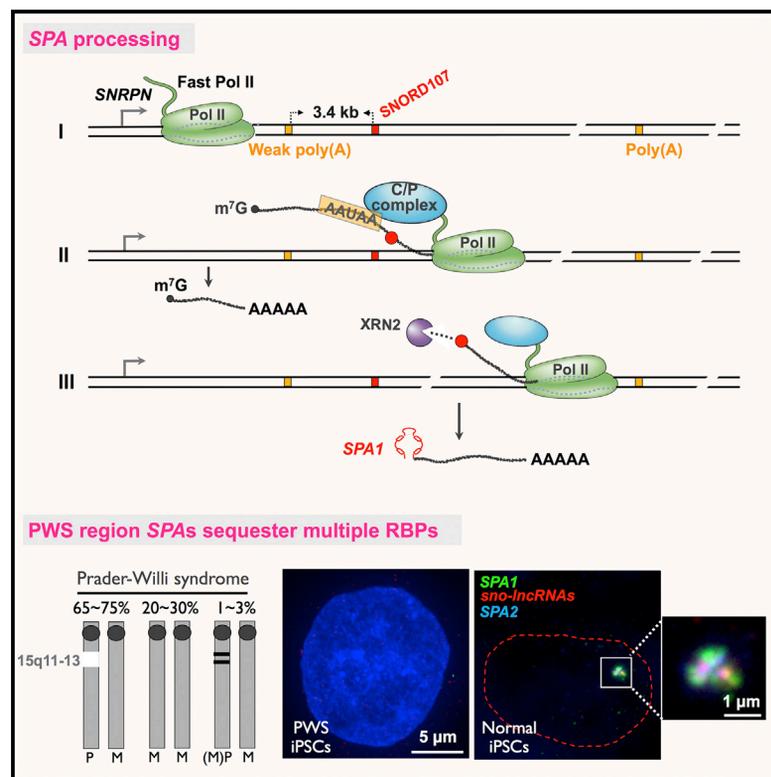


# Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins

## Graphical Abstract



## Authors

Huang Wu, Qing-Fei Yin, Zheng Luo, ..., Jian-Feng Xiang, Li Yang, Ling-Ling Chen

## Correspondence

linglingchen@sibcb.ac.cn

## In Brief

Wu et al. report SPA lncRNAs that are capped by snoRNAs and require snoRNP complexes to protect them from trimming by XRN2. Two SPAs associated with Prader-Willi syndrome can sequester multiple RBPs and regulate alternative splicing.

## Highlights

- Unusual processing generates 5' SnoRNA capped and 3' PolyAdenylated (SPA) lncRNAs
- SPA processing requires a fast Pol II and snoRNP protection from XRN2 trimming
- Two SPAs that sequester multiple RBPs in hESCs are absent in PWS patients
- Knockout SPAs in hESCs leads to altered patterns of RBPs binding and splicing

## Accession Numbers

GSE85851  
GSE85852  
GSE85854



# Unusual Processing Generates *SPA* LncRNAs that Sequester Multiple RNA Binding Proteins

Huang Wu,<sup>1,3,5</sup> Qing-Fei Yin,<sup>1,5</sup> Zheng Luo,<sup>2,3,5</sup> Run-Wen Yao,<sup>1,3</sup> Chuan-Chuan Zheng,<sup>1,3</sup> Jun Zhang,<sup>1</sup> Jian-Feng Xiang,<sup>1,2</sup> Li Yang,<sup>2,3,4</sup> and Ling-Ling Chen<sup>1,3,4,6,\*</sup>

<sup>1</sup>State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Key Laboratory of Molecular Andrology, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup>Key Laboratory of Computational Biology, CAS Center for Excellence in Brain Science and Intelligence Technology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>4</sup>School of Life Science, ShanghaiTech University, Shanghai 200031, China

<sup>5</sup>Co-first author

<sup>6</sup>Lead Contact

\*Correspondence: [linglingchen@sibcb.ac.cn](mailto:linglingchen@sibcb.ac.cn)

<http://dx.doi.org/10.1016/j.molcel.2016.10.007>

## SUMMARY

We identify a type of polycistronic transcript-derived long noncoding RNAs (lncRNAs) that are 5' small nucleolar RNA (snoRNA) capped and 3' polyadenylated (*SPAs*). *SPA* processing is associated with nascent mRNA 3' processing and kinetic competition between XRN2 trimming and Pol II elongation. Following cleavage/polyadenylation of its upstream gene, the downstream uncapped *pre-SPA* is trimmed by XRN2 until this exonuclease reaches the co-transcriptionally assembled snoRNP. This snoRNP complex prevents further degradation, generates a snoRNA 5' end, and allows continuous Pol II elongation. The imprinted 15q11-q13 encodes two *SPAs* that are deleted in Prader-Willi syndrome (PWS) patients. These lncRNAs form a nuclear accumulation that is enriched in RNA binding proteins (RBPs) including TDP43, RBFOX2, and hnRNP M. Generation of a human PWS cellular model by depleting these lncRNAs results in altered patterns of RBPs binding and alternative splicing. Together, these results expand the diversity of lncRNAs and provide additional insights into PWS pathogenesis.

## INTRODUCTION

In eukaryotic cells, nascent pre-mRNA processing, generally including capping, splicing, and cleavage/polyadenylation (C/P), is functionally associated with RNA polymerase II (Pol II) transcription. 7-methyl guanosine (m<sup>7</sup>G) capping at the 5' end occurs during the initial phase of Pol II transcription. The cap structure increases mRNA stability, regulates pre-mRNA processing and nuclear export, and stimulates mRNA translation (Cowling, 2009).

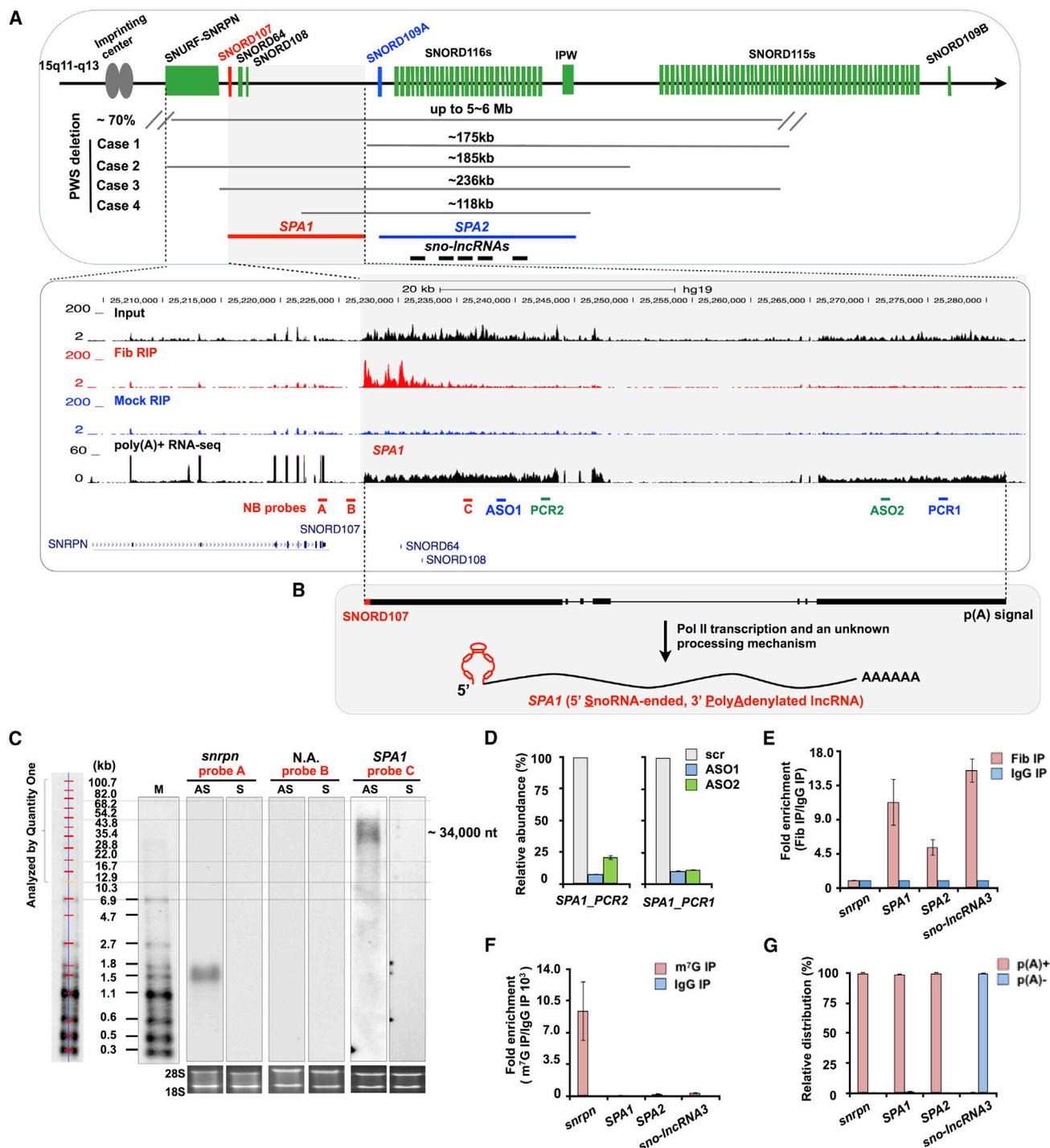
The mechanism of 3' end processing of nearly all Pol II transcribed RNAs in eukaryotic cells involves C/P of nascent transcripts. The C/P machinery first recognizes the AAUAAA

hexanucleotide (or some variants), often together with a downstream G/U-rich sequence present in nascent transcripts, resulting in the endonucleolytic cleavage of the pre-mRNA by the cleavage and polyadenylation specificity factors (CPSFs) (Mandel et al., 2006). A poly(A) tail is subsequently added by poly(A) polymerase to the 3' end of the transcript in a non-templated fashion.

The 3' end maturation of a pre-mRNA is tightly connected with Pol II termination. Transcription of the poly(A) signal triggers the endonucleolytic cleavage of the nascent RNA, generating an upstream cleavage product that is immediately polyadenylated. The remaining downstream cleavage product, with an uncapped phosphate at its 5' end, is highly unstable and is rapidly degraded (Shi et al., 2009). One model explains that the connection between Pol II termination and C/P is the torpedo model (West et al., 2004). The endonucleolytic cleavage at the poly(A) site creates an entry site for XRN2, the 5'→3' exonuclease in humans, to degrade the RNA downstream of the cleavage site. Short nascent RNA induces Pol II elongation complex arrest and promotes termination (Rosonina et al., 2006).

The m<sup>7</sup>G cap and 3' poly(A) are hallmark structures of eukaryotic mRNAs that usually contain multiple exons. These features are also true for most long noncoding RNAs (lncRNAs) (Cabili et al., 2011; Khalil et al., 2009; Ulitsky et al., 2011). However, recent studies have shown that some Pol II transcribed lncRNAs are processed in alternative ways. For example, the excised intron-derived *sno-lncRNAs* are ended by small nucleolar RNAs (snoRNAs) at both ends (Yin et al., 2012).

SnoRNAs are a family of conserved RNAs that are concentrated in Cajal bodies or nucleoli where they either function in the modification of small nuclear RNAs (snRNAs) or ribosomal RNAs (rRNAs), or participate in the processing of rRNAs during ribosomal subunit maturation (Boisvert et al., 2007; Kiss, 2001; Matera et al., 2007). The great majority of snoRNAs are encoded in the introns of protein-coding genes, and usually one intron contains one snoRNA gene (Filipowicz and Pogacić, 2002). SnoRNAs are processed from excised and debranched introns by exonucleolytic trimming and carry out their functions in complex with specific proteins by forming ribonucleoprotein complexes (snoRNPs) (Kiss, 2001). There are two main classes of



**Figure 1. Identification of SPAs from the Imprinted Region of Chr15**

(A) SPAs from the imprinted region of chr15. Top: diagram of the region. Transcription of *SNURF-SNRPN* and downstream noncoding region (green bars) occurs only from the paternal chromosome. 70% of the PWS individuals contain the 15q11-q13 deletion up to 5~6 Mb (Cassidy et al., 2012), and the minimal chromosome deletions reported in four PWS individuals are indicated (gray lines, cases 1-4 (Bieth et al., 2015; de Smith et al., 2009; Duker et al., 2010; Sahoo et al., 2008)). *SPA1*, *SPA2* (red or blue lines), and five *sno-lncRNAs* (black lines) (Yin et al., 2012) are shown underneath. Bottom: Fib RIP RNA-seq (red) and poly(A)<sup>+</sup> RNA-seq (black) revealed *SPA1* in this region in PA1 cells. See also Figures S1 and S2 and Table S1.

(B) A schematic view of *SPA1*. *SPA1* contains seven exons and is a lncRNA with the 5' snoRNA cap and a 3' poly(A) tail.

(C) Validation of *SPA1* expression in hESCs H9. Total RNAs were isolated from H9 cells and resolved on an agarose gel. Probes for NB were shown in (A). Equivalent amounts of RNAs from H9 cells were loaded as indicated by 28S and 18S rRNAs. S, sense; AS, antisense.

(legend continued on next page)

snoRNAs: box C/D snoRNAs and box H/ACA snoRNAs. The box C/D snoRNP contains four key proteins: Fibrillarin (Fib), Nop56, Nop58, and 15.5k (Kiss, 2001). *Sno-IncRNAs* are formed when one intron contains two snoRNA genes. During splicing, the sequences between the snoRNAs are not degraded, leading to the accumulation of lncRNAs flanked by snoRNA sequences but lacking 5' caps and 3' poly(A) tails (Yin et al., 2012). The genomic region encoding the most abundant *sno-IncRNAs* in human embryonic stem cells (hESCs) is specifically deleted in Prader-Willi syndrome (PWS) (Figure 1A) (Yin et al., 2012).

PWS is a neuro-developmental genetic disorder (Cassidy et al., 2012). The minimal deletion associated with PWS has been mapped to the 3' untranslated region (UTR) of *SNURF-SNRPN* at the imprinted region of 15q11-q13 (Figure 1A) (Bieth et al., 2015; de Smith et al., 2009; Duker et al., 2010; Sahoo et al., 2008). This critical deletion region that causes PWS is paternally transcribed and expresses a number of snoRNAs and lncRNAs (Figure 1A), but a detailed annotation of this important genomic region still has been lacking. Here we identified two SPAs, a type of previously unreported lncRNAs that are 5' capped by snoRNAs and 3' polyadenylated. These SPAs assemble into a striking nuclear accumulation and affect binding patterns of RNA binding proteins (RBPs) and alternative splicing in hESCs, suggesting a link between aberrant localizations of multiple RBPs and PWS pathogenesis.

## RESULTS

### Identification of SPAs

Since the minimal PWS deletion region contains many snoRNAs (Figure 1A), we reasoned that RNA-immunoprecipitation (RIP) of snoRNP-associated RNAs followed by RNA-sequencing (RNA-seq) would allow us to identify additional snoRNA-related lncRNAs, if any. Fibrillarin is one component of the box C/D snoRNP (Kiss, 2001). We carried out Fib-RIP followed by RNA-seq (Fib-RIP-seq) from the human ovarian carcinoma cell line PA1 cells using antibodies against Fib. The expressed box C/D snoRNAs (Table S1) and the known PWS region *sno-IncRNAs* (Figures S1A and S1B) could be readily precipitated by Fib-RIP-seq. Unexpectedly, we identified two SPAs (5' snoRNA capped and 3' polyadenylated lncRNAs) in this imprinted region (Figure 1A and Table S1).

*SPA1* and *SPA2* are sequentially located downstream of the protein-coding bicistronic gene *SNURF-SNRPN* (Figure 1A). This region has been previously annotated as the 3'UTR of *SNRPN*. The 5' end of *SPA1* correlates with the snoRNA SNORD107 (Figure 1A, bottom panel). *SPA1* contains seven exons and is 34,000 nt in length (Figure 1B). We have confirmed

the existence of *SPA1* by northern blotting (NB) in H9 cells (Figure 1C) and in PA1 cells (Figure S1C). Furthermore, two phosphorothioate-modified antisense oligodeoxynucleotides (ASOs) that targeted different exons of *SPA1* both reduced *SPA1* to barely detectable levels (Figure 1D), revealing that *SPA1* contains multiple exons. We further confirmed that *SPA1* could only be precipitated with the anti-Fib antibody (Figure 1E), but not with an anti-m<sup>7</sup>G cap antibody (Figure 1F), and that *SPA1* is only present in the polyadenylated fractionation in PA1 and H9 cells (Figure 1G and data not shown). Although capped by a snoRNP at the 5' end, we observed that the half-life of the polyadenylated *SPA1* is similar to that of many m<sup>7</sup>G capped mRNAs (Figure S1D).

In both PA1 and H9 cells, *SPA2* is also produced from the PWS deletion region (Figure 1A). The *SPA2* gene is located 5.6 kb downstream of *SPA1* and contains over 30 exons (Figure S2A). Its 5' end is capped by SNORD109A and is 16,000 nt in length (Figure S2A). The existence of *SPA2* in H9 cells was confirmed by NB with probes that recognize its exons across splicing junctions or its last exon, the *IPW* region (Figure S2B), and by ribonuclease protection assay (RPA) (Figure S2C). It should be noted that the last exon of *SPA2* was previously annotated as the lncRNA *IPW* (Wevrick and Francke, 1997). Together, our analyses show that the 3' UTR of *SNURF-SNRPN* (Kishore and Stamm, 2006; Runte et al., 2001) comprises two additional SPA lncRNAs (Figure S2A).

### SPA Processing Requires an Intact SnoRNA and a Weak Upstream Poly(A) Signal

Nearly all Pol II transcribed exon-containing RNAs in mammals are capped by m<sup>7</sup>G at their 5' ends to protect RNAs from the 5' exonuclease degradation (Ross, 1995). However, the 5' end of *SPA* is not m<sup>7</sup>G capped (Figures 1A and 1F). We asked what mechanism is involved in the 5' end formation of *SPA*. Nascent RNA-seq in PA1 cells (Zhang et al., 2016) revealed that *SPA1* is transcribed from the same promoter as its upstream *SNURF-SNRPN* gene and is processed from the *SNURF-SNRPN* polycistronic transcript (Figure S2D).

We next recapitulated *SPA1* formation in an expression vector to investigate its processing in detail. We cloned the last two exons of the *SNURF-SNRPN* gene (*E9+E10*) with its endogenous poly(A) signals, the intact 3.4 kb long internal sequence between *SNURF-SNRPN* and *SPA1*, and the first 800 nucleotides of *SPA1* including its 5' end SNORD107, followed by a poly(A) signal embedded in the vector (Figure 2A, left panel). We expected that this vector would transcribe a 5,000 nt *pre-RNA*, which could be processed into a mRNA containing *e9+e10* and *SPA* (Figure 2A, top right panel). Transfection of this plasmid into

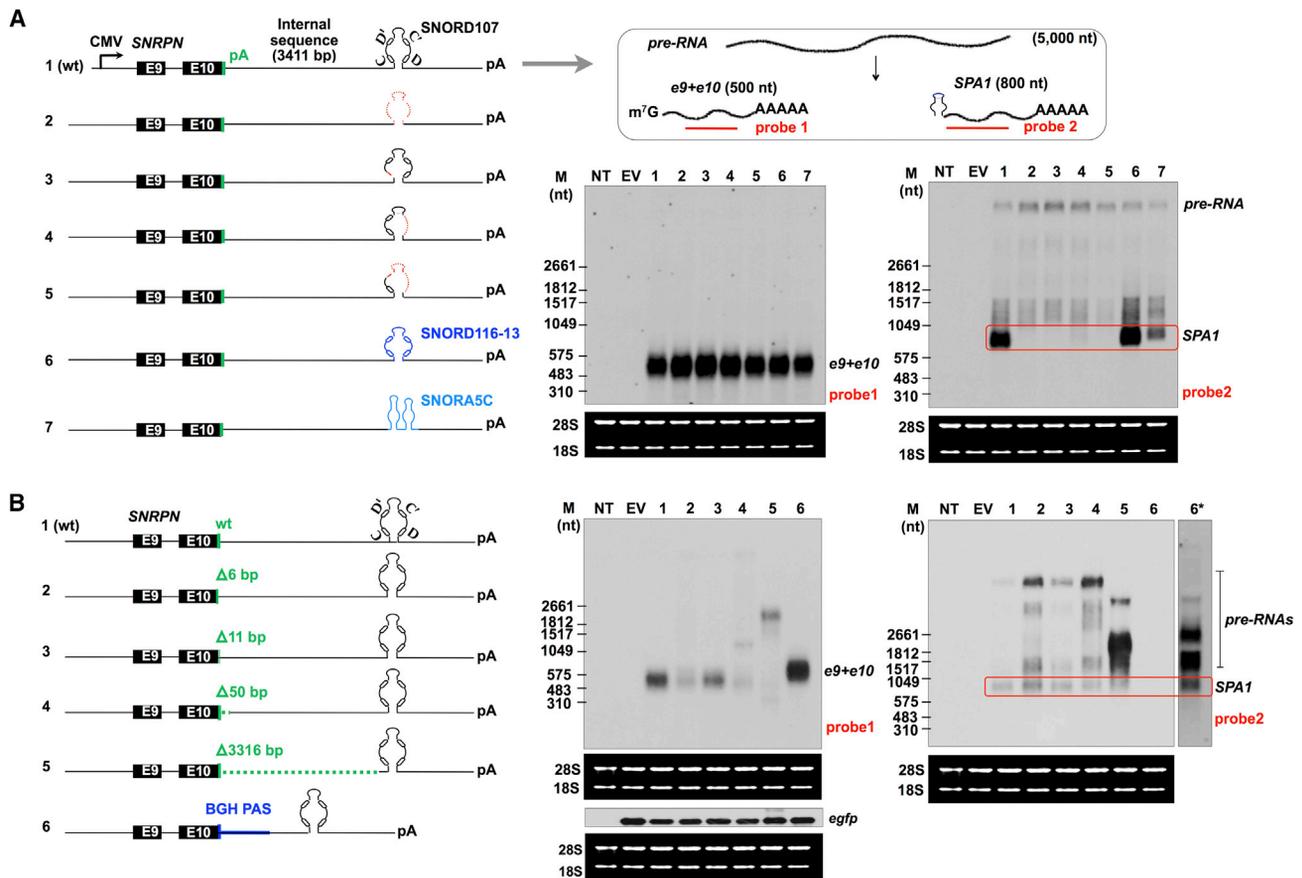
(D) Knockdown of *SPA1* by ASOs targeting two distinct exons followed by examining *SPA1* with primers located in opposite exons confirmed *SPA1*. Locations of ASOs and primers are shown in (A).

(E) SPAs are associated with the snoRNP complexes. RIP was performed from PA1 cells using anti-Fib and anti-IgG followed by qRT-PCR. Bar plots represent fold enrichments of RNAs immunoprecipitated by anti-Fib over IgG.

(F) SPAs are not capped by m<sup>7</sup>G. RIP was performed with total RNAs from PA1 cells using anti-m<sup>7</sup>G cap antibody and IgG followed by qRT-PCR. Bar plots represent fold enrichments of RNAs immunoprecipitated by anti-m<sup>7</sup>G cap over IgG.

(G) SPAs are polyadenylated. Poly(A)<sup>+</sup> and poly(A)<sup>-</sup> RNAs were purified from PA1 cells followed by qRT-PCR to examine the relative abundance of SPAs in each sample. The poly(A)<sup>+</sup> *Snrpn* and poly(A)<sup>-</sup> *sno-IncRNAs* (Yin et al., 2012) were controls.

In (D)–(G), error bars represent SD in triplicate experiments.



### Figure 2. SPA Processing Requires cis-Elements

(A) SPA processing requires a snoRNA structure at its 5' end. Left: a schematic drawing of wild-type (WT) SPA1 and mutants in the expression vector; top right box shows the processing of SPA1 expression vector. Right: total RNA isolated from HeLa cells transfected with each indicated plasmid was resolved on agarose gels for NB with probe1 or probe2, and rRNAs were used as the loading control. See also Figure S3A.

(B) SPA processing requires a weak upstream poly(A) signal. Left: a schematic drawing of a WT SPA1 and mutants containing upstream poly(A) signals in expression vectors. Right: total RNA isolated from HeLa cells transfected with each indicated plasmid was resolved on agarose gels for NB. Efficiency of plasmid transfection was indicated by the co-transfection with pEGFP-C1 vector and detected by NB with an *egfp* probe.

HeLa cells, which lack endogenous SNURF-SNRPN and SPA1, followed by NB revealed that both *e9+e10* and SPA1 could be produced (Figure 2A, lane 1). Importantly, deletion of the intact SNORD107 or its box C, or box C'/D motifs that disrupt the snoRNP structure, could completely eliminate SPA formation, while the expression of *e9+e10* remained unchanged (Figure 2A, lanes 2–5). Replacement of the endogenous SNORD107 with another box C/D or box H/ACA snoRNA still allowed SPA formation (Figure 2A, lanes 6 and 7). Moreover, RPAs also revealed that the 5' end of SPA1 is a snoRNA in H9 cells, as well as in HeLa cells that were transfected with the SPA1 vector (Figure S3A). These results together demonstrate that the essential motifs of a snoRNA at the 5' end are required for SPA processing.

Production of SPA requires continuous Pol II transcription of the downstream region including the snoRNA, after read through of the upstream poly(A) signal. We thus examined whether the strength of the upstream poly(A) signal would have an effect on SPA processing. Deletion of the upstream poly(A) signal, either the AAUAAA C/P recognition sequences (Figure 2B, lane 2), or

the cleavage signal (Figure 2B, lane 3), or both the recognition and cleavage sequences (Figure 2B, lane 4), allowed the upstream RNA to be processed at different levels. A weaker poly(A) strength led to less processed *e9+e10*. Correspondingly, more pre-RNA accumulation could be observed and more SPA1 could be detected (Figure 2B, right panels). It is worth noting, however, that the complete deletion of the poly(A) signal of *e9+e10*, together with any possible cryptic poly(A) signals in the 3.4 kb internal sequence downstream, not only completely inhibited the maturation of the upstream *e9+e10*, but also significantly reduced SPA1 processing (Figure 2B, lane 5). These results suggest that the poly(A) signal of the upstream transcript is required for downstream SPA processing from a polycistronic transcript and that different strengths of the upstream poly(A) sequences can affect the downstream SPA processing. In fact, the replacement of endogenous poly(A) sequence with a strong poly(A) signal from the BGH gene led to the strong processing of the upstream *e9+e10* but dramatically eliminated SPA1 formation (Figure 2B, lanes 6 and 6\*). Altogether, these results reveal that

*SPA* formation requires an intact snoRNA, which is either a box C/D or a box H/ACA snoRNA and a weak poly(A) signal of its upstream gene in the polycistronic transcript.

### **SPA Processing Requires Protein Factors Involved in the Cleavage/Polyadenylation of Pre-mRNA and the Torpedo Model of Pol II Termination**

Transcription of the poly(A) signal triggers endonucleolytic cleavage of the nascent RNA, generating an upstream cleavage product that is immediately polyadenylated (for example, the upstream *e9+e10* mRNA shown in Figures 2A and 2B). The remaining downstream cleavage product, with an uncapped phosphate at its 5' end, is unstable and rapidly degraded. Multiple factors, including CPSF73 and CPSF30, are responsible for the recognition of the poly(A) signal (Chan et al., 2014; Mandel et al., 2006). To determine whether CPSFs play roles in *SPA* processing, we knocked down CPSF73 or CPSF30 in PA1 cells followed by the examination of the relative abundance of *pre-SPA1* and *SPA1*. Depletion of CPSF73 led to the accumulation of *pre-SPA1* after it was normalized to the upstream *snurf-snrpn* in the scramble and CPSF73 shRNA treated cells (Figures S3B and S3C). A similar result was obtained by RPA where the accumulation of *pre-SPA1* was increased after CPSF73 knockdown (Figure 3A). In addition, depletion of CPSF30, another key factor involved in C/P, also led to the accumulation of *pre-SPA1* (Figure S3D).

As the torpedo model explains the connection between Pol II termination and C/P (West et al., 2004), we next examined whether XRN2 is involved in *SPA* processing. Knockdown of XRN2 led to the accumulation of *pre-SPA1* from the PWS deletion region, as revealed by qRT-PCR (Figure S3E) and RPA (Figure 3A) in PA1 cells, as well as by reporter assays in HeLa cells (Figure S3F). It should be noted that in all examined situations, knockdown of XRN2 did not significantly affect transcription or the processing of the upstream mRNA (Figures S3E and S3F), but only the ratio of the relative abundance of *pre-SPA1* and *SPA1* was altered (Figure S3F). Together, these results strongly suggest that *SPA* processing requires factors involved in C/P of the upstream mRNA and XRN2 involved in the torpedo model of Pol II termination.

### **SPA Processing Is Associated with Fast Pol II Transcription Elongation**

Measuring the Pol II elongation speed by nascent RNA-seq (Zhang et al., 2016) revealed that the Pol II transcription elongation rate (TER) at this locus (~3.4 kb/min) was much higher than the average Pol II speed (~2.5 kb/min) in both H9 and PA1 cells (Figure 3B). This observation suggested that the fast Pol II elongation could allow polymerases to reach SNORD107 before XRN2, allowing the co-transcriptional assembly of a snoRNP to occur. This snoRNP would block the arriving XRN2 degradation from the 5' end of nascent *pre-SPA1* and allow the continuous transcription of the polycistronic transcript by fast Pol II. To test this hypothesis, we modulated Pol II TER by expressing the  $\alpha$ -amanitin-resistance ( $Am^r$ ) wild-type, R749H (slow Pol II), or E1126G (fast Pol II) of the human Pol II large subunit implicated in elongation control (Fong et al., 2014; Zhang et al., 2016) in PA1 cells. After transfection with either wild-type or each Pol II

mutant,  $\alpha$ -amanitin was added to block endogenous Pol II elongation. We found that the R749H mutant reduced Pol II TER and the E1126G mutant promoted TER (Figure 3C) at the *SNURF-SNRPN* locus as expected (Figure 3C, left panel). Importantly, the relative abundance of the newly processed *SPA1*, when compared to its upstream *snrpn* mRNA at the nascent level, was significantly decreased or increased in cells with slow or fast Pol II (Figure 3C, right panel). These results thus support our hypothesis that *SPA* processing requires a fast RNA Pol II TER to win the competition with XRN2.

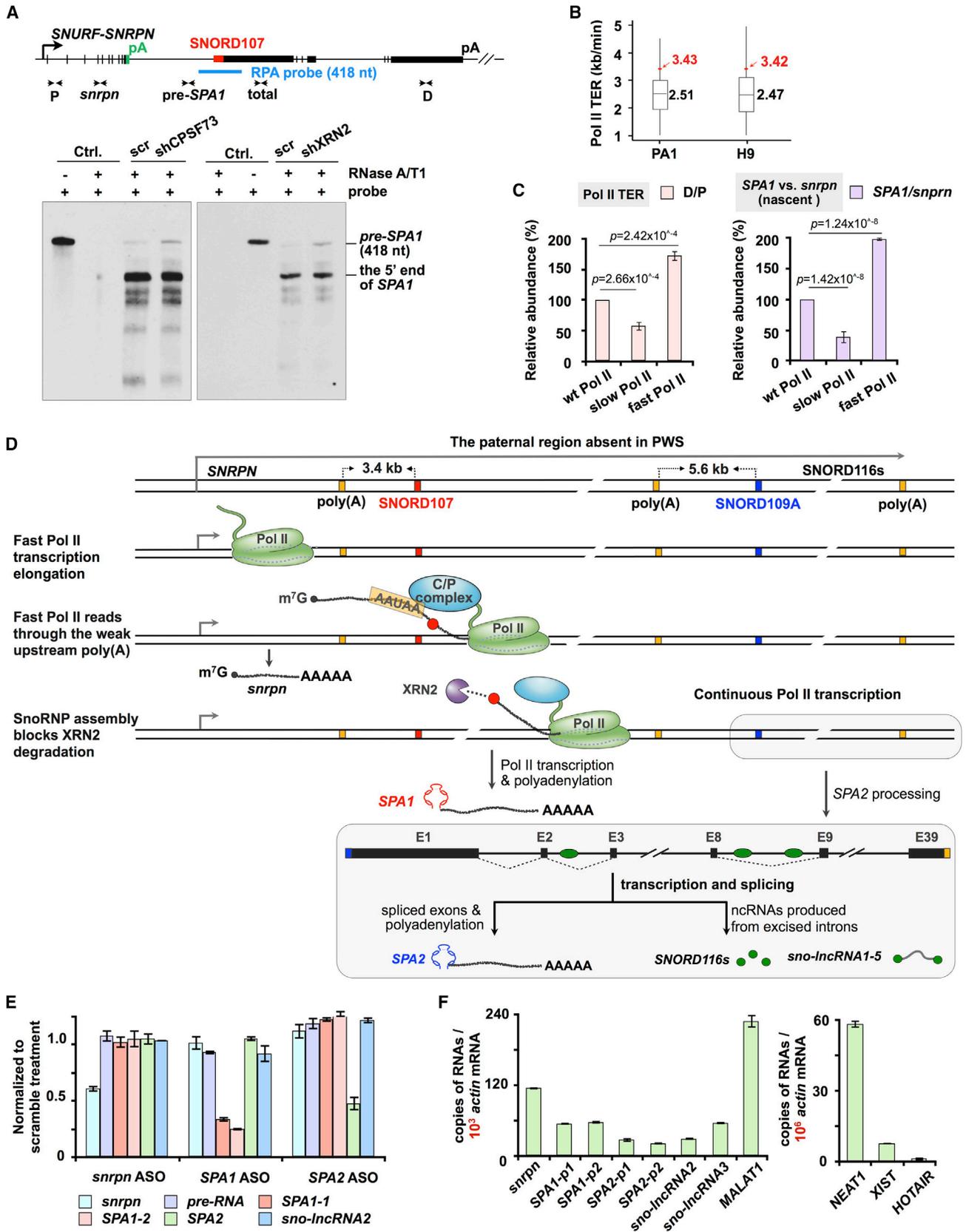
Together, we propose a model for the PWS region *SPA1* processing shown in Figure 3D. The paternal transcription at the 15q11-q13 region (Figure 1A) generates a 180 kb long *snurf-snrpn* polycistronic transcript. It can be processed into both mRNAs and different types of lncRNAs including *SPAs* and *sno-lncRNAs* from the previously thought 3' UTR of *snurf-snrpn*. Fast Pol II transcription elongation coupled with a weak promoter-proximal poly(A) signal allows the continuous transcription of fast Pol II to the downstream region including a snoRNA gene located 3.4 kb away. Following C/P at the 3' end of upstream *snurf-snrpn*, the nascent uncapped pre-RNA associates with XRN2, which chases the Pol II elongation complex in order to destabilize it. However, Pol II reaches SNORD107 first. XRN2 then encounters the co-transcriptionally assembled SNORD107 snoRNP that blocks further degradation of XRN2 and also allows continuous Pol II elongation. Such a mechanism allows the formation of the 5' end snoRNP cap of *SPA1*. Meanwhile, Pol II continues its transcription until it encounters the next poly(A) signal to initiate another round of cleavage and polyadenylation that forms the 3' end poly(A) tail of *SPA1*. In addition, *SPA2*, located 5.6 kb downstream of *SPA1*, is formed by the same mechanism as *SPA1* (gray box, Figure 3D).

Since the *SPA2* gene also comprises 29 copies of SNORD116s and 5 *sno-lncRNAs* (Yin et al., 2012) in its introns, we further examined whether *SPAs* are precursors of these *sno-lncRNAs*. Treated cells with ASOs that target *snrpn* mRNA, *SPA1*, or *SPA2* did not significantly alter *sno-lncRNA* expression (Figure 3E), suggesting that these two types of lncRNAs do not have a precursor-product relationship. Indeed, *snrpn*, *SPAs*, and *sno-lncRNAs* are all derived from the same polycistronic transcript (Figure 3D). *SPA1* and *SPA2* each contains multiple exons and is generated by a mechanism associated with polycistronic transcription, nascent RNA 3' processing, and fast Pol II elongation rate, whereas *sno-lncRNAs* are derived from the excised introns of *SPA2* (Figure 3D).

### **The PWS Region lncRNAs Form a Striking Nuclear Accumulation**

We have measured the copy number of PWS region *SPAs* in hESCs. *SPAs* represent one of the most abundant lncRNAs in hESCs (Figure 3F). For instance, the abundance of *SPAs* is about one third of that of the lncRNA *MALAT1* and is similar to the non-polyadenylated *sno-lncRNAs* in hESCs.

The nuclear/cytoplasmic RNA fractionation revealed that PWS region *SPAs* were nuclear retained (Figure S4A). By RNA fluorescence in situ hybridization (FISH) with tiling probes that are antisense to the full length of *SPA1*, *SPA2*, or *sno-lncRNAs* (Figure 4A), we found that these lncRNAs accumulated to a striking



(legend on next page)

nuclear accumulation in each nucleus of H9 and PA1 cells (Figures 4B and S4B). These accumulations were about 0.4~1.2  $\mu\text{m}$  in diameter, and their volumes were 1~2  $\mu\text{m}^3$  (Figure 4C). Their sizes were larger and greater in H9 cells than those in PA1 cells (Figure 4C). In each subnuclear accumulation, sno-lncRNAs appeared to be located in the middle, surrounded by SPA1 and SPA2 (Figures 4B and S4B).

Double DNA/RNA FISH showed that SPAs accumulated at or near their sites of processing on the paternal chromosome (Figure 4D). To exclude the possibility that the detected signals by RNA FISH were from the partially processed PWS region polycistronic transcript, we blocked Pol II transcription followed by examining the SPA accumulation. 3 hr in Act D or DRB treatment was sufficient to eliminate the polycistronic transcript, with only 5%~12% retaining a barely detectable signal revealed by an intron probe (Figure 4E). In contrast, the SPA1 RNA, although somewhat reduced, remained in 98%~100% of these same nuclei (Figure 4E). Thus, the persistence of PWS region lncRNAs in these transcriptionally inhibited interphase cells is not due to the unprocessed polycistronic transcript. Finally, we found that the striking subnuclear accumulation of SPAs and sno-lncRNAs also existed in induced pluripotent stem cells (iPSCs) derived from normal individuals but was absent in iPSCs derived from PWS patients (Yang et al., 2010) (Figure 4F).

### Conservation of PWS Region SPAs

Comparison of human and mouse genomes revealed that the genome organization of SPA1 is largely conserved (Figure S5A). Similar to human SPA1, the 5' end of mouse SPA1 (*mSPA1*) is *Snord107*. *mSP* contains 9 exons, and its predicted size is about 20,000 nt (Figure S5A). However, the homolog of 5' end SNROD109A for the human SPA2 gene is not present in the mouse genome. Indeed, there is a gap missing in the mouse genome that corresponds to the SPA2 locus in human, suggesting that SPA2 lncRNA is not conserved (Figure S5A). Remarkably, although *Snrpn* is expressed in mouse embryonic stem cells (mESCs) and upon neuronal differentiation (Figure S5B), *mSPA1* is largely absent in mESCs but highly expressed in neurons (Figure S5B). We further confirmed the expression of *mSPA1* in the adult mouse hippocampus tissue, but not in mESCs by NB (Figure S5C). These analyses indicate a different processing of *mSPA1* from human SPA1 and the non-conserved feature of the PWS region lncRNAs between mouse and human.

### Generation of Allele-Specific Isogenic hESC Lines Lacking PWS Region lncRNAs

Because the PWS region SPAs are not conserved (Figure S5) and the current existing mouse models cannot fully recapitulate the clinical features of human PWS (Cassidy et al., 2012; Ding et al., 2008; Tsai et al., 1999), we decided to generate isogenic hESC lines lacking these lncRNAs by CRISPR/Cas9 to create an unbiased human PWS cellular model.

We designed sgRNAs to delete the entire 141 kb genomic region that comprises SPA1 and SPA2 in H9 cells (Figure 5A). As this locus is a genomic imprinting region only transcribed paternally, we selected H9 clones carrying the 141 kb paternal chromosome deletion for a minimum perturbation of the genome. Compared to wild-type (WT) cells, H9 clones having the paternal deletion (P-KO) did not express PWS region lncRNAs (Figures 5A and 5B). The lncRNA-enriched PWS accumulation also disappeared in P-KO cells (Figure S4C). It should be noted that this 141 kb genomic deletion inevitably removed the SPA2 intronic-located SNORD116 snoRNAs, which are orphans and may possibly lack functional potential (Bazeley et al., 2008; Bratkovič and Rogelj, 2014).

H9 P-KO clones of the 141 kb paternal deletion appeared morphologically normal, expressed pluripotency markers, maintained normal karyotype, and could differentiate into all three germ layers (data not shown). RNA-seq and NB revealed little changes in *snrpn* expression in P-KO cells (Figures 5A and 5C), or few alterations in global gene expression (data not shown). These observations suggest that although the PWS region lncRNAs primarily accumulated to their sites of synthesis (Figure 4), they do not act in *cis*. Furthermore, knockout of SPAs had no significant effect on the CpG methylation status at the imprinting center located upstream of the *SNURF-SNRPN* locus (data not shown). Together, we conclude that these PWS region SPAs and sno-lncRNAs are not essential for hESCs, consistent with the fact that PWS patients lacking these lncRNAs are viable.

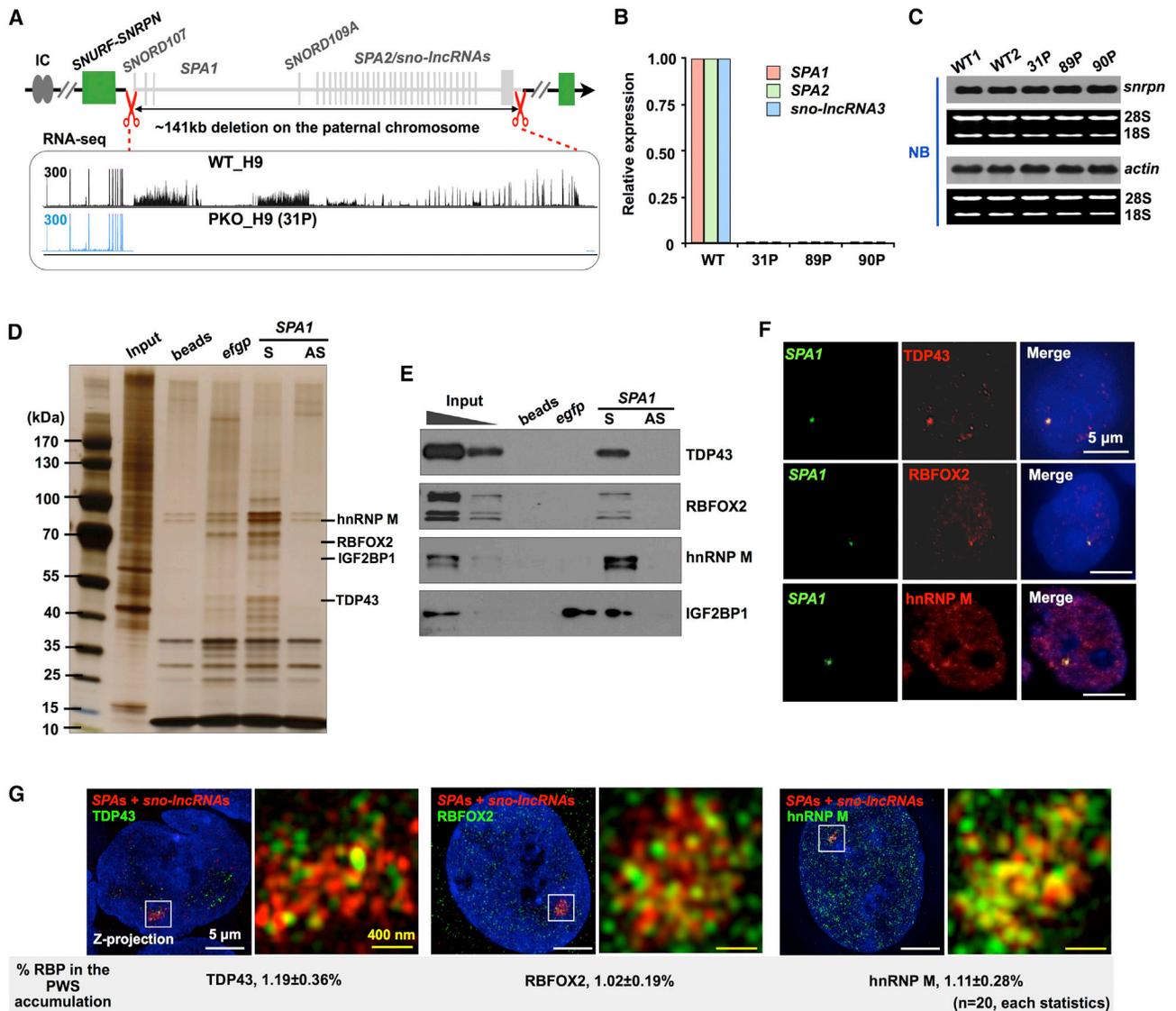
### PWS Region lncRNAs Sequester RBPs with Distinct Binding Preferences

To further explore the role of SPA1, we set up RNA pull-down experiments to identify SPA1-interacting proteins. Since SPA1 is long (Figure 1B), we generated 1,000~2,000 nt sense and antisense biotin-labeled fragments that span almost the full SPA1. After incubation with nuclear extracts isolated from PA1 cells with the biotin-labeled control *egfp*, sense, or antisense SPA1 fragments, we identified three RBPs, TDP43, RBFOX2, and

### Figure 3. SPA Processing Requires Factors Involved in Nascent RNA 3' End Formation and Pol II Termination

(A) Processing of SPA1 requires CPSF73 and XRN2. Top: a schematic drawing of the gene organization of *SNURF-SNRPN* and SPA1. Blue line, the RPA probe; arrowheads, primers used in (C). *pre-SPA1* was accumulated in CPSF73-depleted or XRN2-depleted PA1 cells, as revealed by RPAs. See also Figures S3B–S3F. (B) The Pol II TER at the *SNURF-SNRPN* locus is faster than the average rate in both H9 and PA1 cells, as revealed by 4sUDRB-seq (Zhang et al., 2016). (C) SPA1 processing is associated with rapid Pol II TER. Primer sets were shown in (A). Left, “D” and “P” primer sets were used to quantify pre-mRNA accumulation at distal and proximal region with respect to the transcription start site (TSS). Right, the slow Pol II (R749H) resulted in decreased SPA1 expression and the fast Pol II (E1126G) led to increased SPA1 expression, compared to the level of its upstream *snrpn* mRNA (right). (D) A model for SPA1 processing. See text for details. (E) SPAs are not precursors of sno-lncRNAs. Knockdown of *snrpn* mRNA, SPA1, or SPA2 by ASOs had no significant effect on sno-lncRNA expression, as revealed by qRT-PCR. (F) Relative abundance of PWS region lncRNAs. Copy numbers of PWS region lncRNAs and other lncRNAs were normalized to actin mRNA in H9 cells. In (C), (E), and (F), error bars represent SD in triplicate experiments.





### Figure 5. PWS Region SPAs Interact with Multiple RNA Binding Proteins

(A) A schematic diagram of knockout (KO) the PWS region lncRNAs by CRISPR/Cas9. Top: the gene organization of *SNURF-SNRPN* locus and the sgRNA target sites for the 141 kb KO comprising both *SPA1* and *SPA2* genes. Bottom: poly(A)<sup>+</sup> RNA-seq revealed the successful deletion of SPAs in one of the paternal KO (P-KO) H9 cell lines. See also Figures S4C and S5.

(B) Verification of *SPA1*, *SPA2* and *sno-lncRNAs* KO in H9 cells. RT-qPCR indicates an efficient KO of *SPA1*, *SPA2* and *sno-lncRNAs* in three P-KO H9 cell lines.

(C) KO SPAs did not affect its upstream *snrnp* mRNA expression, revealed by NB.

(D) Identification of proteins associated with *SPA1*. Proteins from PA1 cell extracts were pulled down with the biotin-labeled probes of *SPA1* and controls and then subjected to SDS-PAGE and silver staining.

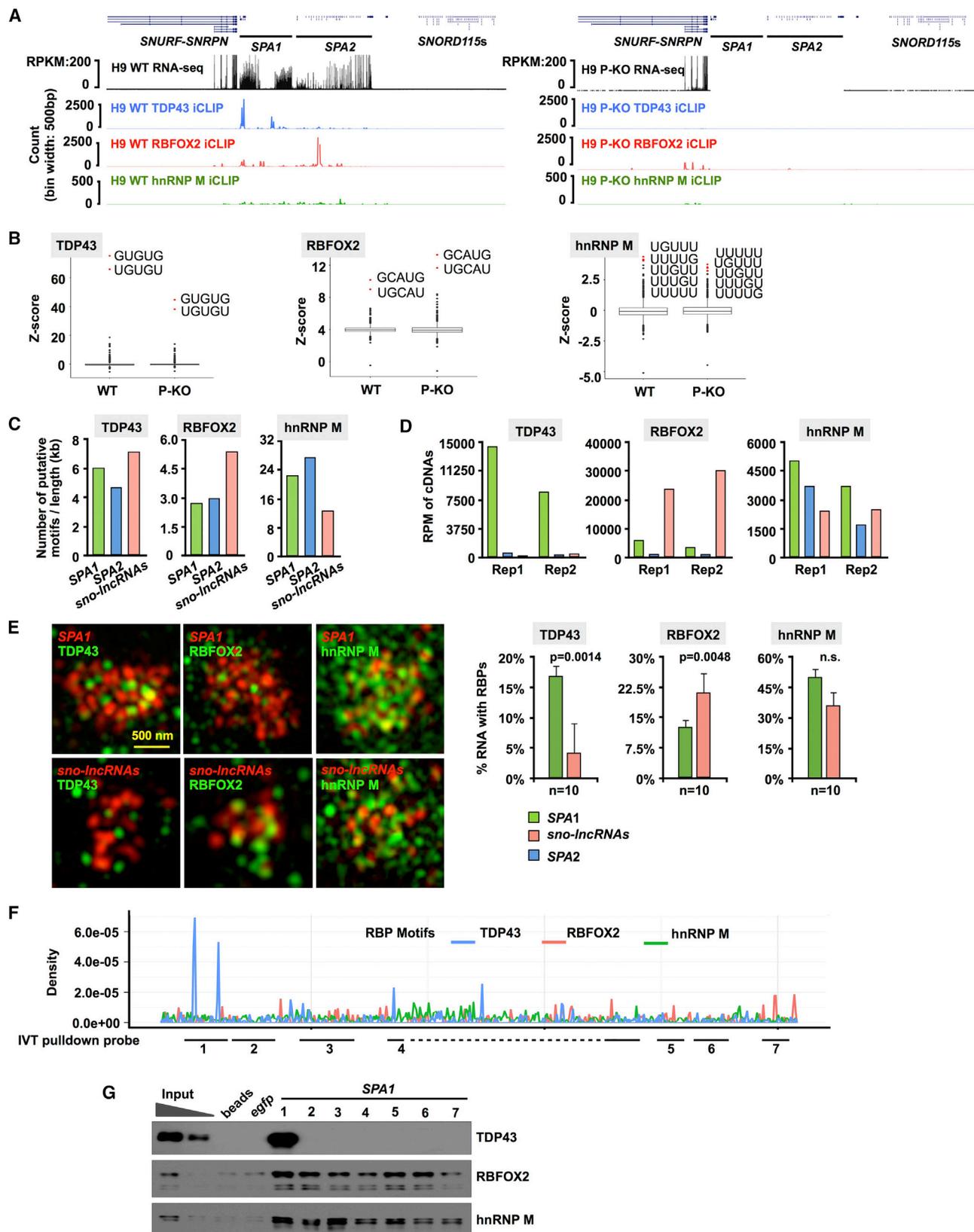
(E) Validation of *SPA1*-interacting proteins. Proteins pulled down by biotin-*SPA1* were analyzed by immunoblotting with TDP43, RBFOX2, hnRNP M, and IGF2BP1 antibodies.

(F) Co-localization of *SPA1* (green) and proteins (red). *SPA1* co-localizes to the enriched staining of TDP43, RBFOX2, and hnRNP M in the nucleus. Representative images are shown.

(G) TDP43, RBFOX2, and hnRNP M are co-localized to the PWS region lncRNA-enriched nuclear accumulation. Representative images of the co-staining of each RBP and PWS-region lncRNAs under 3D SIM. Statistics of the proportion of each RBP enriched in this lncRNA-enriched accumulation are marked underneath. The percentage analysis was processed using the TANGO plug-in for ImageJ/Fiji to generate the 3D mask/segment of each signal channel followed by an integrated density counting. See also Figure S6A.

respectively, the PWS lncRNAs constitute about 0.02%~0.1% of the nuclear volume but can sequester greater than 1% of each RBP.

To further illustrate the binding capability of each RBP with these lncRNAs at the single-nucleotide resolution, we performed individual-nucleotide resolution UV cross-linking and



(legend on next page)

immunoprecipitation (iCLIP) with antibodies for TDP43, RBFOX2, and hnRNP M in WT and P-KO H9 cell lines (Figures S6B–S6D and 6A). TDP43, RBFOX2, and hnRNP M in examined H9 cell lines all exhibited binding motifs similar to those previously reported (Huelga et al., 2012; Tollervey et al., 2011; Yeo et al., 2009) (Figure 6B). For example, TDP43 binds to clusters of UG-rich sequences (Tollervey et al., 2011), RBFOX2 binds to UGCAU and GCAUG sequences (Yeo et al., 2009), and hnRNP M binds to clusters of GU-rich sequences that often contain UU sequences within the motif (Huelga et al., 2012). In addition, comparison of the TDP43 iCLIP results in H9 cells between a previously published dataset (Tollervey et al., 2011), and ours revealed that over 70% of TDP43 targeted cDNAs were overlapped (Figure S6E). All of these analyses suggested successful iCLIP assays performed in our hands. For each antibody, we performed iCLIP in duplicate, and only changes detected in both iCLIPs were counted (Figure S6F).

iCLIP experiments confirmed that all three proteins directly and strongly interacted with *SPA1*, *SPA2*, and *sno-lncRNAs* in WT H9 cells, but no binding signals could be retrieved in P-KO H9 cells lacking these lncRNAs (Figures 6A and S6G). Calculation of putative binding motifs of these RBPs in these lncRNAs revealed remarkable enrichments of their respective motifs without preference (Figure 6C). Strikingly, a strong preference between lncRNA and RBP was revealed when the binding capability of each RBP on individual PWS region lncRNAs was counted: *SPA1* preferred to interact with TDP43, *sno-lncRNAs* preferred to bind RBFOX2, while *SPA2* did not have a preference to these examined RBPs (Figure 6D). SIM further confirmed the binding preference between TDP43 and *SPA1* as well as the preference between RBFOX2 and *sno-lncRNAs* (Figure 6E). Remarkably, *SPA1* bound to TDP43 with a strong bias toward its 5' end but interacted with RBFOX2 and hnRNP M more promiscuously (Figures 6F and 6G). Together, these results show that the PWS region lncRNAs form a striking nuclear accumulation that can sequester a significant proportion of RBPs with distinct preferences.

### hESCs Lacking PWS Region lncRNAs Have Altered Patterns of RBP Binding and Alternative Splicing

The significant enrichment of these RBPs in the PWS region lncRNA-enriched nuclear accumulation in hESCs but not in

PWS patients suggests that the lack of these lncRNAs might alter patterns of RBP binding in PWS patients. Comparing RNA targets of TDP43, RBFOX2, and hnRNP M in WT and P-KO H9 cells revealed changes in the proportion of their associated target RNAs (out of all cDNA counts mapping to the human genome in each iCLIP dataset after removal of PCR duplicates) (König et al., 2010) between WT and P-KO H9 cells (Figures S7A–S7C). We then selected transcripts that have altered RBP binding in P-KO H9 cells; meanwhile, we required that the expression of these RBP associated transcripts remained unchanged in WT and P-KO H9 cell lines (Figures S7A–S7C and Tables S2, S3, and S4). Examples of such transcripts and the proportions of cDNAs that mapped to each transcript in iCLIPs are shown in Figures S7D–S7F.

We then analyzed whether the altered RBP binding would result in alternative splicing changes in P-KO hESCs lacking these lncRNAs. 348 altered splicing events repeatedly occurred from RNA-seq of three P-KO lines (Figure 7A and Table S5). 90 among 348 (26%) showed a corresponding change of RBPs binding to pre-mRNAs (Figures 7B–7E and Table S5). Some correlated only with either TDP43 or RBFOX2, and some with both (Figures 7B–7E and S7G and Table S5), but very few correlated with hnRNP M (Figure 7B). Furthermore, the binding of TDP43 and RBFOX2 to these altered pre-mRNAs could be validated in hESCs (Figure 7F). This observation indicates that *SPA1* and *sno-lncRNAs* can regulate different populations of targets. Remarkably, some alternative splicing changes could be found in human iPSCs derived from a reported PWS patient (Yang et al., 2010) (Figure 7G). Intriguingly, a large number of altered binding sites of RBPs were located to genes that do not have cassette exons (Figure S7H), indicating that the PWS region *SPAs* and *sno-lncRNAs* can sequester RBPs away from their normal functional sites (Figure 7H).

## DISCUSSION

Most well-characterized lncRNAs that contain multiple exons are capped by m<sup>7</sup>G. We report here previously unreported nuclear RNAs that are capped by a snoRNP at their 5' ends. *SPA* processing in the chr15q11–q13 region is associated with fast Pol II transcription elongation and C/P of the upstream gene (Figures

### Figure 6. PWS Region lncRNAs Exhibit Different Binding Preferences to RBPs

(A) TDP43, RBFOX2, and hnRNP M specifically interact with PWS region lncRNAs. Left panels: iCLIP reads were enriched along the PWS deletion region, but not in the adjacent upstream or downstream regions in H9 cells. TDP43 (blue) mainly accumulated at *SPA1*, RBFOX2 (red) mainly accumulated at *sno-lncRNAs*, whereas hnRNP M (green) did not show an obvious preference. Right panels: in P-KO H9 cells, iCLIP reads of each RBP mapped to this region in WT H9 cells were not present in P-KO samples. Counts of iCLIP reads were calculated every 500 bp. See also Figures S6B–S6G.

(B) Z scores of pentamer occurrence within the 61 nucleotide (nt) sequence surrounding all cross-link sites (–30 to +30 nt) for iCLIP assays with individual TDP43 (left), RBFOX2 (middle), and hnRNP M (right) antibodies in WT and P-KO H9 cells. See also Tables S2, S3, and S4.

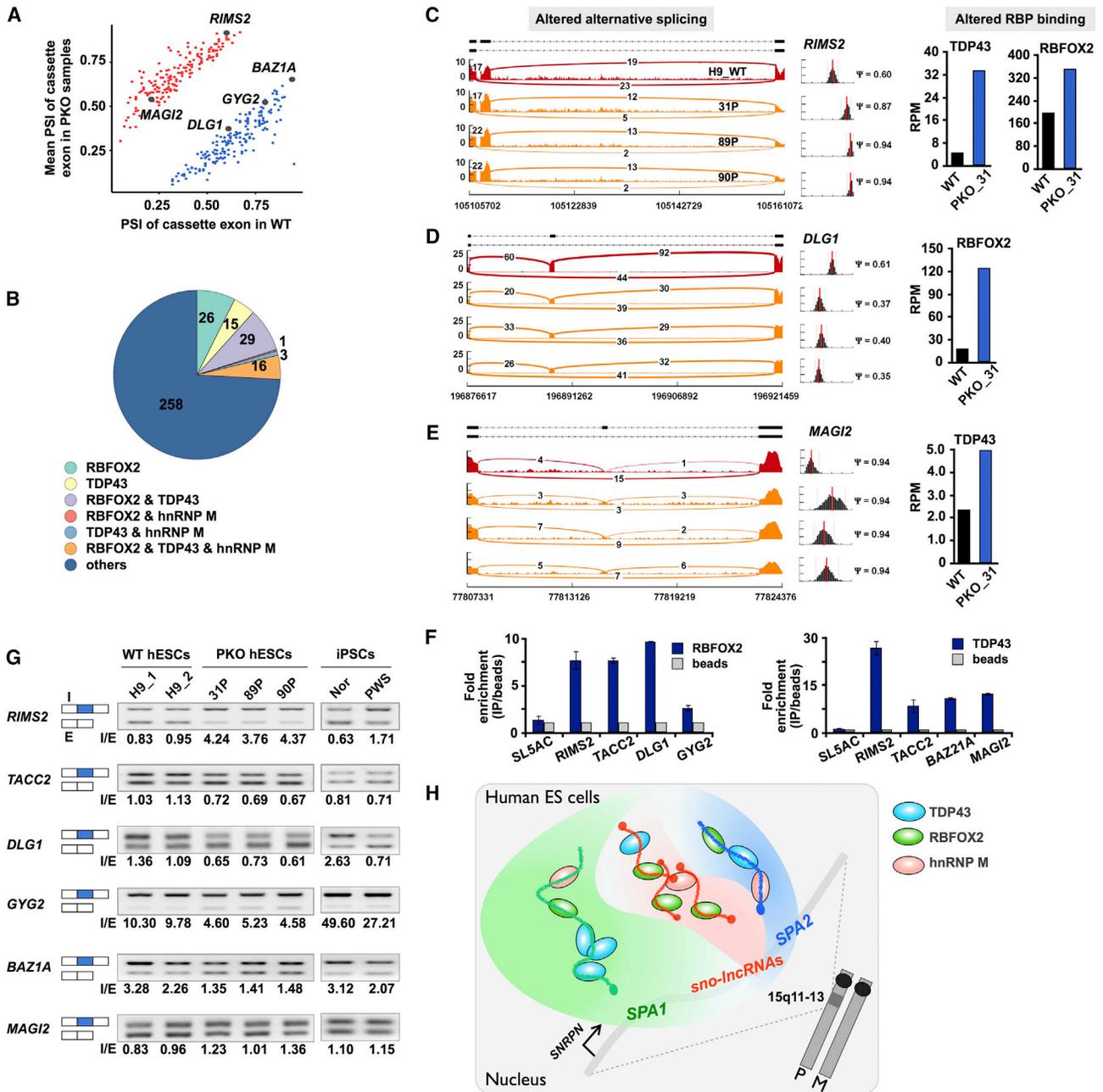
(C) Number of putative motifs of each RBP (TDP43, RBFOX2, and hnRNP M) in *SPA1*, *SPA2*, and *sno-lncRNAs* per kilobase.

(D) Binding capability of TDP43, RBFOX2, and hnRNP M to different PWS region lncRNAs. TDP43 prefers to bind to *SPA1*, RBFOX2 prefers to interact with *sno-lncRNAs*, while hnRNP M binds to all three lncRNAs without an apparent preference. Counts of iCLIP reads were calculated every 500 bp.

(E) PWS region lncRNAs interact with RBPs with preferences. *SPA1* or *sno-lncRNAs* co-staining with TDP43 or RBFOX2 under SIM, respectively. Left panels: representative images of lncRNA and RBP co-staining under SIM. Right panels: statistics of *SPA1* or *sno-lncRNAs* and their interacting RBPs from triplicate experiments. See Figure S5G for details.

(F) iCLIP binding density of TDP43, RBFOX2, and hnRNP M on *SPA1*. Note that there is a notable enrichment of TDP43 binding density at the 5' end of *SPA1*. The biotin-labeled pull-down probes for *SPA1* used in (G) are shown in the bottom.

(G) TDP43 prefers to interact with the 5' end of *SPA1*. Proteins from PA1 cell extracts were pulled down with the biotin-labeled probes of different *SPA1* fragments (F) and then subjected to immunoblotting with antibodies to TDP43, RBFOX2, and hnRNP M.



**Figure 7. HESCs Lacking PWS Region LncRNAs Have Altered Patterns of Alternative Splicing and RBP Binding to Pre-mRNAs**

(A) Altered splicing of cassette exons after depleting PWS region lncRNAs. Cassette exons with significant PSI changes ( $|\text{mean PSI in P-KO samples} - \text{PSI in WT sample}| \geq 0.2$ ) were selected. Red points represent cassette exons with more inclusion in P-KO samples; blue points represent cassette exons with more exclusion in P-KO samples. Cassette exons shown in (C)–(G) are marked in black. See also [Table S5](#).

(B) Analysis of changed alternative splicing events corresponding to altered RBP binding to pre-mRNAs in P-KO hESCs. See also [Tables S2](#), [S3](#), and [S4](#) and [Figures S7A–S7H](#).

(C–E) Examples of alternative splicing changes and their corresponding altered RBP binding in WT and P-KO hESCs. Left panels: wiggle tracks show the RNA-seq mapped reads to altered exons, inferred from the change of PSI on the right (red tracks, WT H9; orange tracks, P-KO H9 lines). Right panels: the altered binding of corresponding RBPs. Inclusion in *RIMS2* corresponded with the altered binding of both TDP43 and RBFOX2 (C); the exclusion in *DLG1* corresponded with the altered binding of RBFOX2 (D); whereas the inclusion in *MAG12* corresponded with the altered binding of TDP43 (E). See also [Figure S7G](#).

(F) TDP43 or RBFOX2 interact with pre-mRNAs with altered patterns of alternative splicing shown in (C)–(E). RIP was performed in H9 cells using anti-TDP43 and anti-RBFOX2 antibodies followed by qRT-PCR. Bar plots represent fold enrichments of RNAs immunoprecipitated by anti-TDP43 and anti-RBFOX2 antibodies, and error bars represent SD in triplicate experiments.

(legend continued on next page)

1, 2, and 3). At the *SPA*-produced *SNURF-SNRPN* locus, Pol II runs faster than its average speed in both PA1 and H9 cells (Figures 3B and 3C). Such a fast Pol II elongation at this locus would shorten the window of opportunity for the upstream 3' end formation to occur and win the race with XRN2 degradation (Figure 3D). It has been very recently shown that fast transcription elongation could shift termination downstream (Fong et al., 2015). Remarkably, the assembly of snoRNPs downstream of cleavage sites impedes XRN2 degradation and delays Pol II termination (Figure 3). These seamlessly coupled processes together unexpectedly generate *SPAs* with 5' snoRNP ends at the *SNURF-SNRPN* locus. As both box C/D and box H/ACA snoRNAs are capable of mediating *SPA* formation (Figure 2A), additional genome-wide analyses will be required to identify more *SPAs* and polycistronic transcripts of this type in different cells. For example, the analysis of publicly available ENCODE datasets revealed a few *SPAs* that were likely to be generated from the SNORD113 and SNORD115 regions (data not shown).

Bacterial and viral mRNAs are often polycistronic (Blumenthal, 1998). The use of alternative translation of bicistronic transcripts to generate protein diversity also exists in mammals (Brubaker et al., 2014), including the *SNURF-SNRPN* locus. *Snurf-snrpn* is known as a bicistronic transcript encoding two proteins: SmN, a spliceosomal protein involved in mRNA splicing (Glenn et al., 1996), and a polypeptide SNURF with unknown function (Gray et al., 1999). However, *snurf-snrpn* produces not only mRNAs and SNOD116 snoRNAs (Bazeley et al., 2008), but also different types of lncRNAs including *SPAs* and *sno-lncRNAs* (Figure 3D). While SNOD116s may lack functional significance (Bazeley et al., 2008; Bratkovič and Rogelj, 2014; Cassidy et al., 2012), the two *SPAs* and five *sno-lncRNAs* localize to a close proximity within the 1~2  $\mu\text{m}^3$  sparkle in the nucleus (Figures 4 and S4). This striking co-localization indicates that all of these lncRNAs originating from the *snurf-snrpn* polycistronic transcript that are not expressed in almost all PWS patients (Figure 1A) might possess similar functions.

At least three important RBPs, TDP43, RBFOX2, and hnRNP M, are sequestered by PWS region lncRNAs (Figures 5 and 6). These RBPs are well known to be involved in multiple aspects of mRNA metabolism regulation (Cho et al., 2014; Lagier-Tourenne et al., 2012; Park et al., 2011; Shi et al., 2009; Tollervey et al., 2011; Wang et al., 2008; Yin et al., 2012). Their aberrant localization and expression have been reported to be associated with neurogenetic disorders, such as Amyotrophic Lateral Sclerosis (Lagier-Tourenne et al., 2012; Tollervey et al., 2011) and autism (Voineagu et al., 2011; Weyn-Vanhenryck et al., 2014). We have previously reported that PWS region sno-lncRNAs interact with RBFOX2 (Yin et al., 2012). Here we show that *SPA1* exhibits a strong preference to TDP43 (Figure 6), suggesting a different role between *sno-lncRNAs* and *SPA1*.

Analysis of RNA-seq from H9 cells with or without these lncRNAs revealed altered splicing events corresponding to

altered TDP43 and RBFOX2 cross-link clusters (Figure 7). Interestingly, gene ontology showed that some genes with altered cassette exons in P-KO hESC lines were associated with synaptosome and neurotrophin signaling pathways (Figure S7I). For example, RIMS2 is a regulator of synaptic vesicles release, pre-synaptic plasticity, and insulin secretion (Kaeser et al., 2012; Kashima et al., 2001). DLG1 has a reported role in synaptogenesis (Bonnet et al., 2013; Parkinson et al., 2013). Importantly, altered cassette exons in RIMS2 and DLG1 mRNAs could be found in iPSCs derived from a human PWS patient (Figure 7G). These observations thus suggest a link between the mislocation of RBPs, alternative splicing, and PWS pathogenesis.

The observed outcomes in PWS region lncRNA-depleted hESCs are subtle. This is indeed consistent with the fact that PWS patients lacking these lncRNAs are viable. At the molecular level, such changes could be due to the requirement of the cooperativity and synergy between distinct RBPs that act in alternative splicing regulation. For example, we have observed that some alternative splicing events were co-regulated by both TDP43 and RBFOX2 (Figure 7 and Table S5). Alternatively, these RBPs might simply be poised in undifferentiated hESCs but could be more actively involved in a greater degree of splicing regulation upon differentiation. Future studies are warranted to investigate roles of these lncRNAs upon differentiation.

Finally, the PWS locus is very complex, with the expression of a number of different noncoding RNAs (*SPAs*, *sno-lncRNAs*, and snoRNAs) (Figures 1A and 3D). These RNAs are abundant and processed by unusual pathways. Understanding PWS pathology will thus require isolating the individual contributions of these transcripts, and the current work represents important progress in this direction. More work is needed to identify additional proteins, RNAs, and DNAs in this lncRNA-enriched nuclear accumulation.

## EXPERIMENTAL PROCEDURES

### RIP

Native RIP was carried out as described (Yeo et al., 2009; Yin et al., 2012). See the Supplemental Experimental Procedures for details.

### RNA FISH

RNA FISH was carried out as described (Yin et al., 2012), and images were acquired on a DeltaVision Elite imaging system. See the Supplemental Experimental Procedures for details.

### 3D SIM

3D SIM on fixed samples was performed on a Deltavision OMX V4 system (GE Healthcare). SI image stacks were captured with a z-distance of 0.125  $\mu\text{m}$  and with 5 phases, 3 angles, and 15 raw images per plane. The raw data were reconstructed with channel-specific OTFs, and a Wiener filter was set to 0.01 for DAPI channel and 0.002 for other channels using softWoRx 6.5. Images were registered with alignment parameters obtained from calibration measurements with 100 nm diameter TetraSpeck beads (Thermo Scientific). See the Supplemental Experimental Procedures for details.

(G) Validation of altered cassette exon in P-KO H9 cells. Semi-quantitative RT-PCRs were performed with total RNA from WT and P-KO H9 cells, as well as normal and PWS iPSCs.

(H) A proposed model. *SPAs* and *sno-lncRNAs* from the PWS deletion region form a striking nuclear accumulation in normal pluripotent cells. These lncRNAs interact with multiple RBPs including TDP43, RBFOX2, and hnRNP M with distinct preferences. HESCs lacking these lncRNAs have altered patterns of RBP binding to pre-mRNAs and alternative splicing. Note that PWS patients don't have these lncRNAs or the lncRNA-enriched nuclear accumulations.

### Generating Allele-Specific SPA KO hESCs

Six sgRNAs (3 targeting upstream and 3 targeting downstream of the deletion) were designed (Table S6) and inserted into one CRISPR/Cas9-nuclease vector px330A by Multiplex CRISPR/Cas9 Assembly System Kit (Addgene) (Sakuma et al., 2014). hESCs were transfected with px330A-SPA using FuGENE HD Transfection Reagent (Roche), and clones were picked up in 2 weeks. Genomic DNA and total RNA were extracted to validate P-KO hESCs with primers listed in Table S6.

### RNA Pull Down

Biotinylated RNA pull down was performed as described (Zhang et al., 2013). The co-immunoprecipitated proteins were analyzed by western blotting. See the Supplemental Experimental Procedures for details.

### iCLIP

iCLIP assays were performed as described (König et al., 2010) with modifications. UV cross-linked immunoprecipitation was carried out with different primary antibodies. Linearized iCLIP cDNA libraries were subjected to high-throughput deep sequencing. See the Supplemental Experimental Procedures for details.

### Alternative Splicing Analysis

Alternative splicing events were identified by comparing PSI (percentage of spliced in) values of cassette exons between different samples. PSI values were calculated using MISO (mixture of isoforms). See the Supplemental Experimental Procedures for details.

### High-Throughput Sequencing

Detailed information about RNA-seq, RIP-seq, and iCLIP-seq is summarized in Table S7.

### ACCESSION NUMBERS

The accession numbers for the sequencing data reported in this study are NCBI Sequence Read Archive: GSE85851, GSE85852, and GSE85854.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2016.10.007>.

### AUTHOR CONTRIBUTIONS

L.-L.C. conceived and supervised the project. H.W. and Q.-F.Y. designed and conducted experiments and analyzed data with the help of C.-C.Z., J.Z., and J.-F.X.; R.-W.Y. conducted images and related analyses; Z.L. and L.Y. analyzed RNA-seq data. L.-L.C. wrote the paper with input from all authors.

### ACKNOWLEDGMENTS

We are grateful to G. Carmichael for reading the manuscript, M. Esteban for normal and PWS iPSCs, X.-O. Zhang and W. Xue for computational support, the Core Facility Center of the Institute of Plant Physiology and Ecology, SIBS, CAS for technical support on 3D SIM, and all lab members for discussion. H9 cells were obtained from the WiCell Research Institute. This work was supported by grants 2016YFA0100701 and 2014CB964802 from MOST and 91440202, 31471241, 31322018, and 31271376 from NSFC.

Received: April 7, 2016

Revised: July 22, 2016

Accepted: October 4, 2016

Published: October 27, 2016

### REFERENCES

- Bazeley, P.S., Shepelev, V., Talebizadeh, Z., Butler, M.G., Fedorova, L., Filatov, V., and Fedorov, A. (2008). snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. *Gene* 408, 172–179.
- Bieth, E., Eddiry, S., Gaston, V., Lorenzini, F., Buffet, A., Conte Aurioi, F., Molinas, C., Cailley, D., Rooryck, C., Arveiler, B., et al. (2015). Highly restricted deletion of the SNORD116 region is implicated in Prader-Willi Syndrome. *Eur. J. Hum. Genet.* 23, 252–255.
- Blumenthal, T. (1998). Gene clusters and polycistronic transcription in eukaryotes. *BioEssays* 20, 480–487.
- Boisvert, F.M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* 8, 574–585.
- Bonnet, S.A.D., Akad, D.S., Samaddar, T., Liu, Y., Huang, X., Dong, Y., and Schlüter, O.M. (2013). Synaptic state-dependent functional interplay between postsynaptic density-95 and synapse-associated protein 102. *J. Neurosci.* 33, 13398–13409.
- Bratkovič, T., and Rogelj, B. (2014). The many faces of small nucleolar RNAs. *Biochim. Biophys. Acta* 1839, 438–443.
- Brubaker, S.W., Gauthier, A.E., Mills, E.W., Ingolia, N.T., and Kagan, J.C. (2014). A bicistronic MAVS transcript highlights a class of truncated variants in antiviral immunity. *Cell* 156, 800–811.
- 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.
- Cassidy, S.B., Schwartz, S., Miller, J.L., and Driscoll, D.J. (2012). Prader-Willi syndrome. *Genet. Med.* 14, 10–26.
- Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., 3rd, Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.* 28, 2370–2380.
- Cho, S., Moon, H., Loh, T.J., Oh, H.K., Cho, S., Choy, H.E., Song, W.K., Chun, J.S., Zheng, X., and Shen, H. (2014). hnRNP M facilitates exon 7 inclusion of SMN2 pre-mRNA in spinal muscular atrophy by targeting an enhancer on exon 7. *Biochim. Biophys. Acta* 1839, 306–315.
- Cowling, V.H. (2009). Regulation of mRNA cap methylation. *Biochem. J.* 425, 295–302.
- de Smith, A.J., Purmann, C., Walters, R.G., Ellis, R.J., Holder, S.E., Van Haelst, M.M., Brady, A.F., Fairbrother, U.L., Dattani, M., Keogh, J.M., et al. (2009). A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum. Mol. Genet.* 18, 3257–3265.
- Ding, F., Li, H.H., Zhang, S., Solomon, N.M., Camper, S.A., Cohen, P., and Francke, U. (2008). SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS ONE* 3, e1709.
- Duker, A.L., Ballif, B.C., Bawle, E.V., Person, R.E., Mahadevan, S., Alliman, S., Thompson, R., Traylor, R., Bejjani, B.A., Shaffer, L.G., et al. (2010). Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur. J. Hum. Genet.* 18, 1196–1201.
- Filipowicz, W., and Pogacić, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* 14, 319–327.
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* 28, 2663–2676.
- Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol. Cell* 60, 256–267.
- Glenn, C.C., Saitoh, S., Jong, M.T.C., Filbrandt, M.M., Surti, U., Driscoll, D.J., and Nicholls, R.D. (1996). Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. *Am. J. Hum. Genet.* 58, 335–346.

- Gray, T.A., Saitoh, S., and Nicholls, R.D. (1999). An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc. Natl. Acad. Sci. USA* *96*, 5616–5621.
- Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* *7*, 167–178.
- Kaesler, P.S., Deng, L., Fan, M., and Südhof, T.C. (2012). RIM genes differentially contribute to organizing presynaptic release sites. *Proc. Natl. Acad. Sci. USA* *109*, 11830–11835.
- Kashima, Y., Miki, T., Shibasaki, T., Ozaki, N., Miyazaki, M., Yano, H., and Seino, S. (2001). Critical role of cAMP-GEFII-Rim2 complex in incretin-potenti-ated insulin secretion. *J. Biol. Chem.* *276*, 46046–46053.
- 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.
- Kishore, S., and Stamm, S. (2006). Regulation of alternative splicing by snoRNAs. *Cold Spring Harb. Symp. Quant. Biol.* *71*, 329–334.
- Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* *20*, 3617–3622.
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* *17*, 909–915.
- Lagier-Tourenne, C., Polymenidou, M., Hutt, K.R., Vu, A.Q., Baughn, M., Huelga, S.C., Clutario, K.M., Ling, S.C., Liang, T.Y., Mazur, C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* *15*, 1488–1497.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* *444*, 953–956.
- Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* *8*, 209–220.
- Park, E., Iaccarino, C., Lee, J., Kwon, I., Baik, S.M., Kim, M., Seong, J.Y., Son, G.H., Borrelli, E., and Kim, K. (2011). Regulatory roles of heterogeneous nuclear ribonucleoprotein M and Nova-1 protein in alternative splicing of dopamine D2 receptor pre-mRNA. *J. Biol. Chem.* *286*, 25301–25308.
- Parkinson, W., Dear, M.L., Rushton, E., and Broadie, K. (2013). N-glycosylation requirements in neuromuscular synaptogenesis. *Development* *140*, 4970–4981.
- Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. *Genes Dev.* *20*, 1050–1056.
- Ross, J. (1995). mRNA stability in mammalian cells. *Microbiol. Rev.* *59*, 423–450.
- Runte, M., Hüttenhofer, A., Gross, S., Kieffmann, M., Horsthemke, B., and Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* *10*, 2687–2700.
- Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W., and Beaudet, A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.* *40*, 719–721.
- Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* *4*, 5400.
- Shi, Y., Di Giandomartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* *33*, 365–376.
- Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Zupunski, V., et al. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* *14*, 452–458.
- Tsai, T.F., Jiang, Y.H., Bressler, J., Armstrong, D., and Beaudet, A.L. (1999). Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader-Willi syndrome. *Hum. Mol. Genet.* *8*, 1357–1364.
- Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* *147*, 1537–1550.
- Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* *474*, 380–384.
- Wang, E.T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* *456*, 470–476.
- West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5'→3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* *432*, 522–525.
- Wevrick, R., and Francke, U. (1997). An imprinted mouse transcript homologous to the human imprinted in Prader-Willi syndrome (IPW) gene. *Hum. Mol. Genet.* *6*, 325–332.
- Weyn-Vanhenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* *6*, 1139–1152.
- Yang, J., Cai, J., Zhang, Y., Wang, X., Li, W., Xu, J., Li, F., Guo, X., Deng, K., Zhong, M., et al. (2010). Induced pluripotent stem cells can be used to model the genomic imprinting disorder Prader-Willi syndrome. *J. Biol. Chem.* *285*, 40303–40311.
- 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, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L., and Chen, L.L. (2013). Circular intronic long noncoding RNAs. *Mol. Cell* *51*, 792–806.
- Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.L., Yang, L., and Chen, L.L. (2016). The Biogenesis of Nascent Circular RNAs. *Cell Rep.* *15*, 611–624.