A), and are involved in the regulation of expression of their parent genes (Figures 4B–4F, Figures S4 and S5). These features distinguish them from circRNAs, which primarily localize to the cytoplasm and function as microRNA sponges (Hansen et al., 2013; Jeck et al., 2013; Memczak et al., 2013). However, some abundant ciRNAs, such as *ci-ankrd52*, also localize to scattered but striking nuclear dots in addition to their sites of synthesis (Figure 6B), strongly indicating that such RNAs have additional roles besides local effects. It is worthy to mention that a recent study showed that after depletion of the nuclear debranching enzyme, many intronic lariats accumulate in the cytoplasm and likely act as decoys to sequester TDP43, thus suppressing TDP43 toxicity in an ALS disease model (Armakola et al., 2012). Therefore, some abundant ciRNAs may function as "molecular sponges" in the nucleus for TDP43 and other RNA binding proteins to regulate gene expression in *trans* under certain circumstances. However, as the loss of function (LOF) of ciRNAs affects their parent gene expression, the genome-wide analysis of such LOF cells is not likely to be generally informative.

Transcription by RNA Pol II is known to be regulated by some ncRNAs. In *C. elegans*, nuclear small regulatory RNAs inhibit Pol II during the elongation phase of transcription and cotranscriptionally silence nuclear localized RNAs (Guang et al., 2010). In mammalian cells, mouse B2 RNA and human Alu RNA bind directly and tightly to Pol II and potentially assemble into preinitiation complexes at promoter regions to repress transcription of mRNA genes during heat shock response (Allen et al., 2004; Espinoza et al., 2004; Yakovchuk et al., 2009). Moreover, recent global run-on analyses revealed that the accumulation and pausing of Pol II at many promoters can be targeted for gene regulation (Core and Lis, 2008; Core et al., 2008). For example, lncRNAs transcribed in upstream regions of *CCND1* (Wang et al., 2008) and *DHFR* (Martianov et al., 2007) mediate transcription suppression by targeting transcription initiation machineries. In our scenario, the coupling between transcription and splicing may allow processed ciRNAs to associate with the elongating Pol II complex at their parent gene loci and enhance transcription activity. In support of this hypothesis, we have observed that ciRNAs partially accumulate to their sites of synthesis (Figure 6) and colocalize with elongation Pol II (Figure S6A). Moreover, they

are specifically associated with phosphorylated Pol II (Figures 7A–7D, S6B, and S6C), and depletion of them leads to significant reduction in parent gene transcription (Figure 7F and Figure S6E).

Finally, it is known that regulatory lncRNAs are less conserved than protein-coding genes (Cabili et al., 2011; Derrien et al., 2012). For example, braveheart, a lncRNA that is required for cardiovascular lineage commitment, is only detected in mouse, but is in no orthologous sequence in human or rat genomes (Klattenhoff et al., 2013). Interestingly, we found that many ciRNAs are species specific to human cells (Figure S7). As such RNAs are tissue and cell specific, further application of approaches similar to those we have used here for different tissues and species may result in the identification of additional ciRNAs. A recent study revealed many stable intronic sequence RNA (sisRNA) in the oocyte nucleus of *Xenopus* tropicalis; however, why sisRNAs are stable and what they do in oocytes have not yet been investigated (Gardner et al., 2012). While some sisRNAs could be potentially produced by a similar mechanism as ciRNAs, several examined ciRNAs exert regulatory roles on their parent genes that are unique to human cells, indicating that this unexpected regulatory function of ciRNAs may be evolutionarily selected.

EXPERIMENTAL PROCEDURES

Detailed experimental procedures are available in the Supplemental Information.

Computational Pipelines for Intronic RNA Annotation and ciRNA Enrichment

See the Supplemental Information and Figures S1–S3 for details.

Identification of Consensus Motifs for ciRNAs

ciRNAs with $RPKM_{poly(A)-/ribo-} \geq 10$ were selected for the identification of RNA-seq reads anchored to 2′,5′-phosphodiester bond regions. See Figure S3 and the Supplemental Information for details. Similar analyses were also applied to six ENCODE data sets (GEO accession number GSE26284).

RNA In Situ Hybridization, Immunofluorescence, and RNA/DNA FISH

RNA in situ hybridyzation using in vitro-transcribed Dig-labeled antisense probes was carried out as previously described (Yin et al., 2012). For colocalization studies with Pol II, cells were fixed for 5 min in 2% formaldehyde after RNA ISH, and immunofluorescence was performed with anti-Pol II CTD phospho Ser2 (Abcam). For RNA/DNA FISH, after RNA ISH, cells were denatured and then hybridized with denatured fluorescence-labeled PCR products recognizing a 20 kb region of chr12q13.3 close to *ANKRD52* genomic locus overnight. PCR primers are listed in Table S4. After hybridization and washes, the nuclei were counterstained with DAPI, image analyses were performed on single z stacks acquired with an Olympus IX70 DeltaVision RT Deconvolution System microscope.

Biotin-Labeled RNAs to Pull Down Proteins

Biotinylated RNA pull-down assay was performed as described (Klattenhoff et al., 2013; Tsai et al., 2010) with modifications. Linearized ciRNAs were in vitro transcribed with the biotin RNA Labeling Mix and were then incubated with nuclear lysates followed by extensive washes. The retrieved proteins were analyzed by western blot.

Native RNA-Protein Complex Immunoprecipitation, UV Crosslinking RIP, and Formaldehyde Crosslinking RIP

Native RNA-protein complex immunoprecipitation (RIP) assays were carried out as previously described (Yin et al., 2012; Yeo et al., 2009; Niranjanakumari et al., 2002) with modifications. See the Supplemental Information for details.

Chromatin Immunoprecipitation

See the Supplemental Information for details.

Nuclear Run-On Assay

The nuclear run-on assays in PA1 cells under different treatments were performed as reported by Core (Core et al., 2008) and Guang (Guang et al., 2010) with minor modifications. See the Supplemental Information for details.

ACCESSION NUMBERS

The NCBI Gene Expression Omnibus accession number for the sequence and the bigWig track file of poly(A)-/ribo-enriched and RNase R-digested RNAs in H9 cells reported in this paper is GSE48003.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, four tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.08.017.

ACKNOWLEDGMENTS

We are grateful to G. Carmichael for critical reading of the manuscript; H. Cheng, J. Hui, and G. Wang for materials; and D. Li and all lab members for helpful discussion. H9 cells were obtained from the WiCell Research Institute. This work was supported by grants XDA01010206 and 2012OHTP08 from CAS, grant 2011CBA01105 from MOST, grant 31271376 from NSFC, and grants 2012SSTP01 and 2012KIP101 from SIBS. L.-L.C. and L.Y. designed the project. Y.Z., T.C., J.-F.X., Q.-F.Y., and Y.-H.X. performed experiments. X.-O.Z. and S.Z. performed bioinformatics analyses. L.-L.C., L.Y., Y.Z., and X.-O.Z. analyzed the data. L.-L.C. wrote the paper with input from the other authors.

Received: April 9, 2013
Revised: June 27, 2013
Accepted: August 6, 2013
Published: September 12, 2013

REFERENCES

Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat. Struct. Mol. Biol. *11*, 816–821.

Armakola, M., Higgins, M.J., Figley, M.D., Barmada, S.J., Scarborough, E.A., Diaz, Z., Fang, X., Shorter, J., Krogan, N.J., Finkbeiner, S., et al. (2012). Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. Nat. Genet. *44*, 1302–1309.

Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2012). Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. Proc. Natl. Acad. Sci. USA *109*, 19202–19207.

Burd, C.E., Jeck, W.R., Liu, Y., Sanoff, H.K., Wang, Z., and Sharpless, N.E. (2010). Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. PLoS Genet. *6*, e1001233, http://dx.doi.org/10.1371/journal.pgen.1001233.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Chen, L.-L., and Carmichael, G.G. (2010). Long noncoding RNAs in mammalian cells: what, where, and why? Wiley Interdiscip. Rev. RNA *1*, 2–21.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science *308*, 1149–1154.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat. Struct. Mol. Biol. *11*, 822–829.

Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. Nucleic Acids Res. *36*, 2257–2267.

Gardner, E.J., Nizami, Z.F., Talbot, C.C., Jr., and Gall, J.G. (2012). Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. Genes Dev. *26*, 2550–2559.

Guang, S., Bochner, A.F., Burkhart, K.B., Burton, N., Pavelec, D.M., and Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature *465*, 1097–1101.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. *28*, 503–510.

Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., and Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J. *30*, 4414–4422.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Jeck, W.R., 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.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. *13*, 700–712.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA *106*, 11667–11672.

Kiss, T., and Filipowicz, W. (1995). Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. Genes Dev. *9*, 1411–1424.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell *152*, 570–583.

Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O., and Lai, E.C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. Genome Res. *22*, 1634–1645.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670.

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.

Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M.A. (2002). Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. Methods *26*, 182–190.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89–100.

Petfalski, E., Dandekar, T., Henry, Y., and Tollervey, D. (1998). Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. Mol. Cell. Biol. *18*, 1181–1189.

Qian, L., Vu, M.N., Carter, M., and Wilkinson, M.F. (1992). A spliced intron accumulates as a lariat in the nucleus of T cells. Nucleic Acids Res. *20*, 5345–5350.

Rearick, D., Prakash, A., McSweeny, A., Shepard, S.S., Fedorova, L., and Fedorov, A. (2011). Critical association of ncRNA with introns. Nucleic Acids Res. *39*, 2357–2366.

Rebbapragada, I., and Lykke-Andersen, J. (2009). Execution of nonsense-mediated mRNA decay: what defines a substrate? Curr. Opin. Cell Biol. *21*, 394–402.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rodríguez-Trelles, F., Tarrío, R., and Ayala, F.J. (2006). Origins and evolution of spliceosomal introns. Annu. Rev. Genet. *40*, 47–76.

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, http://dx.doi.org/10.1371/journal.pone.0030733.

Stefansson, B., Ohama, T., Daugherty, A.E., and Brautigan, D.L. (2008). Protein phosphatase 6 regulatory subunits composed of ankyrin repeat domains. Biochemistry *47*, 1442–1451.

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res. *19*, 347–359.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. Nucleic Acids Res. *34*, e63, http://dx.doi.org/10.1093/nar/gkl151.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689–693.

Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831–845.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130.

Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell *135*, 919–932.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. *23*, 1494–1504.

Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.D., Joshua-Tor, L., and Sharp, P.A. (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. Genes Dev. *26*, 2392–2407.

Wu, T.T., Su, Y.H., Block, T.M., and Taylor, J.M. (1998). Atypical splicing of the latency-associated transcripts of herpes simplex type 1. Virology *243*, 140–149.

Wu, Q., Kim, Y.C., Lu, J., Xuan, Z., Chen, J., Zheng, Y., Zhou, T., Zhang, M.Q., Wu, C.I., and Wang, S.M. (2008). Poly A- transcripts expressed in HeLa cells. PLoS ONE *3*, e2803, http://dx.doi.org/10.1371/journal.pone.0002803.

Yakovchuk, P., Goodrich, J.A., and Kugel, J.F. (2009). B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. Proc. Natl. Acad. Sci. USA *106*, 5569–5574.

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.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell *38*, 662–674.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat. Struct. Mol. Biol. *16*, 130–137.

Yin, Q.F., Yang, L., Zhang, Y., Xiang, J.F., Wu, Y.W., Carmichael, G.G., and Chen, L.L. (2012). Long noncoding RNAs with snoRNA ends. Mol. Cell *48*, 219–230.

Zhang, Y., Zhang, H., Liang, J., Yu, W., and Shang, Y. (2007). SIP, a novel ankyrin repeat containing protein, sequesters steroid receptor coactivators in the cytoplasm. EMBO J. *26*, 2645–2657.