

## Molecular determinants that govern scaRNA processing by Drosha/DGCR8

Douglas M. McLaurin, Madelyn K. Logan, Katheryn E. Lett and Michael D. Hebert\*

Department of Cell and Molecular Biology, The University of Mississippi Medical Center,  
Jackson, MS 39216-4505, USA

\*Address correspondence to:

Michael D. Hebert

Department of Cell and Molecular Biology

The University of Mississippi Medical Center

2500 North State Street

Jackson, MS 39216-4505

Tel: (601) 984-1526

FAX: (601) 984-1501

mhebert@umc.edu

Keywords: Cajal Body (CB), Small Cajal Body-specific RNA (scaRNA), microRNA (miRNA)

Summary statement: Our findings clearly establish that the Drosha/DGCR8 complex processes a subset of scaRNAs, and further strengthens the connection between CBs and the miRNA processing machinery.

## Abstract

The Cajal body (CB) is a subnuclear domain that participates in the biogenesis of many different types of ribonucleoproteins (RNPs), including small nuclear RNPs (snRNPs), small Cajal body-specific RNPs (scaRNPs) and telomerase. Most scaRNAs, the RNA component of scaRNPs, accumulate in CBs. However, there are three scaRNAs (scaRNA 2, 9, and 17) that are known to be processed into small, nucleolar-enriched fragments. Evidence suggests that these fragments are packaged into a new class of RNPs, called regulatory RNPs (regRNPs), and may modify small nucleolar RNP (snoRNP) activity, thus playing a role in rRNA modification. However, the mechanism by which these fragments are produced is unknown. Previous work has reported the involvement of Drosha and DGCR8 in the cleavage of primary-scaRNA9. Here, we expand on that knowledge by identifying sequence elements necessary for the efficient production of these RNA fragments and demonstrate that primary scaRNA 2 and 17 are also processed by the Drosha-DGCR8 complex. Collectively, our work establishes new factors in the scaRNP biogenesis pathway and adds to the ever-expanding list of noncanonical functions for the microprocessor complex.

## Introduction

The Cajal body (CB) is a subnuclear domain that participates in the biogenesis of many different types of ribonucleoproteins (RNPs), including small nuclear RNPs (snRNPs), small Cajal body-specific RNPs (scaRNPs) and telomerase (Meier, 2017). CBs are easily detected in neuronal cells and transformed cells, but are less abundant in other cell types (Sawyer et al., 2016). In cell types lacking CBs, the activities that take place in the CB occur in the nucleoplasm, but at a reduced efficiency compared to cells with CBs (Nizami et al., 2010; Sawyer et al., 2016). One of the activities that take place in the CB is the modification of the small nuclear RNA (snRNA) component of snRNPs. Like rRNA, snRNAs contain numerous modifications such as ribose methylation and pseudouridylation, and these modifications are crucial for ribosomal and snRNP function (Bohnsack and Sloan, 2018; Sloan et al., 2017a). In rRNA, ribose methylation and pseudouridylation modifications are conducted by small nucleolar RNPs (snoRNPs) (Sloan et al., 2017b). There are two classes of snoRNPs which are defined by conserved *cis* elements present in the snoRNA. Namely, box C/D snoRNAs form box C/D snoRNPs and box H/ACA snoRNAs form box H/ACA snoRNPs. Box C/D snoRNPs contain fibrillarin (conducts ribose methylation) and box H/ACA snoRNPs contain dyskerin (conducts pseudouridylation) (Burke et al., 2018; Lykke-Andersen et al., 2018). These modifications are guided by base pairing of the snoRNA to rRNA. Similarly, the snRNA component of snRNPs is modified by scaRNPs. There are three different classes of scaRNPs

defined by conserved *cis* elements present in the scaRNA: Box C/D contain fibrillarin (conducts ribose methylation), box H/ACA contain dyskerin (conducts pseudouridylation), and mixed domain contain both fibrillarin and dyskerin (conducts ribose methylation and pseudouridylation) (Burke et al., 2018). Modifications are guided by scaRNA base pairing to snRNA. An additional *cis* element present in box H/ACA scaRNAs is the CAB (Cajal body localization) motif. The CAB motif is bound by the protein WRAP53 (TCAB1/WDR79) to target this class of scaRNA, which includes the telomerase RNA component, to the CB (Izumikawa et al., 2019; Tycowski et al., 2009). It is unclear how box C/D scaRNAs, which lack a CAB motif, accumulate in the CB but TDP-43 may be involved (Izumikawa et al., 2019).

Most scaRNAs and snoRNAs are encoded in the introns of host genes. However, two scaRNAs (scaRNA2 and scaRNA17) are independently transcribed (Gerard et al., 2010; Tycowski et al., 2004). Very interestingly, scaRNA 2 and 17 can be further processed to generate fragments 70-80 nt in length (Tycowski et al., 2004). In addition to scaRNA 2 and 17, scaRNA9 has also been shown to undergo processing, generating two fragments of similar size to that observed for scaRNA 2 and 17 (Tycowski et al., 2004). These fragments mostly accumulate in the nucleolus, but are also localized in the nucleoplasm (CB) (Tycowski et al., 2004). The function of these stable fragments is unclear. It is possible that the processing of primary-scaRNA 2, 9 and 17 is a method to regulate snRNA modifications and hence govern snRNP biogenesis. For example, full-length scaRNA 2, 9 and 17 localize to the CB and there guide modifications present in U2, U4 and U12 snRNAs. The processing of primary-scaRNA 2, 9 and 17 into fragments alters the amount of full-length scaRNA 2, 9 and 17 present in the CB, which in turn would be expected to impact snRNA modification. Alternatively, it is also possible that the fraction of scaRNA 2, 9 and 17-derived fragments that accumulate in the nucleolus interact with snoRNPs and regulate their activity. In other words, some of the fragments generated from scaRNA 2, 9 and 17 may form regulatory RNPs (regRNPs), and we have published data in support of this possibility (Burke et al., 2018; Burke et al., 2019; Poole et al., 2017). Efforts to understand the mechanisms that generate scaRNA 2, 9 and 17 fragments are therefore important to both snRNP and ribosome biogenesis.

We have recently published that Drosha and DGCR8, well-characterized components of the miRNA processing pathway (Han et al., 2006; Macias et al., 2013; Nguyen et al., 2015), take part in the biogenesis of fragments derived from primary-scaRNA9 (Logan et al., 2020). In addition, we have shown that CBs can associate with a miRNA gene cluster (the Chromosome 19 MicroRNA Cluster: C19MC) and positively regulate miRNA biogenesis

(Logan et al., 2020). These findings demonstrate a functional relationship between CBs and the miRNA processing machinery. Here we report that Drosha/DGCR8 also take part in the biogenesis of fragments derived from primary-scaRNA2 and primary-scaRNA17. We also define sequence elements in primary-scaRNA 2, 9 and 17 that are important molecular determinants that regulate processing by Drosha/DGCR8. These findings clearly establish that some scaRNAs are substrates for the Drosha/DGCR8 complex, and further strengthen the connection between CBs and the miRNA processing machinery.

## Results

### Processed scaRNAs

Currently, there are three scaRNAs (scaRNA 2, 9 and 17) known to be processed which generates nucleolar enriched fragments (Tycowski et al., 2004). A schematic of these scaRNAs and fragments is shown in Figure 1A. White rectangles represent domains that are not processed into smaller fragments. Note that scaRNA 2, 9 and 17 are box C/D scaRNAs, so they guide ribose methylation modifications. Colored rectangles denote the stable stem/loop fragments derived from these primary scaRNAs, and the nomenclature indicates the site of modification in a given snRNA. For example, the mgU2-61 region of scaRNA2 (red rectangle in Figure 1A) serves as a **m**ethylation **g**uide for **U2** snRNA at position **61** (Tycowski et al., 2004). Importantly, the stem/loop scaRNA 2, 9 and 17 fragments (Figure 1A, colored rectangles) are approximately the same size (70-80 nt) (Tycowski et al., 2004) as precursor-miRNAs generated by Drosha/DGCR8, indicating that some primary-scaRNAs may be unorthodox substrates for the miRNA processing machinery. Our recent work supports this hypothesis (Logan et al., 2020). To further investigate the mechanisms that generate scaRNA 2, 9 and 17 fragments, we have generated expression constructs in the pCDNA3.1 vector which yield sufficient amounts of ectopically expressed full-length (FL) and fragment (frag) that can be easily detected by Northern blotting using specific DIG-labelled probes (Figure 1B). In contrast, endogenous scaRNA 2, 9 and 17 fragments are more difficult to detect using non-radioactive probes. Although ectopically expressed scaRNA 2, 9 and 17 can yield the appropriate processed fragments (Figure 1B), it is important to appreciate that these plasmid constructs do not fully recapitulate endogenous expression conditions. In particular, important *cis* elements that promote or facilitate processing may be lacking in these constructs. Altered expression conditions as a result of ectopic scaRNA 2, 9 and 17 transcription from a pCDNA3.1 vector (Figure 1A) may account for the observed differences in the amount of processed fragment present for each scaRNA. For example, fragments derived from scaRNA9 (mgU2-19 and mgU2-30) are easily detected whereas the amount of fragment

derived from scaRNA2 (mgU2-61) is relatively less abundant considering the high amount of FL scaRNA2 signal (Figure 1B). It is possible that scaRNA9 fragments are more efficiently generated because scaRNA9 is present within an intron and splicing of this ectopically expressed pre-mRNA allows for scaRNA9 to be appropriately targeted to the Drosha/DGCR8 processing machinery. In contrast, scaRNA2 and scaRNA17 expression from pCDNA3.1 creates a hybrid mRNA consisting of various elements, including a poly A tail, that are not present in endogenous full-length scaRNA 2 and 17. We speculate that these scaRNA 2 and 17 hybrid mRNAs are not efficiently targeted to the Drosha/DGCR8 processing machinery. Inefficient Drosha/DGCR8 processing of ectopically expressed scaRNA 2, 9 and 17 may also result in the accumulation of primary scaRNAs or intermediates that are larger in size than the full-length scaRNAs (Figure 1B).

### **Downstream sequences are important for scaRNA 2 and 17 fragment biogenesis**

It is known that, in general, the microprocessor complex identifies the basal junction between single stranded RNA and double stranded RNA in primary-miRNA as a target for processing (Han et al., 2006). Additionally, there are many conserved motifs present in primary-miRNAs that promote processing by Drosha/DGCR8 (Auyeung et al., 2013). To evaluate the importance of single stranded RNA downstream of the processed fragments in scaRNA 2 and 17, we generated expression constructs lacking sequence downstream of the end of the fragments (Figure 2A). These constructs were transfected into cells and RNA was isolated 24 hrs later. Northern blotting of isolated RNA showed that deletion constructs had a reduced amount of scaRNA2 and scaRNA17 fragment detected relative to full-length scaRNA compared to that obtained with wild-type (WT) constructs (Figure 2B, Histograms). These findings demonstrate that sequences downstream of the fragment region of scaRNA 2 and 17 impact fragment processing.

### **The scaRNA9 intronic context influences mgU2-30 fragment levels**

We next investigated the upstream and downstream sequences of scaRNA9 that may impact the biogenesis of the mgU2-30 fragment. Unlike scaRNA2 and scaRNA17, which are independently transcribed, scaRNA9 is encoded in the intron of the *CEP295* host gene (Figure 1A, Figure 3A). Gene expression analysis from GTEx (Release V6) which includes 53 tissues (University of California, Santa Cruz, Genome Browser) shows that the *CEP295* host gene is uniformly expressed in most tissues examined. An approximate 2-fold induction of *CEP295* expression in cerebellar hemisphere, cerebellum and testis was noted relative to most other tissues examined. All of these tissues, therefore, would be expected to produce

intronicly-encoded scaRNA9. Previous work has shown that miRNA processing of intronicly expressed primary-miRNAs can take place before splicing (Janas et al., 2011; Kim and Kim, 2007). Indeed, introns in which the pre-miRNA has been removed by Drosha/DGCR8 are still subjected to splicing (Kim and Kim, 2007). In other words, the intron encoded primary-miRNA does not need to be excised by the spliceosome before miRNA processing. To determine if the intronic context of scaRNA9 is important for the biogenesis of the mgU2-30 fragment, a DNA expression construct was engineered that deletes the intronic sequence upstream of the scaRNA9 coding sequence, leaving the intronic sequence downstream of the coding region intact (3' ext) (Figure 3A). Additionally, we also generated a construct that deletes intronic sequences upstream and downstream of the scaRNA9 coding sequence ( $\Delta$ ) (Figure 3A). Expression of WT and mutant scaRNA9 constructs in cells, followed by RNA isolation and Northern blotting with a mgU2-30 probe reveals that relatively little mgU2-30 fragment is generated from the upstream and downstream deletion construct ( $\Delta$ ) (Figure 3B, compare the amount of the mgU2-30 signal in lane 2 to that in lane 4). The amount of mgU2-30 fragment is also reduced compared to WT for the construct lacking sequences upstream of the scaRNA9 coding sequence (3' ext), although we note that the expression of the primary-scaRNA9-3' ext transcript is less abundant than the other primary transcripts tested. This finding may indicate that the primary-scaRNA9-3' ext transcript is less stable than the ectopically expressed WT and ( $\Delta$ ) primary transcripts. The relative lack of mgU2-30 fragment with the  $\Delta$  construct despite the presence of a sizable amount of the primary-scaRNA9- $\Delta$  transcript clearly demonstrates that intronic sequences upstream and downstream of the scaRNA9 coding region are necessary for efficient biogenesis of the mgU2-30 fragment (Figure 3B, histogram).

### ***In vitro* processing of scaRNA2 and scaRNA17 by Drosha/DGCR8**

We have published that Drosha/DGCR8 complexes can process a primary-scaRNA9 substrate *in vitro*, generating both mgU2-19 and mgU2-30 fragments (Logan et al., 2020). We have now examined if Drosha/DGCR8 can generate the mgU2-61 fragment using primary-scaRNA2 and the mgU4-8 fragment using primary-scaRNA17 *in vitro* transcribed substrates in an *in vitro* processing assay (Figure 4A). As shown in Figure 4B, processing assays using primary-scaRNA9 substrate incubated with FLAG beads complexed with FLAG-DGCR8/Drosha generate a fragment consistent with mgU2-30 (left panel, lanes 4 and 5), consistent with our previously published results (Logan et al., 2020). A mgU2-30 fragment is not observed when using FLAG beads complexed with lysate from un-transfected cells (left panel, lane 3), or incubated with water alone (lane 2). Full-length scaRNA9 and mgU2-30

fragment from ectopically expressed primary-scaRNA9 is shown in lane 6. Processing assays using FLAG beads from un-transfected cells (UT) and FLAG-DGCR8/Drosha transfected cells (Trans) were also conducted using primary-scaRNA2 (middle panel) and primary-scaRNA17 (right panel). As with primary scaRNA9 (Logan et al., 2020) (Figure 4B, left panel), fragments corresponding to mgU2-61 (middle panel) and mgU4-8 (right panel) are detected only in reactions using FLAG beads complexed with Drosha/DGCR8 (Trans) and not with FLAG beads from un-transfected cells (UT) or incubated with water alone. These *in vitro* generated fragments are approximately the same size as found in cells ectopically expressing scaRNA2 and scaRNA17 constructs (lane 5, middle and right panel, respectively). Hence Drosha/DGCR8 can directly generate fragments from primary-scaRNA 2 and 17 *in vitro*, similar to our published report showing that the microprocessor complex can also generate scaRNA9 fragments (Logan et al., 2020).

### **Identification of a mutant in scaRNA9 that disrupts *in vivo* mgU2-30 biogenesis**

Secondary structure predictions of scaRNA9 indicate that the mgU2-19 and mgU2-30 regions form stem/loop structures (Tycowski et al., 2004). The unpaired loop region of these structures is predicted to serve as the guide RNA component in the scaRNP for snRNA modification, in accordance with generally accepted rules as to the function of box H/ACA and box C/D snoRNPs and scaRNPs (Kiss, 2004). To disrupt this stem/loop structure, we have generated a construct with an 8 nt insertion in the C' motif (Figure 5A, mutant #2) and tested if this mutation would alter mgU2-30 biogenesis in cells and by *in vitro* processing assays. These constructs were then transfected into cells, followed by RNA isolation, Northern blotting and hybridization with a probe that detects FL scaRNA9 and the mgU2-30 fragment. As shown in Figure 5B, the amount of the mgU2-30 fragment is reduced relative to FL scaRNA9 when using mutant #2 compared to the fragment/FL ratio observed with the WT scaRNA9 construct. As expected, the mgU2-30 fragment derived from the mutant #2 construct, which contains an 8 nt insertion, migrates slightly slower than the WT mgU2-30 fragment. Quantification of this and other data is shown in Figure 5C. *In vitro* processing assays, however, show that Drosha/DGCR8 complexes are still able to generate the mgU2-30 fragment with the mutant #2 substrate (Figure 5D).

## Mutation of the downstream region of primary-scaRNA9 decreases *in vivo* mgU2-30 fragment biogenesis

It is known that many *cis* elements influence Drosha/DGCR8 primary-miRNA processing efficiency. One of these elements is the CNNC which is located downstream of the basal junction (Creugny et al., 2018). Primary-scaRNA9 has three CNNC motifs in the single stranded region downstream of the mgU2-30 region (Figure 6A). To monitor if these CNNC motifs impact mgU2-30 biogenesis, we generated three different mutants that progressively disrupt all of the CNNC motifs (mutant #3, mutant #4 and mutant #5 in Figure 6A). These constructs were then transfected into cells, followed by RNA isolation, Northern blotting, and hybridization with a probe that detects FL scaRNA9 and the mgU2-30 fragment to monitor the *in vivo* processing of these mutated RNAs. As shown in Figure 6B, and quantified in Figure 6C, progressive mutation of the scaRNA9 3' region containing the CNNC motifs decreases the efficiency of mgU2-30 fragment production or stability. Notably, mutant #4, which has the same length of scaRNA9 downstream region as WT scaRNA9 but has mutations in all three CNNC motifs, has a statistically significant reduction in the amount of mgU2-30 fragment relative to FL scaRNA9 compared to that obtained with WT scaRNA9. *In vitro* assays demonstrate that all of the mutants in the scaRNA9 3' region can still be processed by Drosha/DGCR8 with no dramatic loss of efficiency (Figure 6D). Collectively, the results of these experiments demonstrate that the CNNC motifs present in the scaRNA9 3' region impact the *in vivo* biogenesis of mgU2-30. It is likely that the lack of specificity factors in the *in vitro* processing assay account for the observation that scaRNA9 3' region mutants are still effectively processed by this method, as detailed in the discussion below.

## Discussion

The data presented here expand the number of scaRNAs known to be processed by the Drosha/DGCR8 complex. Our previous work demonstrates that primary-scaRNA9 is a substrate for Drosha/DGCR8 (Logan et al., 2020), and we show here that primary-scaRNA2 and primary-scaRNA17 can also be processed by this complex. Specifically, we show by *in vitro* processing assays that the biogenesis of previously reported scaRNA 2, 9 and 17 fragments (Tycowski et al., 2004) can be achieved by the Drosha/DGCR8 complex. In addition, we identify sequence elements that facilitate scaRNA fragment generation. For example, an important determinant for processing by Drosha/DGCR8 is a single stranded RNA region upstream and downstream of the double stranded RNA region (Creugny et al., 2018).



This was observed in our previous study, which showed that the mgU2-30 fragment is not efficiently generated in an *in vitro* processing assay when using a scaRNA9 3' deletion substrate (Logan et al., 2020). Here we report that the 3' region of both scaRNA2 and scaRNA17, which resides downstream of the processed fragment location, positively influences fragment formation. A previous report (Gerard et al., 2010) also examined the 3' region of scaRNA2, but this study focused on the role of 3' region in the formation of full-length scaRNA2 and not the mgU2-61 fragment. Specifically, this report showed that the 3' region downstream of the end of the scaRNA2 coding region does not impact the formation of full-length scaRNA2 (Gerard et al., 2010). In contrast, we show here that the 3' region of scaRNA2 does play a role in the processing of primary-scaRNA2 and generation of the mgU2-61 fragment.

We also demonstrate that the 3' region of scaRNA9 contains CNNC motifs, which are associated with increased processing of primary-miRNAs by Drosha/DGCR8. Mutation of these motifs in primary-scaRNA9 reduces the amount of the mgU2-30 fragment as assayed by *in vivo* processing experiments. Reduced levels of the mgU2-30 fragment were also observed *in vivo* when an 8 nt insertion was placed in the scaRNA9 C' box, clearly indicating that the Drosha/DGCR8 complex requires an optimal configuration of single stranded RNA regions upstream and downstream of a double stranded RNA region. Interestingly, *in vitro* processing assays using scaRNA9 mutants #2, #4, and #5 do not show a large reduction in the amount of mgU2-30 fragment production as observed in the *in vivo* processing assays. One possible explanation for the difference between our *in vivo* and *in vitro* processing assays is that a specificity factor is present in the *in vivo* assays which is not part of the Drosha/DGCR8 immuno-complex used in the *in vitro* assays. One such specificity factor could be SRSF3 (SRP20), a splicing factor that has been shown to play a role in the identification of the CNNC motif in primary-miRNAs to help facilitate processing by Drosha (Auyeung et al., 2013; Kim et al., 2018). SRSF3 is not part of the Drosha/DGCR8 immuno-complex, however (Auyeung et al., 2013), so it is likely that DGCR8 can serve in the *in vitro* processing assays as an anchor by binding to the RNA substrate to direct Drosha processing. In contrast, in the environmentally more complex *in vivo* processing assays, SRSF3 likely ensures specificity of Drosha/DGCR8 activity. Collectively, the mutational analysis conducted here expands the known *cis* elements that impact scaRNA 2, 9 and 17 fragment generation, which now includes the leader sequence of scaRNA9, the GU-rich regions in scaRNA 2 and 9, (Enwerem et al., 2015; Poole et al., 2016; Poole et al., 2017), and the 3' region of scaRNA 2, 9 and 17. Interestingly, the scaRNA9 leader sequence is important for full-length scaRNA9

and mgU2-19 fragment accumulation, but does not impact mgU2-30 fragment levels (Poole et al., 2017). This finding indicates that Drosha/DGCR8 processing of mgU2-30 is not dependent upon mgU2-19 processing.

An important next step will be to assess how the cell regulates the amount of primary-scaRNA 2, 9 and 17 that used to generate full-length scaRNAs that form scaRNPs and localize to the CB versus that which is processed by Drosha/DGCR8 and form RNPs that are primarily nucleolar in localization. Conceivably, the regulation of this cellular decision point will have ramifications for both snRNP biogenesis and rRNA modification. As such, additional studies into the processing of scaRNAs by Drosha/DGCR8 will clarify regulatory mechanisms that impact both the splicing and translation machinery.

## Materials and Methods

### Cell lines, plasmids, and transfections

HeLa cells were obtained from the American Type Culture Collection (ATCC). Cells were cultured as previously described (Enwerem et al., 2014). ScaRNA2 pcDNA3.1+ (Enwerem et al., 2014), scaRNA9 pcDNA3.1+ (Enwerem et al., 2015), and scaRNA17 pcDNA3.1+ (Poole et al., 2017) have been previously described and are represented by  $\Delta$  in this manuscript. For wild-type (WT) or 3' ext constructs, scaRNAs were amplified from genomic DNA isolated from HeLa-ATCC using standard molecular biology techniques and cloned into either pBluescript KS+ for *in vitro* transcription or pcDNA3.1+ for cell transfection. ScaRNA9 mutant constructs were obtained by performing site-directed mutagenesis of either pcDNA3.1+ or pBluescript KS+ scaRNA9 WT using the Q5 Site-Directed Mutagenesis Kit (New England BioLabs, Ipswich, MA, USA) following the manufacturer's protocol. Primers, obtained from Integrated DNA Technologies (Coralville, IA, USA), and restrictions sites used in the cloning and mutagenesis procedures are detailed below. Note that endogenous scaRNA fragments are only faintly visible (Poole et al., 2016), thus the majority of the detected full-length and fragment scaRNA is from the processing of ectopically expressed primary-scaRNA. DNA transfections were conducted using FuGene HD (Promega, Madison, USA) according to the manufacturer's protocol.

## Cloning and mutagenesis primers

To clone WT scaRNA2, Forward 5'-GGCGGATCCGTTTTAGGGAGGGAGAGCGGCCTG-3' and Reverse 5'-TATGGAATTCGGAAAGTGGGAGGAAAGTATATC-3' primers were used. The underlined sequences denote a *Bam*HI and an *Eco*RI site, respectively. To clone WT scaRNA9, Forward 5'-GCCGAATTCCTTAAGTTATGCTGTGGAGGAAG-3' and Reverse 5'-GGCGCGGCCGCTTTCATAACTTAAAAGGCTCC-3' primers were used. The underlined sequences denote an *Eco*RI and a *Not*I site, respectively. To clone scaRNA9 3' ext, Forward 5'-GGCGGATCCCTTTCTGAGATCTGCTTTTAGTGA-3' and Reverse 5'-GGCGAATTC AACAGTTGCTGAAGATAATGG-3' primers were used. The underlined sequences denote a *Bam*HI and an *Eco*RI site, respectively. To clone WT scaRNA17, Forward 5'-TATGGGATCCAGAGGCTTGGGCCGCCGAGCT-3' and Reverse 5'-TATGGAATTCGTGTTTGAAAAGCAGGATTCTAG-3' primers were used. The underlined sequences denote a *Bam*HI and an *Eco*RI site, respectively. The following primers were used for scaRNA9 mutagenesis: Mutant #2, Forward 5'-ATGGGTTTCTACACTTGACCTG-3' and Reverse 5'-GCTCAATAGTTACAAAGATCAGTAGTAAAC-3'; Mutant #3, Forward 5'-TGCTGGGGTTGGTGATTT-3' and Reverse 5'-AAGCTTTCACCTTCTGAGCTCAGGTC-3'; Mutant #4, Forward 5'-GGATCCGTGGTATGATATTCCCCTTAATTTTC-3' and Reverse 5'-AAAACCCCAAATCACCAACC-3'; Mutant #5, Forward 5'-ATATTCCCCTTAATTTCTAGGC-3' and Reverse 5'-AAGCTTTCACCTTCTGAGCTCAGGTC-3'.

## *In vitro* Drosha/DGCR8 Processing Assay

*In vitro* processing assays were conducted using Drosha/DGCR8 immunoprecipitates. Myc-tagged Drosha and FLAG-tagged DGCR8 plasmids were obtained from Addgene (Watertown, MA, USA). HeLa cells were co-transfected with myc-Drosha and FLAG-DGCR8 DNA for 24 hours. After 24 hours, cells were washed with PBS and harvested using a KCl lysis buffer (20mM Tris-HCl, 180mM KCl, 0.2mM EDTA) followed by sonication with a Fisher Scientific sonic dismembrator (model 100) for 6 times for 5 seconds each using the output setting of 1, and finally centrifuged at 12000 rpm for 15 min at 4°C. The FLAG-DGCR8/Drosha complex was immunoprecipitated using anti-FLAG-M2 affinity agarose beads (Sigma Aldrich, St. Louis, MO, USA) overnight and washed 3-5 times with KCl lysis buffer. The precipitated complex was then incubated for 90 min at 37°C with RNA substrate, along with 6.4mM MgCl<sub>2</sub>, RNase Inhibitor (ThermoFisher, Waltham, MA, USA), and RNase-free water in a total volume of 30µL.

Control reactions include substrate incubated with water alone and substrate incubated with FLAG beads from non-transfected cell lysate. The RNA was then isolated using Tri-Reagent (Molecular Research Center, Cincinnati, OH, USA) following the manufacturer's protocol and subjected to Northern Blotting using 5' DIG probes specific for full-length scaRNA and the processed fragment, previously described (Poole et al., 2017), as follows: scaRNA9 mgU2-30 (5'-TAGAAACCATCATAGTTACAAAGATCAGTAGTAAACCTTTTCATCATTGCC-3') and mgU2-19 (5'-GTAGACTGGAAAGACTTCTGATGCTCAGATTTGGCTAGTTTCATCATTGA-3'); scaRNA2 mgU2-61 (5'-AGTGGCCGGGGACAAGCCCGGCCTCGTCTATCTGATCAATTCATCACTTCT-3'); scaRNA17 mgU4-8 (5'-AACTCAGATTGCGCAGTGGTCTCGTCATCA-3'). DIG labeled probes were obtained from Integrated DNA Technologies (Coralville, IA, USA).

RNA substrate used in the *in vitro* processing assay was generated by the following method. First, scaRNA constructs in pBS KS+ were linearized at the 3' end with *EcoRI* (for scaRNA2 and scaRNA17) or *HindIII* (for scaRNA9) and then gel purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Linearized DNA was subjected to *in vitro* transcription with the Megascript T7 Transcription Kit (Thermo Fisher, Waltham, MA, USA). Lastly, the *in vitro* generated RNA was gel purified using the ZR-small RNA PAGE Recovery Kit (ZYMO Research, Carlsbad, CA, USA).

### Northern Blotting

Total RNA was isolated using TRI-Reagent (Molecular Research Center, Cincinnati, OH, USA). For total RNA, typically 10-15 µg was run on a 6% denaturing polyacrylamide gel (Invitrogen, Carlsbad, CA, USA) in 1X Tris-Borate-EDTA (TBE) at 200V. The gel was then washed in 1X TBE and then transferred onto a positively charged nylon membrane (Invitrogen, Carlsbad, CA, USA) with the iBlot Gel Transfer device (Life Technologies, Grand Island, NY, USA) using program 5 for 5 minutes. After transfer, the membrane was rinsed in ultrapure water, allowed to dry, and then subjected to a UV cross-linker (UVP, Upland, CA) at a setting of 120,000µJ/cm<sup>2</sup>. The membrane was then placed in a hybridization bottle and pre-hybridized using UltraHyb Ultrasensitive Hybridization buffer (Ambion Life Technologies, Grand Island, NY, USA) for 30 minutes at 42°C in a hybridization oven. The DNA oligo probes used for the detection of full-length and fragment scaRNAs 2, 9, and 17 were 5' DIG labeled and previously described (Poole et al., 2017). Membranes were then prepared for detection using the DIG Wash and Block kit (Invitrogen, Carlsbad, CA, USA) following the

manufacturer's suggested protocol with the Anti-DIG antibody used at 1:10,000. Detection was carried out using CSPD (Roche, Mannheim, Germany) following the manufacturer's suggested protocol. Blots were imaged using a Chemidoc imager (Bio-Rad, Hercules, CA, USA). Where noted, adjustments to images were made using the transformation settings on QuantityOne software and applied across the entire image.

### **Acknowledgements**

Funding for this work was provided by the Intramural Research Support Program of The University of Mississippi Medical Center.

## References

- Auyeung, V. C., Ulitsky, I., McGeary, S. E. and Bartel, D. P.** (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**, 844-58.
- Bohnsack, M. T. and Sloan, K. E.** (2018). Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol Chem* **399**, 1265-1276.
- Burke, M. F., Logan, M. K. and Hebert, M. D.** (2018). Identification of additional regulatory RNPs that impact rRNA and U6 snRNA methylation. *Biol Open* **7**.
- Burke, M. F., McLaurin, D. M., Logan, M. K. and Hebert, M. D.** (2019). Alteration of 28S rRNA 2'-O-methylation by etoposide correlates with decreased SMN phosphorylation and reduced Drosha levels. *Biol Open* **8**.
- Creugny, A., Fender, A. and Pfeffer, S.** (2018). Regulation of primary microRNA processing. *FEBS Lett* **592**, 1980-1996.
- Enwerem, II, Wu, G., Yu, Y. T. and Hebert, M. D.** (2015). Cajal body proteins differentially affect the processing of box C/D scaRNPs. *PLoS One* **10**, e0122348.
- Enwerem, I. I., Velma, V., Broome, H. J., Kuna, M., Begum, R. A. and Hebert, M. D.** (2014). Coilin association with Box C/D scaRNA suggests a direct role for the Cajal body marker protein in scaRNP biogenesis. *Biology Open* **3**, 240-249.
- Gerard, M. A., Myslinski, E., Chylak, N., Baudrey, S., Krol, A. and Carbon, P.** (2010). The scaRNA2 is produced by an independent transcription unit and its processing is directed by the encoding region. *Nucleic Acids Res* **38**, 370-81.
- Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T. and Kim, V. N.** (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887-901.
- Izumikawa, K., Nobe, Y., Ishikawa, H., Yamauchi, Y., Taoka, M., Sato, K., Nakayama, H., Simpson, R. J., Isobe, T. and Takahashi, N.** (2019). TDP-43 regulates site-specific 2'-O-methylation of U1 and U2 snRNAs via controlling the Cajal body localization of a subset of C/D scaRNAs. *Nucleic Acids Res* **47**, 2487-2505.
- Janas, M. M., Khaled, M., Schubert, S., Bernstein, J. G., Golan, D., Veguilla, R. A., Fisher, D. E., Shomron, N., Levy, C. and Novina, C. D.** (2011). Feed-forward microprocessing and splicing activities at a microRNA-containing intron. *PLoS Genet* **7**, e1002330.
- Kim, K., Nguyen, T. D., Li, S. and Nguyen, T. A.** (2018). SRSF3 recruits DROSHA to the basal junction of primary microRNAs. *Rna* **24**, 892-898.

**Kim, Y. K. and Kim, V. N.** (2007). Processing of intronic microRNAs. *Embo J* **26**, 775-83.

**Kiss, T.** (2004). Biogenesis of small nuclear RNPs. *J Cell Sci* **117**, 5949-51.

**Logan, M. K., McLaurin, D. M. and Hebert, M. D.** (2020). Synergistic interactions between Cajal bodies and the miRNA processing machinery. *Mol Biol Cell*, mbcE20020144.

**Lykke-Andersen, S., Ardal, B. K., Hollensen, A. K., Damgaard, C. K. and Jensen, T. H.** (2018). Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-mRNA. *Mol Cell* **72**, 99-111.e5.

**Macias, S., Cordiner, R. A. and Caceres, J. F.** (2013). Cellular functions of the microprocessor. *Biochem Soc Trans* **41**, 838-43.

**Meier, U. T.** (2017). RNA modification in Cajal bodies. *RNA Biol* **14**, 693-700.

**Nguyen, D., Grenier St-Sauveur, V., Bergeron, D., Dupuis-Sandoval, F., Scott, M. S. and Bachand, F.** (2015). A Polyadenylation-Dependent 3' End Maturation Pathway Is Required for the Synthesis of the Human Telomerase RNA. *Cell Rep* **13**, 2244-57.

**Nizami, Z., Deryusheva, S. and Gall, J. G.** (2010). The Cajal body and histone locus body. *Cold Spring Harbor perspectives in biology* **2**, a000653-a000653.

**Poole, A. R., Enwerem, II, Vicino, I. A., Coole, J. B., Smith, S. V. and Hebert, M. D.** (2016). Identification of processing elements and interactors implicate SMN, coilin and the pseudogene-encoded coilp1 in telomerase and box C/D scaRNP biogenesis. *RNA Biol* **13**, 955-972.

**Poole, A. R., Vicino, I., Adachi, H., Yu, Y. T. and Hebert, M. D.** (2017). Regulatory RNPs: a novel class of ribonucleoproteins that potentially contribute to ribosome heterogeneity. *Biol Open* **6**, 1342-1354.

**Sawyer, I. A., Sturgill, D., Sung, M. H., Hager, G. L. and Dundr, M.** (2016). Cajal body function in genome organization and transcriptome diversity. *Bioessays* **38**, 1197-1208.

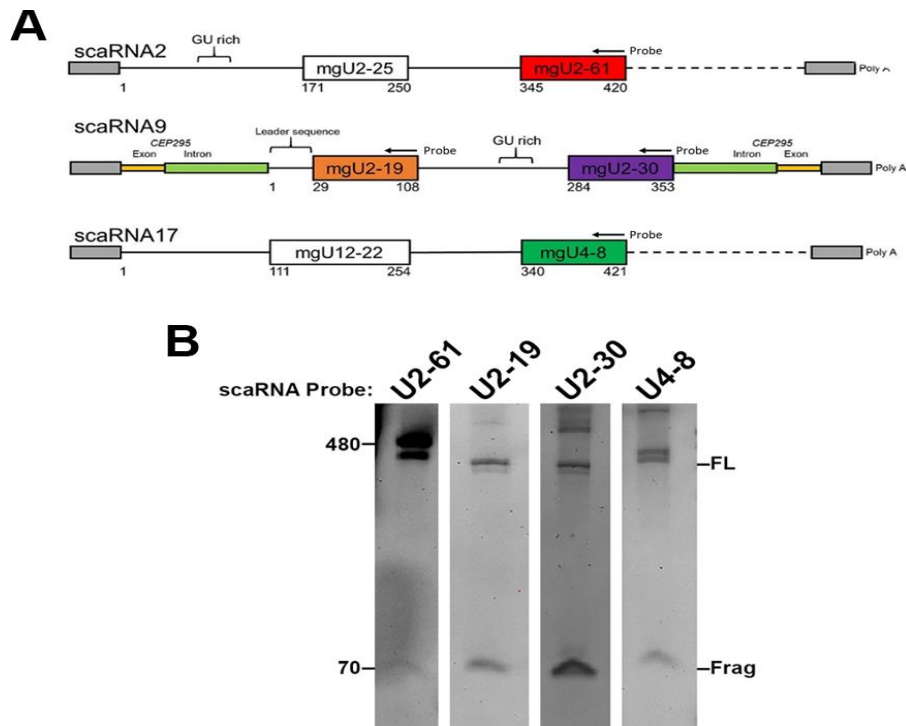
**Sloan, K. E., Warda, A. S., Sharma, S., Entian, K.-D., Lafontaine, D. L. J. and Bohnsack, M. T.** (2017a). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**, 1138-1152.

**Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J. and Bohnsack, M. T.** (2017b). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**, 1138-1152.

**Tycowski, K. T., Aab, A. and Steitz, J. A.** (2004). Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr Biol* **14**, 1985-95.

**Tycowski, K. T., Shu, M.-D., Kukoyi, A. and Steitz, J. A.** (2009). A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Molecular cell* **34**, 47-57.

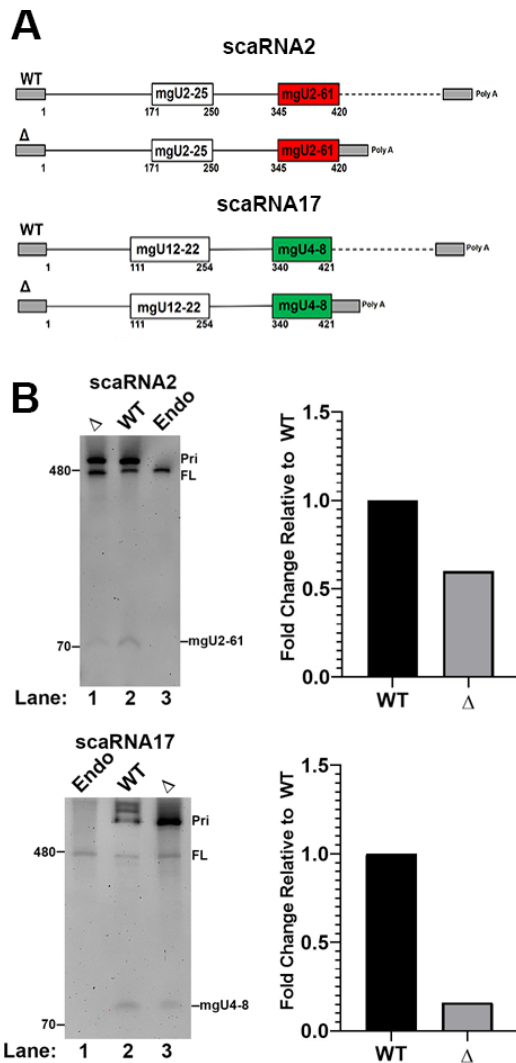
## Figures



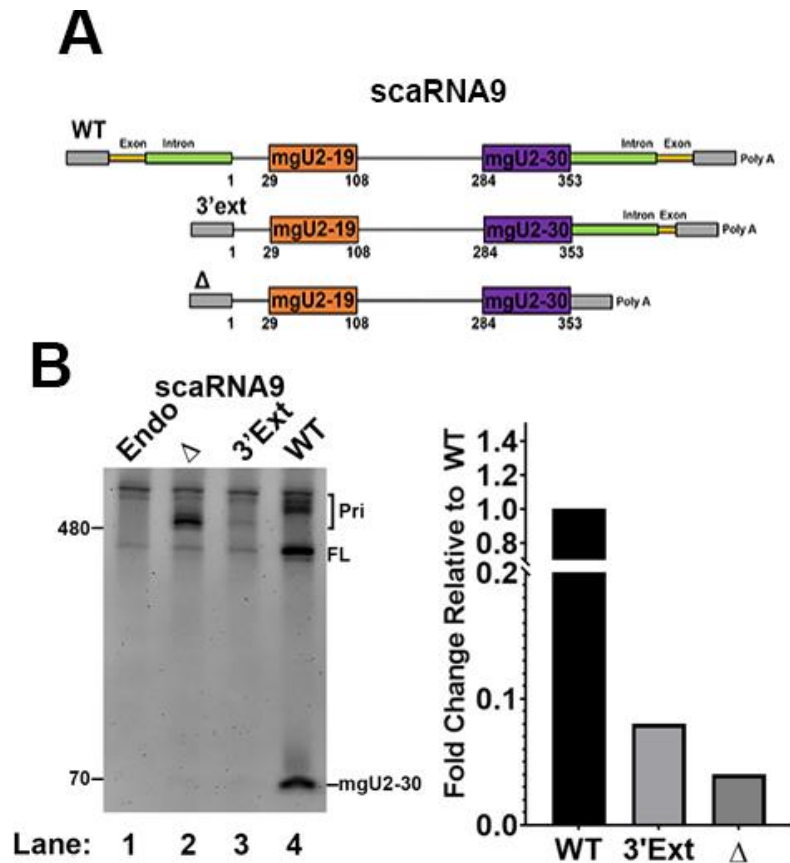
**Figure 1. ScaRNA 2, 9, and 17 expression constructs and processing.** (A) Schematic of expression constructs used to evaluate *in vivo* processing of scaRNA 2, 9, and 17. Sequences encoding these scaRNAs were cloned into pcDNA3.1+. Transcription will generate RNAs that include vector sequences at the 5' and 3' ends of the RNA (gray rectangles), including the addition of a poly A tail. The solid black line represents the endogenous full-length (FL) scaRNA detectable by Northern blotting. The dashed line represents sequences downstream of the FL scaRNA that are important for processing. The colored rectangles denote processed endogenous fragments for each scaRNA that are detectable by Northern blotting. In contrast to scaRNA 2 and 17, which are independently transcribed in the genome, scaRNA9 is encoded in the intron of the *CEP295* host gene. Our pcDNA3.1+scaRNA9 expression construct contains the entire *CEP295* intron that harbors scaRNA9 (green rectangles), along with partial exonic sequences upstream and downstream of this intron (gold rectangles), ensuring proper splicing of the transcribed RNA. The probes used for Northern blotting are indicated. Other notable sequence elements, such as the leader sequence in scaRNA9 and the GU rich regions in scaRNA2 and scaRNA9 (Enwerem et al., 2015; Poole et al., 2016; Poole et al., 2017), are denoted. (B) Detection of ectopically expressed FL and processed fragments by Northern blotting. HeLa cells were transfected with the DNA expression constructions described above, followed by RNA isolation, TBE-



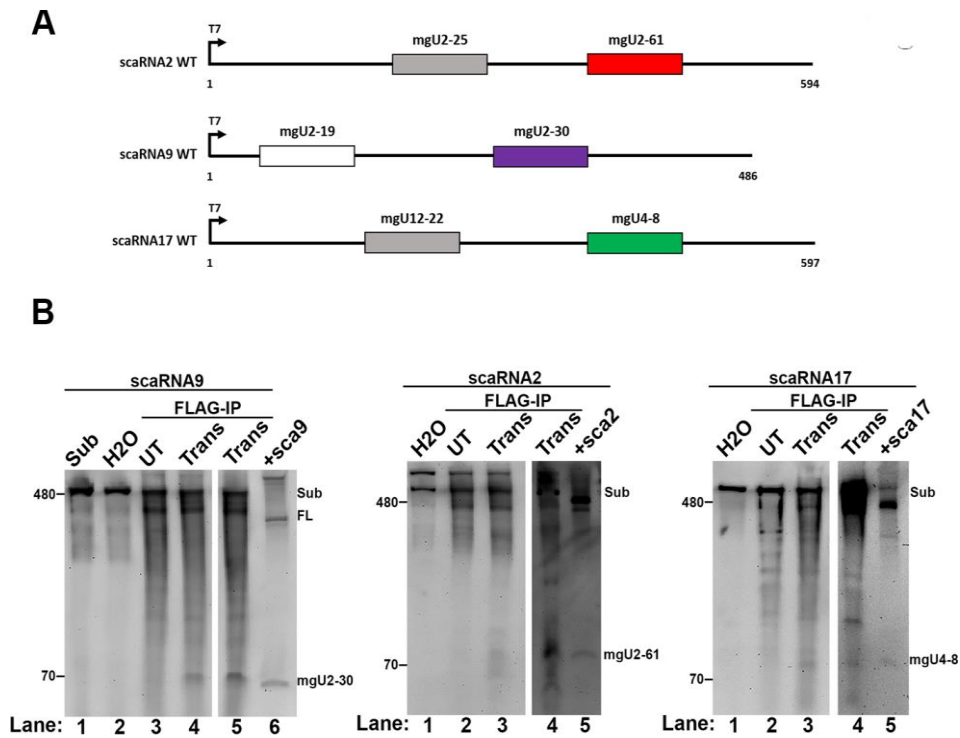
Urea gel electrophoresis, Northern blotting and hybridization with probes that bind the indicated region of the scaRNA. The detected signals corresponding to ectopically expressed FL or fragment (frag) for each scaRNA are noted.



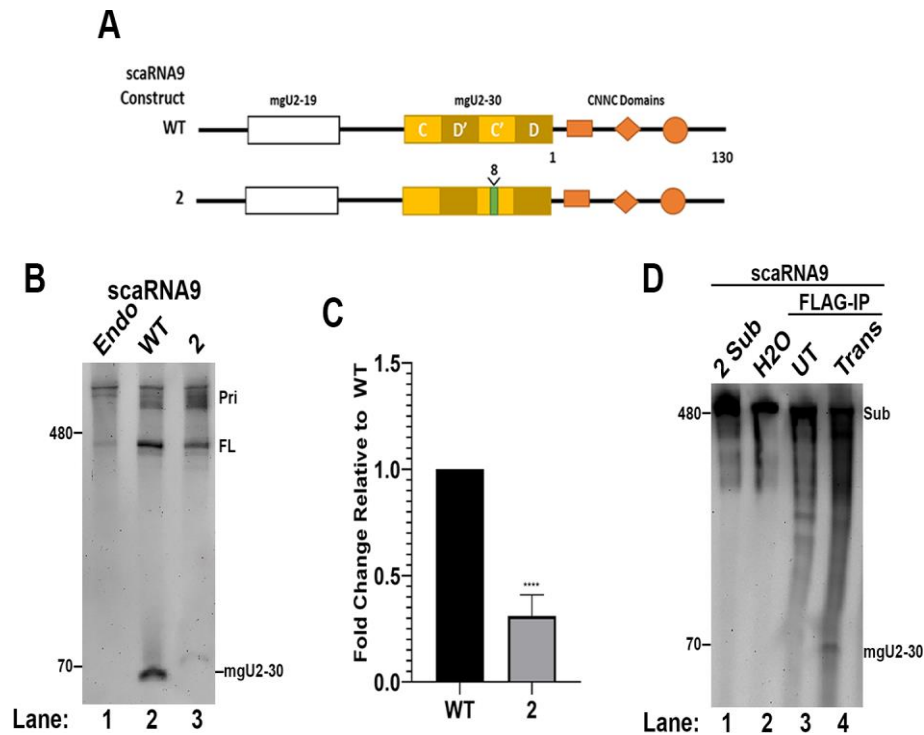
**Figure 2. An extended 3' sequence is required for the efficient generation of scaRNA fragments.** DNA constructs expressing wild-type (WT) or 3' deletion ( $\Delta$ ) scaRNA2 or scaRNA17 were transfected into HeLa cells followed by RNA isolation, Northern blotting and detection with probes described in Figure 1. Full-length (FL) scaRNA 2 and 17 and fragments thereof are indicated. Endogenous signals are present in lanes containing RNA from untransfected cells (Endo). Primary-scaRNA signal is denoted (Pri). The data shown in the gels was quantified and displayed in histograms. Specifically, the fragment signal was divided by the FL + Pri signal and the WT value was set to 1. Both the FL + Pri signals were included in the calculation in order to account for all ectopic expression products.



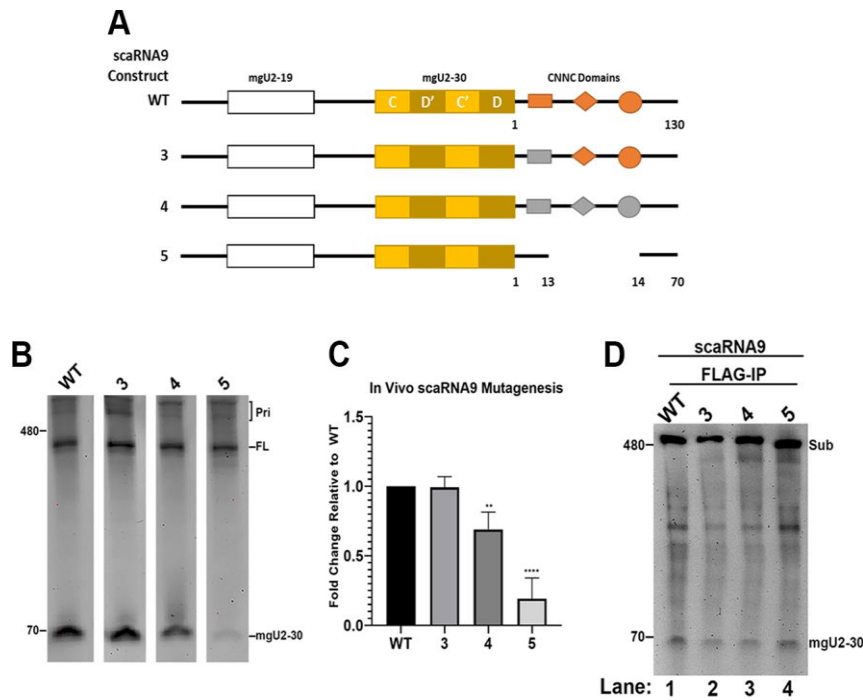
**Figure 3. Intronic sequences are required for the efficient processing of scaRNA9.** (A) Schematic of scaRNA9 constructs illustrating the absence of 5' intronic sequence in the 3' extension construct (3' ext) and the lack of both 5' and 3' intronic sequences in the deletion construct ( $\Delta$ ). (B) Detection of full-length (FL) scaRNA9 and the mgU2-30 fragment by Northern blot utilizing RNA from untransfected HeLa cells (Endo) or HeLa cells ectopically-expressing WT, 3' ext, or  $\Delta$  scaRNA9. Primary-scaRNA signal is denoted (Pri). The data shown in the gel was quantified (histogram). Specifically, the fragment signal was divided by the FL + Pri signal and the WT value was set to 1. Both the FL + Pri signals were included in the calculation in order to account for all ectopic expression products.



**Figure 4. *In vitro* processing of scaRNAs by Drosha/DGCR8.** (A) Schematic of scaRNA 2, 9, and 17 substrates used for *in vitro* processing assays with immunoprecipitated Drosha/DGCR8. Substrates were obtained by *in vitro* transcription of linearized DNA encoding the indicated scaRNA followed by RNA gel purification. (B) *In vitro* transcribed scaRNA9 (left), scaRNA2 (middle) and scaRNA17 (right) substrates were used in a processing assay with immunopurified (FLAG) Drosha/DGCR8. RNA isolated from the processing reactions was subjected to Northern blotting with probes that detect the various full-length scaRNAs and fragments. Substrate was incubated with FLAG beads from Drosha/FLAG-DGCR8 transfected lysate (Trans), FLAG beads from un-transfected cell lysate (UT), or water (H<sub>2</sub>O). Note that two different Trans processing reactions are shown for each substrate (lanes 4 and 5 for scaRNA9 and lanes 3 and 4 for scaRNA2 and scaRNA17). A positive control of ectopically expressed scaRNA9 RNA is shown in the left panel, lane 6 (+sca9). Ectopically expressed scaRNA2 and scaRNA17 are shown in lane 5 of the middle (+sca2) and right (+sca17) panels, respectively.



**Figure 5. Disruption of the scaRNA9 C' box impedes *in vivo* biogenesis of the mgU2-30 fragment.** (A) Schematic of WT and mutant scaRNA9 constructs. The 8 nt insertion in the C' box of scaRNA9 mutant 2 is shown. Also indicated are CNNC motifs that are known to impact Drosha/DGCR8 processing. (B) Northern blot detection of the mgU2-30 fragment and FL scaRNA9 in RNA isolated from HeLa cells that ectopically express the WT or mutant #2 scaRNA9 constructs. Primary-scaRNA signal is denoted (Pri). (C) Histogram generated from the quantification of data shown in (B). The mgU2-30 signal was divided by the FL scaRNA9 signal in that lane, and normalized to that obtained for WT. N = 4 biological repeats, \*\*\*\* =  $p < 0.0001$ , error bars represent standard deviation. (D) Northern blot detection of RNA isolated from *in vitro* processing assays with Drosha/DGCR8 and scaRNA9 mutant #2 as the substrate.



**Figure 6. The CNNC domains impact *in vivo* mgU2-30 biogenesis.** (A) Schematic of scaRNA9 constructs annotating the three CNNC domains found downstream of the mgU2-30 fragment and the mutations that replace or delete these domains in mutants #3, #4, and #5. (B) Northern blot detection of the mgU2-30 fragment and FL scaRNA9 from RNA ectopically-expressing the WT, or mutant #3, #4, or #5 scaRNA9 construct. (C) Histogram generated from the quantification of data shown in (B). For each condition, the mgU2-30 signal was divided by the FL scaRNA9 signal in that lane, and normalized to that obtained from WT. N = 4 biological repeats, \*\* =  $p < 0.0025$ , \*\*\*\* =  $p < 0.0001$ , error bars represent standard deviation. (D) Northern blot detection of RNA isolated from *in vitro* processing assays with Drosha/DGCR8 using WT and mutant #3, #4 and #5 scaRNA9 constructs as substrates. The mgU2-30 fragment generated by the *in vitro* processing assay is denoted.