

UC San Diego

UC San Diego Previously Published Works

Title

Sm complex assembly and 5' cap trimethylation promote selective processing of snRNAs by the 3' exonuclease TOE1

Permalink

<https://escholarship.org/uc/item/82k2w12d>

Journal

Proceedings of the National Academy of Sciences of the United States of America, 121(3)

ISSN

0027-8424

Authors

Ma, Tiantai
Xiong, Erica S
Lardelli, Rea M
[et al.](#)

Publication Date

2024-01-16

DOI

10.1073/pnas.2315259121

Peer reviewed



Sm complex assembly and 5' cap trimethylation promote selective processing of snRNAs by the 3' exonuclease TOE1

Tiantai Ma¹, Erica S. Xiong¹, Rea M. Lardelli^{a,1}, and Jens Lykke-Andersen^{a,1}

Edited by Roy Parker, University of Colorado, Boulder, CO; received September 1, 2023; accepted December 6, 2023

Competing exonucleases that promote 3' end maturation or degradation direct quality control of small non-coding RNAs, but how these enzymes distinguish normal from aberrant RNAs is poorly understood. The Pontocerebellar Hypoplasia 7 (PCH7)-associated 3' exonuclease TOE1 promotes maturation of canonical small nuclear RNAs (snRNAs). Here, we demonstrate that TOE1 achieves specificity toward canonical snRNAs through their Sm complex assembly and cap trimethylation, two features that distinguish snRNAs undergoing correct biogenesis from other small non-coding RNAs. Indeed, disruption of Sm complex assembly via snRNA mutations or protein depletions obstructs snRNA processing by TOE1, and *in vitro* snRNA processing by TOE1 is stimulated by a trimethylated cap. An unstable snRNA variant that normally fails to undergo maturation becomes fully processed by TOE1 when its degenerate Sm binding motif is converted into a canonical one. Our findings uncover the molecular basis for how TOE1 distinguishes snRNAs from other small non-coding RNAs and explain how TOE1 promotes maturation specifically of canonical snRNAs undergoing proper processing.

TOE1 | snRNA | 3' end processing | Sm complex | 5' cap trimethylation

Small non-coding (nc)RNAs play key roles at all levels of gene regulation. In eukaryotes, the majority of small ncRNAs are transcribed as precursor RNAs that need to undergo further processing to become functional mature RNAs, including trimming of 3' end extensions (1, 2). These extensions serve as central hubs for quality control, where competing 3'-to-5' exonucleases that promote maturation or degradation dictate whether the small ncRNAs undergo processing to functional molecules (3–6) or are subjected to decay (5, 7, 8). The competition between these processes can be influenced by post-transcriptional oligo(A) or -(U) tailing by terminal nucleotidyl transferases (5, 9–12). A key question is what dictates the specificity of the exonucleases and terminal nucleotidyl transferases that act on small ncRNAs to ultimately control their fate. A prevailing hypothesis is that enzymes that promote maturation recognize specific features of RNAs undergoing proper biogenesis, while more promiscuous decay enzymes degrade those RNAs that fail to conform to canonical processing. This predicts that maturation enzymes recognize common features of canonical RNAs that signify their correct biogenesis.

Recent studies have uncovered several 3' to 5' exonucleases that promote deadenylation and 3' end maturation of human small ncRNAs. Two such exonucleases are the DEDD family deadenylase target of EGR1 protein 1 (TOE1; also known as CAF1Z) (13) and poly(A)-specific ribonuclease (PARN) (14). While TOE1 and PARN are homologous proteins and are both activated by oligo(A) tails (13, 14), they differ in their specificity for small ncRNAs. PARN has been reported to process 3' ends of several types of small ncRNAs, including small nucleolar RNAs (snoRNAs) (4, 15) and the telomerase RNA component (TERC) (16–18). By contrast, TOE1 is known to process 3' ends of RNA Polymerase II-transcribed small nuclear RNAs (Pol II snRNAs; i.e., all snRNAs except U6 and U6atac) (3, 5), and has also been reported to process some snoRNAs, scaRNAs, and TERC (4, 19). Yet another DEDD deadenylase, the germ-cell specific PNLDC1, has been implicated in processing of Piwi-interacting (pi)RNAs (20–22), and USB1, an exonuclease of the 2H phosphodiesterase family, is known to process U6 and U6atac snRNAs and some miRNAs (23–25). Functions of these enzymes are central to human health, with TOE1 mutations associated with *Pontocerebellar Hypoplasia Type 7* (PCH7) (3), PARN mutations with *Dyskeratosis Congenita* and other human disorders (26, 27), PNLDC1 mutations with *azoospermia* (28), and USB1 mutations with *Poikiloderma with neutropenia* (29). Despite their importance in RNA metabolism and human disease, little is known about the molecular basis for the RNA specificity of these exonucleases.

snRNAs are central components of the spliceosome that carries out pre-mRNA splicing. The maturation of Pol II snRNAs includes multiple steps in the nucleus and cytoplasm (30). During transcription, a 7-methyl guanosine (m7G) cap is co-transcriptionally added

Significance

RNA quality control pathways rid cells of defective RNAs by mechanisms that are often poorly understood. In the case of human small nuclear (sn)RNAs, core components of the spliceosome, previous studies identified a central role in quality control of competing 3'-to-5' exonucleases that promote either 3' end maturation or degradation. However, how these exonucleases distinguish functional from defective snRNAs has remained unknown. Here, we demonstrate that the 3' end maturation exonuclease TOE1 distinguishes functional from defective snRNAs through two snRNA features, Sm-complex assembly and cap trimethylation, that signify successful snRNA biogenesis. These findings suggest that snRNA quality control relies on the specificity of TOE1 for correctly assembled snRNAs, leaving defective snRNAs to be degraded by more promiscuous decay exonucleases.

Preprint Servers: This manuscript has been deposited to bioRxiv as a preprint (doi: <https://doi.org/10.1101/2023.08.15.553431>) under a CC-BY-NC-ND 4.0 International license.

Author contributions: T.M., R.M.L., and J.L.-A. designed research; T.M., E.S.X., and R.M.L. performed research; T.M., E.S.X., R.M.L., and J.L.-A. analyzed data; and T.M. and J.L.-A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: reamartine@gmail.com or jlykkeandersen@ucsd.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2315259121/-DCSupplemental>.

Published January 9, 2024.

to the snRNA 5' end and the 3' end is cleaved by the integrator complex, leaving a short 3' end tail (31, 32). The m7G cap subsequently promotes nuclear export via the export factor PHAX in association with the nuclear cap-binding complex (CBC) (33). In the cytoplasm, the Sm complex is loaded onto the snRNAs with the help of the SMN-Gemin2 complex, and protein arginine methyl transferases (PRMTs) (34–42), and the m7G cap is hypermethylated to a trimethylguanosine (TMG) cap by trimethylguanosine synthase 1 (TGS1) (43–45). The Sm complex and the TMG cap are subsequently recognized by Snurportin (SNUPN), which promotes nuclear import via Importin β (46, 47). Following nuclear import, the snRNAs undergo nucleotide modifications directed by scaRNAs and assemble with other snRNA-protein complexes (snRNPs) to form the spliceosome (48).

Evidence suggests that TOE1 acts on Pol II snRNAs at least twice during their maturation (5). During early biogenesis, prior to or upon nuclear export, TOE1 partially trims Pol II snRNA 3' ends in a process that involves oligoadenylation (5). The remainder of the tail is removed later, during or after nuclear import, when TOE1 acts on snRNAs a second time (5, 49). While TOE1 acts on all canonical Pol II snRNAs, tested unstable snRNA variants transcribed from snRNA pseudogenes and mutant U1 snRNA deleted of its Sm binding site escape 3' end processing and are instead subjected to decay by degrading exonucleases (5, 12, 49, 50). The features of canonical snRNAs recognized by TOE1 during early and late processing steps and how TOE1 distinguishes canonical snRNAs from unstable snRNA variants and other small ncRNAs of the cell has remained unknown.

In this study, we addressed how TOE1 achieves specificity toward canonical Pol II snRNAs. Global 3' end sequencing of newly transcribed small RNAs in the absence or presence of TOE1 confirms that TOE1 specifically processes canonical Pol II snRNAs over other classes of small RNAs. Dissecting the features of Pol II snRNPs recognized by TOE1 revealed the Sm complex and the TMG cap as two key features characteristic of Pol II snRNPs that promote TOE1 processing. An unstable U1 snRNA variant known to escape TOE1-processing, U1v15, is fully processed by TOE1 when the variant Sm binding motif is converted into that of canonical snRNAs. Our findings demonstrate that TOE1-mediated snRNA maturation is driven by Sm complex assembly and cap trimethylation, features that are specific to canonical Pol II snRNAs undergoing proper biogenesis.

Results

TOE1 Specifically Processes Pol II snRNAs. RNAs affected at steady state by TOE1 depletion have been previously globally monitored (4). Since the processing of stable RNAs is better captured in the newly transcribed RNA population, to further delineate the repertoire of small RNA targets of TOE1, we isolated newly transcribed small RNAs, 100 to 500 nucleotides in length, from human embryonic kidney 293 T (HEK293T) cells under normal or TOE1-depleted conditions and subjected them to global 3' end sequencing (Fig. 1*A* and *SI Appendix, Fig. S1 A and B*) (51). Comparing mean genomic-encoded tail lengths of small RNAs in the absence or presence of TOE1 revealed Pol II snRNAs as the primary targets of TOE1-mediated 3' end processing (Fig. 1*B* and *C*). A subset of Pol II snRNAs also accumulated short oligo(A) tails in the absence of TOE1 (Fig. 1*D* and *E*). Only one small ncRNA that was not an snRNA, scaRNA20, was observed to be significantly extended in the absence of TOE1 (Fig. 1*B* and *SI Appendix, Fig. S1C*), primarily due to an extended oligo(A) tail (*SI Appendix, Fig. S1D*). While our global sequencing assays captured the majority of Pol II snRNAs as targets of TOE1,

U5 and U4atac snRNAs were of too low abundance in the sequencing assays to include in the analysis; however, these have both previously been identified as TOE1 targets in gene-specific sequencing experiments (3, 5). Collectively, our global 3' end sequencing analyses of newly transcribed small RNAs show TOE1 specificity toward Pol II snRNAs as a class, although TOE1 may also act on a subset of other small ncRNAs, including scaRNA20, consistent with previous reports (4, 19). This raised the question of how TOE1 achieves specificity for Pol II snRNAs over other small ncRNAs.

Mutations in Sm Complex and U1-70 K Binding Motifs of U1 snRNA Impair TOE1-Mediated 3' End Processing. To investigate how TOE1 achieves substrate specificity, we turned to U1 snRNA (Fig. 2*A* and *SI Appendix, Fig. S2A*). Reasoning that TOE1 may recognize specific protein components of the U1 snRNP, we introduced previously described mutations that disrupt the interaction of U1 snRNA with U1A, U1-70 K, and the Sm complex (Fig. 2*B*) (40, 52, 53). These mutations were introduced into a bar-coded exogenous U1 snRNA (3, 50), which allowed us to monitor their effect on U1 snRNA processing by 3' end sequencing. Consistent with early observations in *Xenopus* oocytes (49), disruption of the Sm complex binding motif led to a strong defect in U1 snRNA 3' end processing (Fig. 2*C* and *D*). In addition, mutating the U1-70 K binding motif caused a minor, but statistically significant, defect in U1 snRNA 3' end processing, whereas disrupting the U1A binding motif did not impair processing (Fig. 2*C* and *D*). Further analyses of the mutant U1 snRNA 3' ends revealed that the disruption of the Sm complex binding motif caused a significant fraction of the U1 snRNA population to accumulate with oligo(U) tails (Fig. 2*E* and *F*), consistent with recent observations (12). Interestingly, U1 snRNA mutated for U1-70 K binding also accumulated oligo(U) tails, and both Sm and U1-70 K binding mutants showed oligo(A) tailing as well (Fig. 2*E* and *G*). Monitoring 3' end processing of the mutant U1 snRNAs in the presence or absence of TOE1 depletion revealed that disruption of the Sm complex binding motif fully abolished the ability of TOE1 to process U1 snRNA (Fig. 2*H* and *I*), whereas processing by TOE1 is only partially defective for the U1-70 K binding mutant (*SI Appendix, Fig. S2 B and C*).

Sm Complex Depletion Impairs 3' End Processing of Multiple Pol II snRNAs. The experiments above were performed using exogenous bar-coded U1 snRNAs. To verify the importance of snRNP proteins in snRNA processing, we next monitored effects of their depletion on processing of endogenous snRNAs. Sm complex assembly is a feature characteristic of all Pol II snRNAs. To test whether Sm complex assembly is necessary for 3' end processing of Pol II snRNAs in general, we depleted the Sm complex component SmB (*SI Appendix, Fig. S3A*) and performed targeted 3' end sequencing of newly transcribed Pol II snRNAs. All tested Pol II snRNAs showed defects in 3' end processing following SmB depletion as compared to the negative control U3 snoRNA, although some snRNAs were affected more than others (Fig. 3*A* and *SI Appendix, Fig. S3*). U1 and U2 snRNAs showed strong defects in 3' end processing, whereas U4, U5, and U4atac snRNAs showed significant, but less extensive, defects. A majority of the Pol II snRNAs also accumulated oligo(U) tails to various degrees upon SmB depletion (Fig. 3*B* and *C*). These tails were most prominent for U1, U2, and U5 snRNAs and observed at much lower levels for U4 and U4atac snRNAs (Fig. 3*B* and *C* and *SI Appendix, Fig. S3*). The control U3 snoRNA showed a very minor increase in uridylation upon SmB depletion (Fig. 3*B* and *C*). Collectively, our findings demonstrate a general role for the Sm complex in 3'

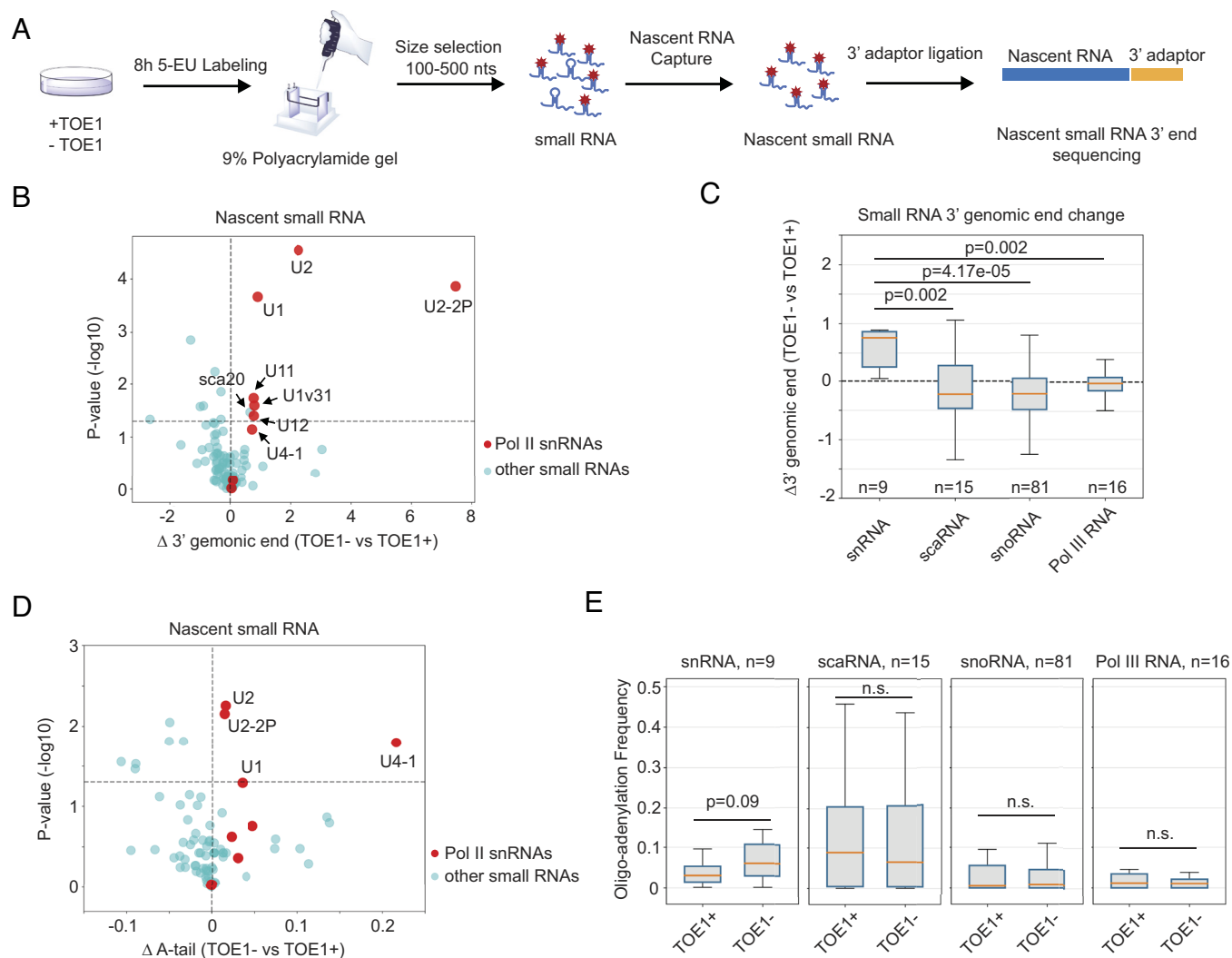


Fig. 1. TOE1 shows specificity toward Pol II snRNAs. (A) Schematic of the newly transcribed small RNA global 3' end sequencing workflow. Control (TOE1+) or TOE1 depleted (TOE1-) cells were metabolically labeled with 5-ethynyl uridine (5-EU) for 8 h followed by size selection and nascent RNA capture. Newly transcribed RNA 3' ends were determined by global small RNA sequencing. (B) Scatter plot showing changes in nascent small RNA 3' end lengths in TOE1- versus TOE1+ cells and corresponding *P*-values. Each dot represents an individual small ncRNA with Pol II snRNAs in red and other small ncRNAs in cyan. *P*-values were calculated from three individual biological repeats by Student's two-tailed *t*-test and plotted on a $-\log_{10}$ scale. The horizontal dashed line represents *P* = 0.05. Small ncRNAs on the *Right* side of the plot have extended 3' end tails upon TOE1 depletion. (C) Box and whisker plots showing changes in 3' end lengths of groups of related small ncRNAs in TOE1- versus TOE1+ cells. The number (*n*) of RNAs in each group is indicated. Boxes represent quartiles around median values shown as orange lines, with whiskers extending to maximum and minimum values within 1.5 times the interquartile range from the boxes. *P*-values were calculated by comparing each group of small ncRNAs to the group of Pol II snRNAs using the two-sample Kolmogorov-Smirnov test. (D) Scatter plot of changes in frequencies of nascent small RNAs containing 3' end posttranscriptional oligo(A) tails, defined as two or more post-transcriptional adenosines, in TOE1- versus TOE1+ cells. Dots and axes are labeled as in panel B. (E) Box and whisker plots showing distributions of 3' end oligoadenylation (two or more adenosines) frequencies for each group of small ncRNAs in TOE1+ or TOE1- cells. Shown *P*-values were calculated using Student's two-tailed *t*-test comparing TOE1- to TOE1+ conditions.

end processing of Pol II snRNAs, although some snRNAs are more sensitive to Sm complex levels than others.

U1-70 K Compensates for a Suboptimal Sm Binding Motif to Stimulate U1 snRNA 3' End Processing.

In contrast to the Sm complex, which assembles with all Pol II snRNAs, U1-70 K is unique to U1 snRNA. Previous studies have shown that U1-70 K helps promote assembly of the Sm complex onto U1 snRNA, which contains a suboptimal Sm binding site (34) (Fig. 4A). Consistent with the effect of the U1-70 K binding-site mutation, depletion of the U1-70 K protein impaired 3' end processing of newly transcribed U1 snRNA and caused low-level oligo(A) and oligo(U) tailing (Fig. 4 B-E and *SI Appendix*, Fig. S4A). To test whether U1-70 K promotes U1 snRNA 3' end processing via its stimulation of Sm complex assembly, we tested the effect of converting the suboptimal U1 snRNA Sm binding motif into a

canonical one (Fig. 4F, superU1), which is known to promote assembly of the Sm complex independently of U1-70 K (34). The canonical Sm binding site rescued the 3' end processing defect of the U1-70 K binding site mutation and strongly reduced adenylation and uridylation (Fig. 4 G-I). Thus, U1-70 K promotes U1 snRNA 3' end processing by compensating for a suboptimal Sm complex binding site. In contrast to U1-70 K and the Sm complex, depletion of U1 snRNP components U1A and U1C showed no defect in U1 snRNA 3' end processing (*SI Appendix*, Fig. S4 B-G), consistent with our observations with the U1A-binding mutant U1 snRNA (Fig. 2).

The U1 snRNA Variant U1v15 Escapes TOE1 Processing due to Variant Nucleotides in Its Sm-Binding Motif.

The human U1 snRNA variant U1v15 is known to evade TOE1 recognition and undergo rapid degradation (5). To identify the nucleotide

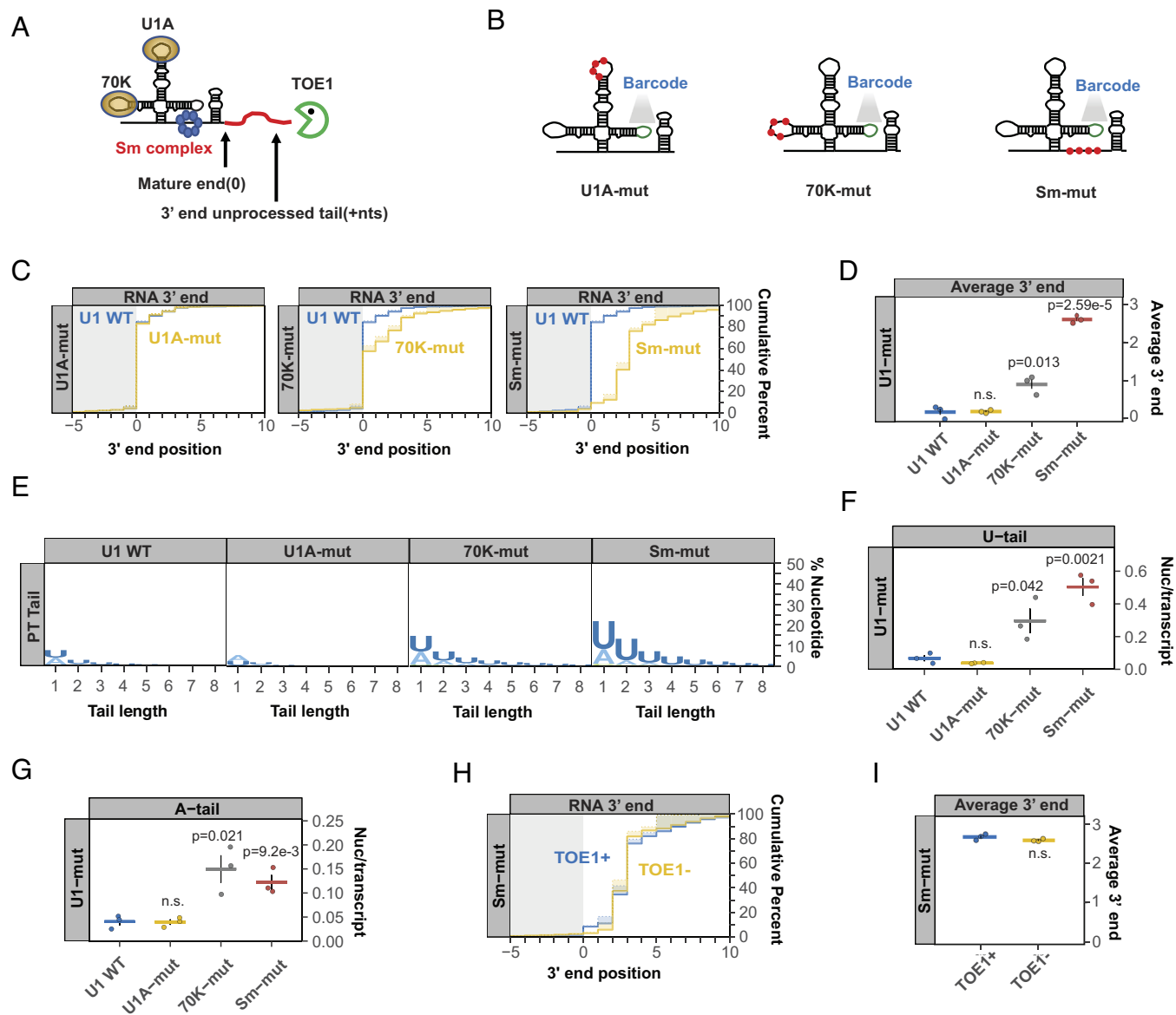


Fig. 2. Mutations in Sm complex- and U1-70 K-binding motifs of U1 snRNA inhibit 3' end processing. (A) Schematic of the U1 snRNP with 70 K, U1A, and Sm complex components shown at their binding motifs and a red line representing a 3' end unprocessed tail. The snRNP schematic is not proportional to size. A proportional U1 snRNA-Sm complex model (54) can be found in *SI Appendix, Fig. S2A*. (B) Schematics of barcoded U1 snRNAs with mutations known to abolish binding of U1A (U1A-mut), U1-70 K (70 K-mut) or the Sm complex (Sm-mut) (*SI Appendix, Table S2*) shown as red dots. A barcode sequence in stem loop 3 is indicated in green. (C) Cumulative plots showing distributions of 3' end positions for each U1 snRNA mutant (yellow) as compared to wild-type U1 snRNA (U1 WT, blue) as determined by U1 snRNA gene-specific 3' end sequencing of steady-state exogenous snRNAs. Mature 3' ends are represented as position 0 with positive numbers representing unprocessed 3' end tails. Solid lines represent actual snRNA 3' end positions, while dotted lines represent genome-encoded 3' end positions with shading in between representing post-transcriptionally added nucleotides. Only RNAs terminating at or downstream of the -10 position were analyzed. Three individual biological repeats were averaged for each plot. (D) Average 3' end positions for each U1 WT or mutant snRNAs with each dot representing an individual biological repeat. Black vertical lines represent SEM and *P*-values were calculated using Student's two-tailed *t*-test comparing U1 mutants to U1 WT (*P* > 0.05 is noted as n.s.). (E) Sequence logo plots representing percentages of post-transcriptional added nucleotides for U1 WT and U1 mutant snRNAs. (F and G) Average post-transcriptional added uridines (F) and adenosines (G) per transcript for U1 WT and U1 mutants plotted as in panel D. (H) Cumulative plot showing 3' end distributions for the Sm mutant U1 snRNA under TOE1+ (blue) or TOE1- (yellow) conditions. (I) Average 3' end positions for the Sm mutant U1 snRNA in TOE1+ (blue) and TOE1- (yellow) conditions.

variations within U1v15 RNA that inhibit processing by TOE1, we introduced variant nucleotides found within the U1A or Sm binding motifs of U1v15 into the bar-coded canonical U1 snRNA (Fig. 5A). The variant nucleotides of the U1v15 U1A binding motif showed no effect on 3' end processing when introduced into the canonical U1 snRNA (Fig. 5B and C). By contrast, the variant nucleotides of the Sm binding motif were sufficient to trigger a 3' end processing defect similar to that seen for the variant U1v15 snRNA (Fig. 5B and C), and promote similar levels of oligo(U) and -(A) tailing (Fig. 5D–F). We next converted the variant Sm binding site of U1v15 RNA

into the Sm binding motif found in U1 snRNA (Fig. 5A, v15-SmWT) or the canonical Sm binding motif (Fig. 5A, v15-super), while leaving all other variant nucleotides intact. These modifications rescued 3' end processing of the U1v15 snRNA variant (Fig. 5B and C), although a minor residual level of mono-adenylation and -uridylation could be observed (Fig. 5D–F). Taken together, these observations show that an intact Sm complex binding site is not only necessary for the processing of canonical Pol II snRNAs by TOE1, but can also rescue the processing of a normally unprocessed and unstable snRNA variant.

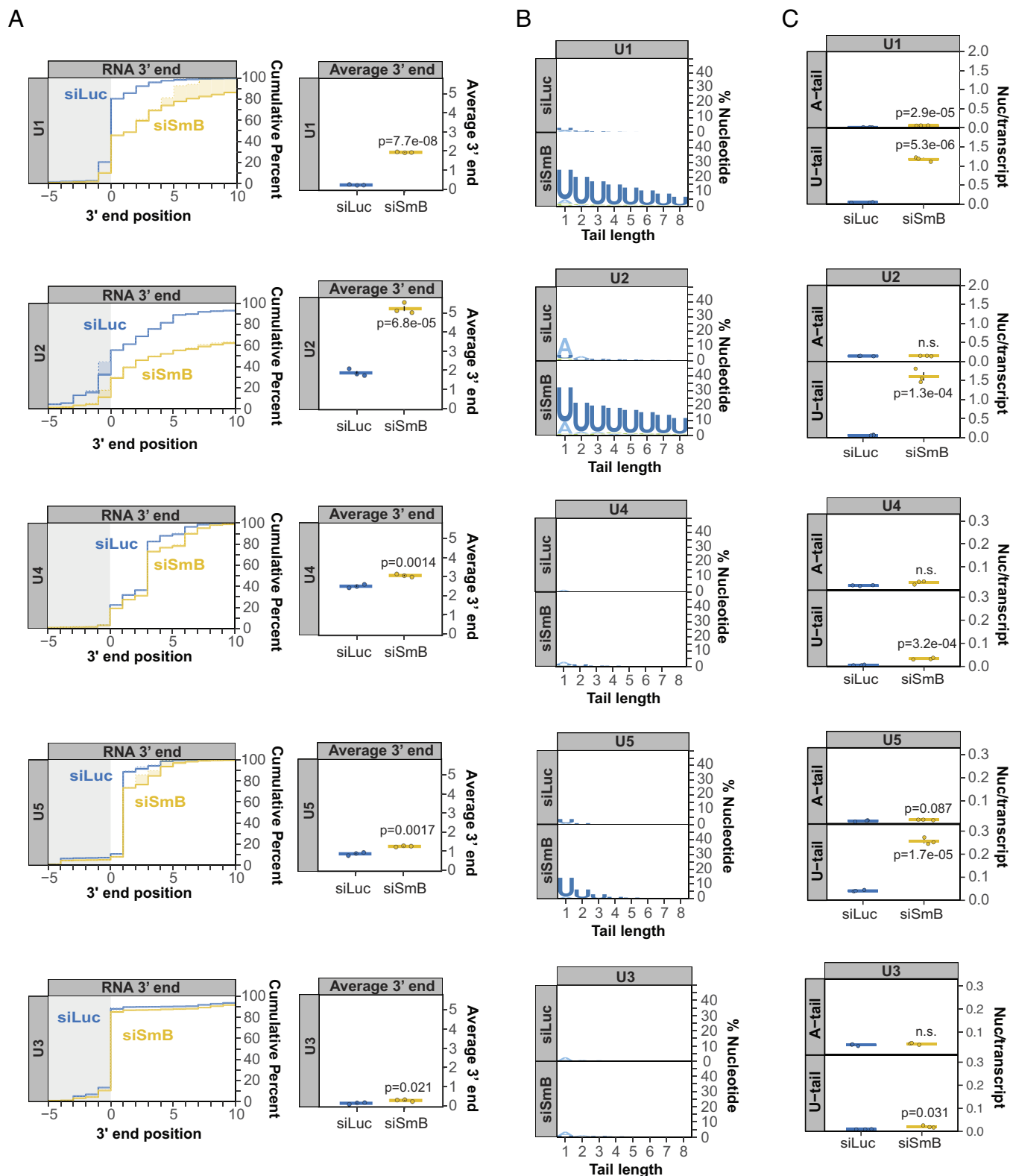


Fig. 3. Sm complex depletion impairs 3' end processing of multiple Pol II snRNAs. (A) Cumulative plots showing 3' end distributions of newly transcribed U1, U2, U4, and U5 snRNAs in control (siLuc; blue) versus SmB (siSmB; yellow) knock-down conditions. U3 snoRNA serves as a negative control. The average of three individual biological repeats is plotted and average 3' end positions are graphed on the *Right* with dots representing individual biological repeats and SEMs shown as vertical black lines. *P*-values were calculated by Student's two-tailed *t*-test. (B) Sequence logo plots representing the percentage of post-transcriptionally added nucleotides for snRNAs and U3 snoRNA under control (siLuc) or SmB knock-down (siSmB) conditions. (C) Average number of post-transcriptionally added adenosine or uridine nucleotides per transcript for major class snRNAs and U3 snoRNA plotted as in panel A.

TOE1 Directly Recognizes the Sm Complex-Assembled U1 snRNP.
The assembly of the Sm complex onto Pol II snRNAs takes place in the cytoplasm and triggers subsequent 5' cap trimethylation

and nuclear import (43–47) (Fig. 6A). Since TOE1 is known to concentrate in the nucleus (4, 13), our findings raised the question of whether TOE1 directly recognizes the Sm complex-assembled

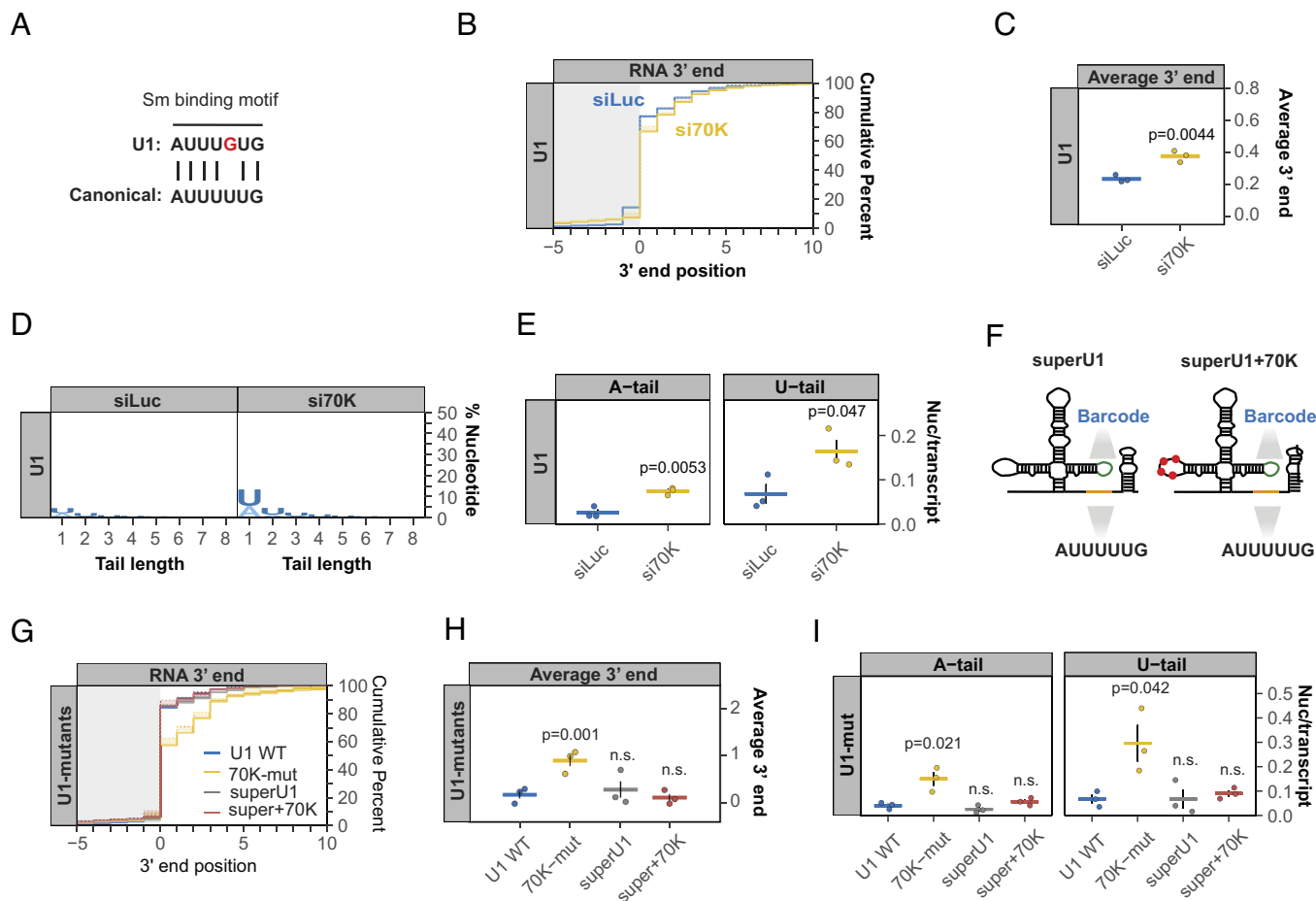


Fig. 4. U1-70 K compensates for a suboptimal Sm binding motif to stimulate U1 snRNA 3' end processing. (A) Comparison of the U1 snRNA Sm binding motif with the canonical Sm binding motifs found in other Pol II snRNAs. (B) Cumulative plot showing the 3' end distributions of newly transcribed U1 snRNA in control (siLuc, blue) or U1-70 K depletion (si70K, yellow) conditions. Averages of three individual biological repeats are plotted. (C) Average newly transcribed U1 snRNA 3' end positions from the experiments plotted in panel B with dots representing individual biological repeats and SEMs shown as vertical black lines. The *P*-value was calculated by Student's two-tailed *t*-test. (D) Sequence logo plot representing the percentage of post-transcriptionally added nucleotides in control (siLuc) or U1-70 K depleted (si70K) conditions. (E) Average post-transcriptionally added adenosines and uridines per transcript plotted as in panel C. (F) Schematics of barcoded exogenous U1 snRNAs with canonical Sm binding motif (superU1) and superU1 with mutations that abolish U1-70 K binding shown as red dots (superU1+70 K). (G) Cumulative plot showing 3' end distributions for indicated exogenous U1 snRNAs at steady state. (H) Average 3' end positions for indicated exogenous U1 snRNAs, plotted as in panel C. (I) Average post-transcriptionally added adenosine and uridine nucleotides per transcript for indicated exogenous U1 snRNAs, plotted as in panel E.

snRNP. Alternatively, the Sm complex could promote TOE1-mediated processing by indirect means such as via snRNA nuclear import. To address this question, we tested the ability of recombinant TOE1 to process wild-type or Sm-mutant snRNPs isolated from cells. A 2'-OMe-RNA oligonucleotide (*SI Appendix, Table S1*) hybridizing to the U1 snRNA 5' splice site recognition motif was used to isolate bar-coded U1 wild-type or Sm-mutant snRNP complexes exogenously expressed in TOE1-depleted HEK293T cells (Fig. 6B and *SI Appendix, Fig. S5*). Isolated snRNPs were subsequently incubated with Flag-tagged TOE1 protein *in vitro* and analyzed by 3' end sequencing. We found that wild-type U1 snRNP was increasingly processed at the 3' end with increasing concentrations of TOE1, but not with a previously characterized catalytically inactive TOE1 mutant, TOE1-DE (3) (Fig. 6C–E). By contrast, the Sm-mutant U1 snRNP showed little processing by TOE1 (Fig. 6F–H). These observations demonstrate that TOE1 directly recognizes the Sm complex-assembled U1 snRNP. This could occur either via direct interaction with the Sm complex or by recognition of snRNP modifications downstream of Sm complex assembly.

snRNA Processing by TOE1 Is Activated by the 5' TMG Cap. In addition to promoting nuclear import, Sm complex assembly

also stimulates trimethylation of the snRNA cap (43–45). Interestingly, the TOE1 homolog PARN is known to be activated by an m7G-monomethyl cap (55–57). To test whether TOE1 activity is affected by the snRNA cap structure, we generated uncapped, m7G-capped, or TMG-capped U1 snRNAs by *in vitro* transcription (*SI Appendix, Fig. S6 A–C*). Each U1 snRNA contained a 20-nucleotide genomic-encoded tail to allow for monitoring of TOE1 activity (U1+20, Fig. 7A). Incubating each of the differently capped U1 snRNAs with Flag-tagged TOE1 isolated after overexpression in HEK293T cells (*SI Appendix, Fig. S6D*), we found that TOE1 processed the 5' TMG-capped U1 snRNA much more efficiently than the corresponding 5' uncapped or m7G-capped U1 snRNA substrates (Fig. 7B and C). This effect was independent of co-purified proteins from human cells, as His-tagged TOE1 isolated from *E. coli* was similarly stimulated by the TMG cap (Fig. 7B and C and *SI Appendix, Fig. S6D*). The TMG cap analog used in this assay produces an RNA population of which at most half are TMG capped (*SI Appendix, Fig. S6 A and B*), which may explain why about half of the RNA population remains unprocessed by TOE1 (Fig. 7B). Collectively, our observations demonstrate that TOE1 achieves substrate specificity through Sm complex assembly and

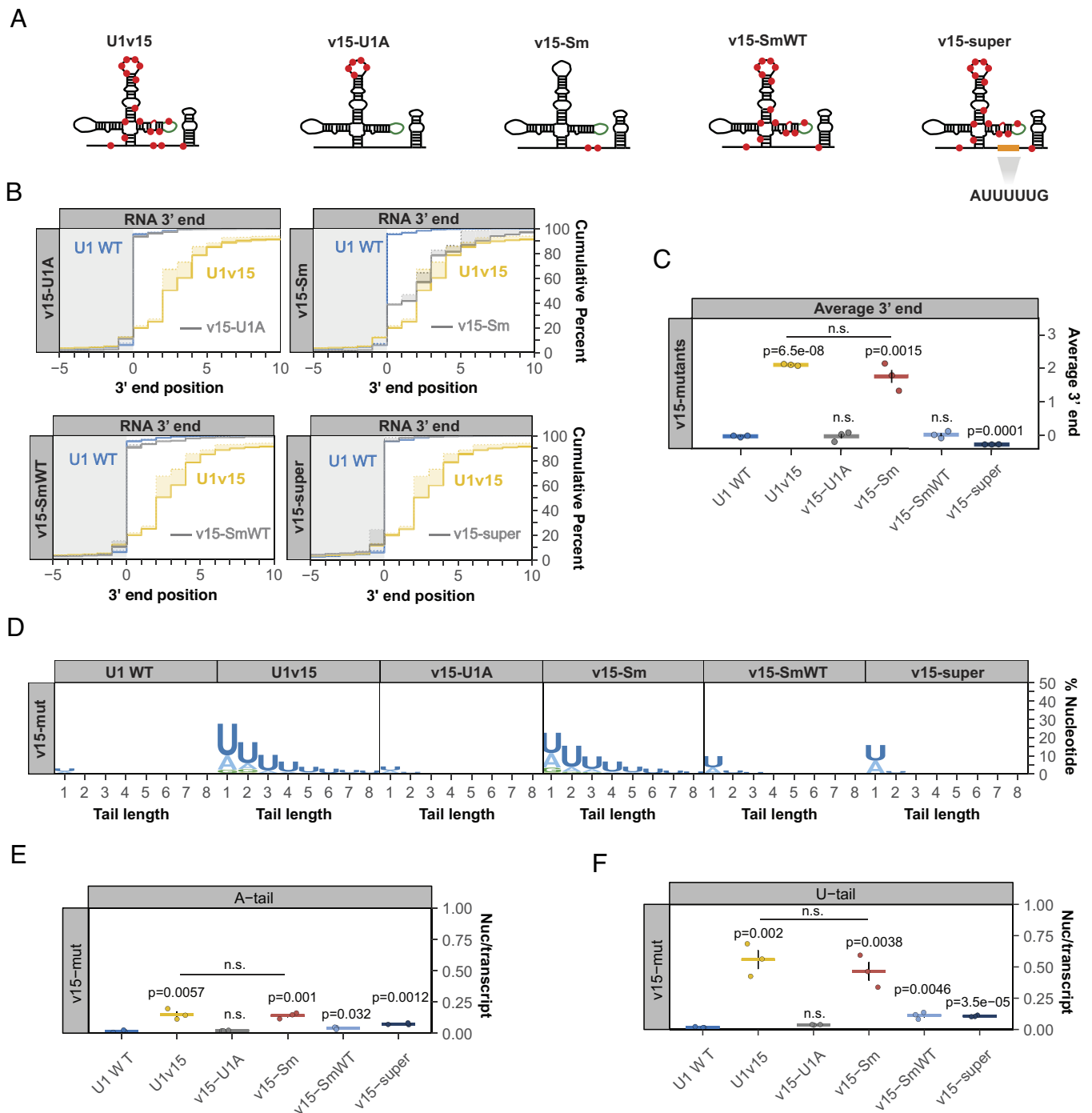


Fig. 5. U1 snRNA variant U1v15 escapes TOE1 processing due to variant nucleotides in its Sm binding motif. (A) Schematic of U1 snRNA variant U1v15 and U1 snRNA mutants (v15-U1A, v15-Sm, v15-SmWT) (*SI Appendix, Table S2*) containing mutations corresponding to U1v15 nucleotide variations indicated by red dots. v15-super corresponds to U1v15 with a canonical Sm binding motif. (B) Cumulative plots showing 3' end distributions for U1 WT (blue), U1v15 (yellow), and each v15 mutant (gray) at steady state. The average of three individual biological repeats is plotted. (C) Average 3' end positions for U1 WT, U1v15, and v15 mutants. Dots represent individual repeats and vertical lines are SEM. *P*-values were calculated by Student's two-tailed *t*-test by comparing v15 mutants to U1 WT. (D) Sequence logo plots representing the percentage of posttranscriptionally added nucleotides for U1 WT, U1v15, and v15 mutants. (E and F) Average post-transcriptionally added adenosines (E) and uridines (F) per transcript plotted as in panel C.

cap trimethylation, two features that distinguish Pol II snRNAs undergoing proper assembly from other small ncRNAs of the cell.

Discussion

Pontocerebellar Hypoplasia 7-associated protein TOE1 is a key 3' end maturation factor for canonical Pol II snRNAs, but how it distinguishes substrate from non-substrate RNAs has remained

unknown. Here, we present evidence that TOE1 selectively processes Pol II snRNAs over other classes of small non-coding RNAs (Fig. 1). This specificity is imparted in part by Sm complex assembly, a distinguishing feature of Pol II snRNAs. Indeed, manipulations known to inhibit Sm complex assembly, including U1 snRNA mutations and nucleotide variations (Figs. 2 and 5), and depletion of an Sm complex component or U1-70 K (Figs. 3 and 4), impair TOE1-mediated snRNA 3' end processing. TOE1

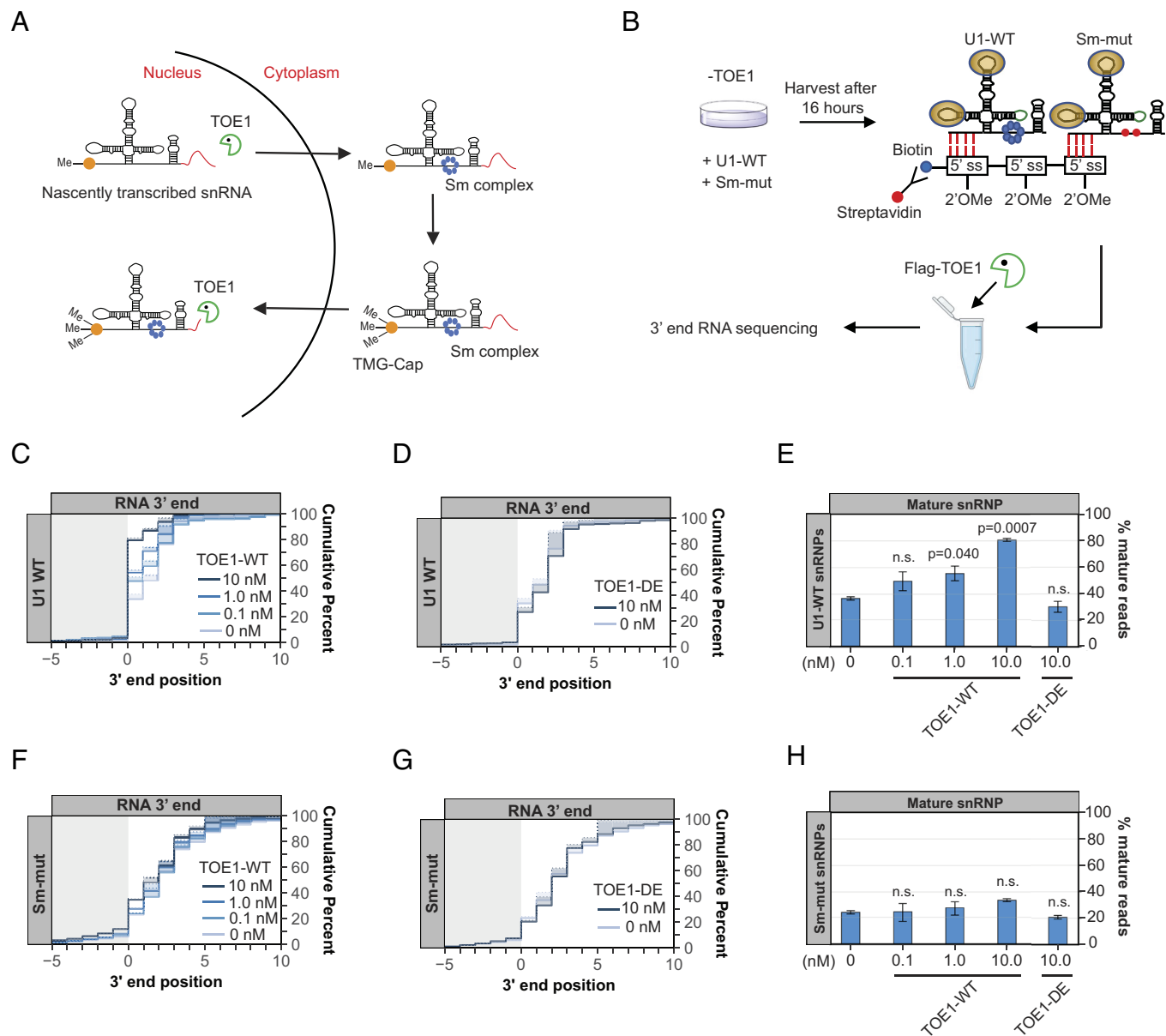


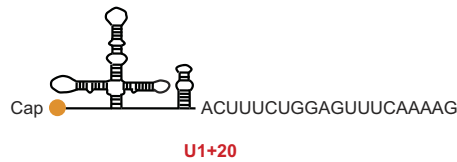
Fig. 6. TOE1 directly recognizes the Sm complex-assembled U1 snRNP. (A) Schematic of the Pol II snRNA biogenesis pathway Sm complex assembly, 5' cap trimethylation, and TOE1-mediated processing steps. The red lines at the 3' end represent 3' end extensions. snRNP models are not proportional to actual size. (B) Schematic of the in vitro U1 snRNP pull-down and TOE1 processing assay. See *Materials and Methods* for details. (C and D) Cumulative plots showing 3' end distributions of wild-type U1 (U1 WT) snRNPs after incubation with TOE1 (panel C) or catalytically inactive TOE1 (TOE1 DE, panel D) at indicated concentrations. The average of two individual biological repeats is plotted. (E) Percentage of fully processed U1 WT snRNPs after TOE1 or TOE1 DE incubation at indicated concentrations. *P*-values were calculated by Student's one-tailed *t*-test by comparing with the 0 nM TOE1 condition (n.s.: $P > 0.05$). (F–H) Same as panels C–E but analyzing the U1 Sm-mut snRNP.

directly recognizes the Sm complex-assembled snRNP as demonstrated by its specificity toward wild-type over Sm-mutant U1 snRNP in an in vitro 3' end processing assay (Fig. 6). This recognition could occur via direct interaction with the Sm complex and/or with a downstream Sm complex-dependent modification of the snRNP. Indeed, TOE1 directly recognizes the trimethylated snRNA cap as evidenced by the specificity of TOE1 for a 5' TMG capped snRNA in vitro (Fig. 7). These findings reveal the molecular basis for how TOE1 distinguishes canonical Pol-II snRNAs from other small ncRNAs of the cell.

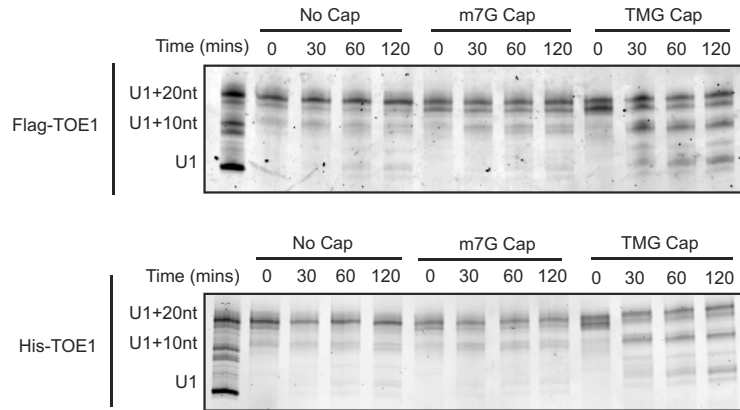
When and where during their maturation does TOE1 process snRNAs? The biogenesis of Pol II snRNAs involves both nuclear and cytoplasmic processes (Fig. 6A). We previously presented evidence that TOE1 acts on snRNAs at least twice during their biogenesis, by first partially processing snRNAs prior to or during

nuclear export, before completing maturation during or after nuclear re-import (5). Our findings in this study provide a molecular basis for how TOE1 achieves specificity toward snRNAs during late-stage biogenesis. At this point, snRNAs have undergone Sm complex assembly and cap trimethylation, the two features shown here to promote processing by TOE1. Since snRNA cap trimethylation occurs as a consequence of Sm complex assembly (43–45), it is possible that this late-stage processing by TOE1 occurs solely through TOE1 recognition of the trimethylated cap. Consistent with a key role for the TMG cap, depletion of TGS1, the enzyme that carries out cap trimethylation, has been reported to lead to accumulation of 3' end extended snRNAs (58, 59). It remains to be determined how the TMG cap stimulates TOE1 activity. We observed no evidence for allosteric activation of TOE1 on an uncapped U1 snRNA by a TMG cap analog added in trans

A



B



C

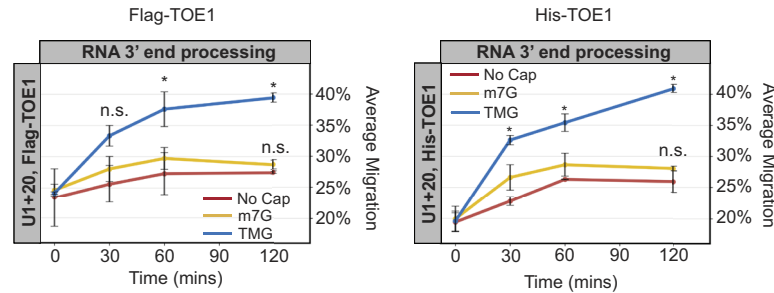


Fig. 7. snRNA processing by TOE1 is stimulated by the 5' TMG cap. (A) Schematic of in vitro transcribed U1 snRNA with a 20-nucleotide (nt) genomic-encoded tail (U1+20). (B) Representative denaturing gels showing uncapped, m7G-, or TMG-capped U1+20 incubated with Flag-TOE1 or His-TOE1 in vitro for indicated times. The RNA ladder consists of U1 snRNA with no tail, a 10-nt, or a 20-nt tail. (C) Quantification of Flag-TOE1 and His-TOE1 processing in panel B. The average RNA migration for each condition was quantified by ImageJ (with the U1+20-nt marker corresponding to 0% migration and U1 corresponding to 100% migration). The average of two individual biological repeats is plotted. *P*-values were calculated by Student's two-tailed *t*-test comparing m7G or TMG U1+20 migration to the uncapped U1+20 migration at each timepoint (*: *P* < 0.05; n.s.: *P* > 0.1).

(SI Appendix, Fig. S7A). In addition to recognizing the TMG cap, TOE1 may also directly recognize the Sm complex. Consistent with this additional complexity in snRNP recognition, depletion of the Sm complex component SmB differentially affected the processing of different snRNAs by TOE1 (Fig. 3), but we observed no correlation with the extent to which cap trimethylation was affected (SI Appendix, Fig. S7B). The early-stage processing of snRNAs by TOE1, which occurs prior to Sm complex assembly and cap trimethylation, is likely a consequence of the activity of TOE1 as a deadenylase (3–5, 13). Consistent with this idea, TOE1 can process a U1 snRNA containing an oligo(A)-tail to maturity in vitro, independent of its cap structure (SI Appendix, Fig. S7 C–E).

Hundreds of unstable snRNA variants are encoded in the human genome (60, 61). We previously observed that TOE1 distinguishes canonical snRNAs from tested unstable snRNA variants, leaving the latter unprocessed (5). Our observations in this study demonstrate that U1v15 RNA escapes TOE1 recognition specifically due to nucleotide variations in the Sm binding motif (Figure 5). These same nucleotide variations also trigger 3' end oligo(U) and -(A) tailing (Fig. 5), features known to promote degradation by exonucleases and the decapping machinery

(10–12, 50). Thus, multiple layers of quality control act on U1v15 RNA to prevent it from assembling into spliceosomes. This likely represents general mechanisms that serve to repress the subset of transcribed snRNA variants that have acquired nucleotide variations that impair Sm complex assembly, either in the Sm binding motif or in other regions of the snRNA that affect Sm binding, such as the U1-70 K binding site (Figs. 2 and 4).

A large number of small non-coding RNAs are processed by enzymes of the DEDD deadenylase family, including TOE1 and PARN (3, 4). Despite their homology and shared preference for oligo(A) tails, TOE1 and PARN show different specificities for small RNA substrates (3–5). How do these two enzymes achieve specificity toward different substrates? Our findings suggest that this may be explained for a subset of target RNAs by their cap structures. While we find here that TOE1 recognizes the TMG cap characteristic of Pol II snRNAs, PARN was previously shown to be activated by the m7G cap (55–57), a 5' modification found on a subset of small ncRNAs targeted by PARN, including TERC and a subset of snoRNAs. There are reports that TERC can also be observed with a TMG cap (62) and serve as a substrate for TOE1 (19), although we have been unable to detect processing of TERC by TOE1 in HEK293T cells (3). While cap structures

may explain the differential activity of TOE1 and PARN toward a subset of substrates, how PARN achieves specificity for other substrates, including a majority of snoRNAs that are processed from introns and therefore remain uncapped (63), remains unresolved. A contributing factor may well be subcellular localization, with TOE1 known to co-localize with snoRNAs in Cajal Bodies and PARN with snoRNAs and other substrates in the nucleolus. Similarly, the molecular basis for target preferences of other maturation exonucleases PNLDC1 and USB1 remains an important question for future study.

The 3' ends of newly transcribed small non-coding RNAs serve as quality control hubs for competing exonucleases to drive 3' end maturation or degradation. In the case of snoRNAs, previous studies have shown that USB1 processes the Pol-III transcribed U6 and U6atac snoRNAs leading to their stabilization (23, 64). For Pol II snoRNAs, we previously identified TOE1 and the nuclear exosome as competing exonucleases acting on their 3' ends to distinguish canonical snoRNAs from unstable snoRNA variants (5). Our findings here, that TOE1 is activated by late snoRNA biogenesis features, Sm complex assembly, and cap trimethylation, explain how TOE1 can achieve specificity toward canonical snoRNPs undergoing correct assembly. Unstable snoRNA variants (5, 61) and canonical snoRNAs experiencing defects in snoRNP assembly (12, 50), likely undergo degradation because they fail to reach this late stage of snoRNP biogenesis. This may represent a general principle in RNA quality control whereby substrate specificity relies on maturation enzymes that specifically recognize canonical RNAs undergoing proper biogenesis in competition with more promiscuous degradation enzymes to distinguish normal from aberrant RNAs and ultimately dictate their fate.

Materials and Methods

Cell Culture, RNA Interference, and Plasmid Transfections. All cells were maintained in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% Fetal Bovine Serum (FBS, Gibco) and 1% penicillin/streptomycin (Gibco) at 37 °C, 5% CO₂. Mycoplasma testing was routinely performed. TOE1 was depleted in HEK 293 T-REX-derived TOE1-degron cells (5) by incubation with 600 μM auxin hormone Indole-3-Acetic Acid (IAA, Sigma) for 8 h. RNA interference was performed in HEK 293 T Flp-In cells (FIRT, Thermo Fisher) with 20 nM small interfering (si)RNA targeting genes of interest, or luciferase as a control (*SI Appendix, Table S1*), using siLentFect (Bio-Rad) transfection reagent according to the manufacturer's recommendations at 72 and 24 h before cell harvest. Plasmid transfections were performed using 2 μg plasmid per 3.5-cm well plates using Lipofectamine 2000 (Life Technologies) transfection reagent according to the manufacturer's recommendations at 48 h before harvest, unless specified otherwise.

Global 3' End Sequencing of Newly Transcribed Small RNAs. HEK 293 T-REX TOE1-degron cells were treated to deplete TOE1 as described above. A total of 1.2 × 10⁶ TOE1-depleted, or control non-depleted, cells were incubated with 0.2 mM Ethynyl Uridine (EU; Thermo Fisher) for 8 h and harvested in 1 mL TRIzol reagent (Thermo Fisher). Total RNA was isolated according to the manufacturer's recommendation. Small RNA from 35 μg of total RNA from each sample was isolated by separation in a 9% polyacrylamide/6 M urea denaturing gel. After Sybr Gold staining (Thermo Fisher), small RNAs between 100 and 500 nucleotides in length were cut out of the gel and eluted into 400 μL gel elution buffer (0.3 M sodium acetate pH 5.3, 1 mM EDTA, 0.1% SDS) by end-over-end rotation overnight at 4 °C. Eluted small RNAs were subsequently subjected to phenol:chloroform:isoamyl alcohol extraction followed by ethanol precipitation as previously described (3). Genomic DNA was removed using the Turbo DNA-free kit (Thermo Fisher) and ribosomal (r)RNAs were depleted using the RiboCOP rRNA depletion kit (Lexogen) per the manufacturer's recommendations. The rRNA-depleted small RNA samples were subsequently subjected to FastAP/PNK treatment to remove RNA 5' and 3' phosphates by incubating samples with 2.5 μL 10× FastAP Buffer (Thermo Fisher), 0.5 μL RNaseOUT (Thermo Fisher), and 2.5 μL FastAP phosphatase (Thermo Fisher)

in 25 μL total volume at 37 °C for 30 min, followed by supplementation with 10 μL 10× PNK buffer (NEB), 1 μL 0.1 M DTT, 1 μL Turbo DNase (2 U/μL; Thermo Fisher), 1 μL RNaseOUT (40 U/μL; Thermo Fisher), and 7 μL Polynucleotide Kinase (10 U/μL; NEB) in a total of 100 μL and incubation for 20 min at 37 °C. RNA samples were purified using an RNA Clean & Concentrator kit (Zymo Research). AG10N or AG11N RNA adaptors (10 μM; *SI Appendix, Table S1*) were ligated to the RNA 3' ends of each sample by incubation at 25 °C for 75 min in a 20 μL total volume containing 1.8 μL DMSO (Sigma), 2 μL 10× T4 ligase buffer (NEB), 0.2 μL 0.1 M ATP (Thermo Fisher), 0.2 μL RNaseOUT (40 U/μL; Thermo Fisher), 8 μL 50% PEG 8000 (NEB), 1.3 μL T4 RNA ligase (30 U/μL; NEB). RNA samples were then subjected to extraction of newly transcribed RNAs using the Click-it nascent RNA capture kit (Thermo Fisher) as previously described (5). Reverse transcription was performed on the RNA capture beads in a total of 20 μL with 0.5 μL 20 μM AR17 primer (*SI Appendix, Table S1*), 2 μL 10× AffinityScript buffer (Agilent), 2 μL 0.1 M DTT, 0.8 μL 100 mM dNTPs, 0.3 μL RNaseOUT (40 U/μL; Thermo Fisher), 0.9 μL AffinityScript Reverse Transcriptase (Agilent) at 55 °C for 45 min, followed by 15 min incubation at 70 °C and 5 min of incubation at 85 °C to release cDNA. Excess primers and RNA was removed from the first-strand cDNA by incubating with 3.5 μL ExoSAP-IT (Thermo Fisher) at 37 °C for 15 min followed by addition of 3 μL of 1 M NaOH at 70 °C for 12 min. Then, 3 μL of 1 M HCl was added to the sample after the clean-up to adjust pH. A 3Tr3 adaptor (*SI Appendix, Table S1*) was ligated to cDNA 5' ends by adding 0.8 μL of the adaptor at 80 μM to the cDNA sample along with 1 μL 100% DMSO (Sigma), 2 μL 10× T4 ligase buffer (NEB), 0.2 μL 0.1 M ATP (Thermo Fisher), 1.5 μL T4 RNA ligase (30 U/μL; NEB), and 1.1 μL double distilled water, and incubating the reaction at 25 °C for 16 h. The cDNA was amplified in two stages of PCR using Q5 DNA polymerase (NEB). For the first PCR reaction, the cDNA library was amplified using 3' adaptor primer (AR17) and a primer complementary to the 5' adaptor (RC_3Tr3) (*SI Appendix, Table S1*) for 8 cycles. The PCR product was purified by AMPure XP beads (Beckman Coulter) per the manufacturer's recommendation. The second PCR reaction was performed using Illumina Truseq D50X and D70X primers (*SI Appendix, Table S1*) for 18 cycles. Amplified cDNA was subsequently isolated by separation in a 3% agarose gel and isolation of cDNAs 200 to 600 base pairs in length (corresponding to 100 to 500 nucleotide RNAs plus adaptors) using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's recommendation. The library quality was monitored by qPCR for select genes and TapeStation (Agilent) analyses. In addition, 100-bp paired-end sequencing was performed on an Illumina NovaSeq 6000 platform per the manufacturer's recommendation.

Sequencing Data Analyses. FASTQ files were first subjected to 3' adaptor and PCR duplicate removal using custom python scripts (<https://pypi.org/project/jla-demultiplexer/>). For highly abundant small RNAs (U1, U2, U3, 7SL, and 7SK), 100,000 reads were selected for subsequent analyses by averaging three random samplings using seqtk (<https://github.com/lh3/seqtk>). This was done to prevent undercounting of these RNAs since their abundance in the full library exceeded the depth of the randomizer used for detecting PCR duplicates (*SI Appendix, Fig. S1b*). For all other RNAs, the entire library was used. Reads were mapped to the human genome (version hg38) using STAR 2.7.8a (65). To diminish mis-mapping of canonical small RNA reads to small RNA variant genes, a three-pass alignment was applied. Briefly, reads were first mapped to the human hg38 genome (STAR --outFilterMultimapNmax 1000 --alignIntronMin 9999999 --outFilterMultimapScoreRange --outFilterMismatchNoverLmax 0.2) and reads mapping to small RNA genes were extracted using bedtools (66) and samtools (67). The mapped small RNA reads were then aligned to a custom FASTA database of canonical small RNA genes, each including 50 base pair upstream and downstream sequences (STAR --outFilterMultimapNmax 1000 --outFilterMultimapScoreRange 0 --outFilterMismatchNoverLmax 0.2 --outFilterMismatchNoverReadLmax 0.05 --clip5pNbases 20 0 --clip3pNbases 0 20 --alignIntronMin 9999999 --alignMatesGapMax 500 --alignEndsType EndToEnd --outReadsUnmapped Fastx). Reads that mapped to the canonical small RNA gene database were subsequently re-aligned to the full human hg38 genome (STAR --outFilterMultimapNmax 1000 --outFilterMultimapScoreRange 0 --outFilterMismatchNoverLmax 0.025 --alignIntronMin 9999999 --alignMatesGapMax 500 --alignEndsType Local) and those that again mapped to canonical small RNA genes were extracted with bedtools. This step was performed to remove canonical small RNA reads with sequencing errors that may misalign with small RNA variant genes. Reads that failed to map to the canonical small RNA gene database were also re-aligned with the full human

hg38 genome using the same settings and combined with the reads mapping to canonical small RNA genes. Gene-specific 3' end information and graphs were subsequently generated using Tailor (68) (<https://github.com/TimNicholsonShaw/tailor>) using the global alignment mode.

RNA Gene-Specific 3' End Sequencing. RNA was isolated using TRIzol (Thermo Fisher) per the manufacturer's recommendation. RNA was subsequently treated with DNase I (1 U/ μ L; Zymo research) in DNase buffer (10 mM Tris-HCl pH 7.5, 2.5 mM MgCl₂, 0.5 mM CaCl₂) at 25 °C for 15 min, followed by extraction with phenol:chloroform:isoamyl alcohol and ethanol precipitation as previously described (5). For sequencing of newly transcribed RNAs, nascent RNA was captured as previously described (5). Gene-specific RNA 3' end sequencing libraries were generated using gene-specific primers (SI Appendix, Table S1) and sequenced on an Illumina MiSeq platform as previously described (5).

Isolation of Flag-Tagged TOE1. HEK 293 T Flp-In cells (FIRT) cells were transfected with pcDNA5-Flag-TOE1 plasmid (13) as described above. First, 1 μ g/mL of tetracycline was added 24 h before harvest to induce Flag-TOE1 expression. Then, 2 \times 10⁷ Flag-TOE1-induced cells were lysed in 2.5 mL of isotonic lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 0.1% Triton-X100, 1 mM PMSF, 2 μ g/mL Aprotinin, 2 μ g/mL Leupeptin) with 125 μ g/mL RNase A. The lysate concentration was measured using BCA protein assay (Thermo Fisher) following the manufacturer's protocol. Then, 40 μ L of 50% anti-FLAG M2 agarose slurry (Sigma) was washed twice with 500 μ L NET-2 (10 mM Tris HCl pH 7.4, 150 mM NaCl, 0.1% Triton-X100) before incubation with 1 to 5 mg of cell lysate with end-to-end rotation for 2 h at 4 °C. Beads were subsequently washed with 500 μ L NET-2 eight times. Flag-TOE1 was eluted with 100 μ L of NET-2 containing 150 ng/ μ L FLAG peptide (ApexBio) and 10% glycerol by rotating at 4 °C for 30 min. The isolated protein was detected by silver staining (Thermo Fisher) per the manufacturer's recommendation. Eluates were aliquoted and stored at -80 °C until use.

In Vitro snRNP Pull-Down and 3' End Processing Assay. HEK293T-Rex TOE1-degron cells were treated to deplete TOE1 as described above and co-transfected with 0.2 μ g U1-WT and 2 μ g Sm-mutant (SI Appendix, Table S2) expression plasmids as described above, 16 h prior to cell harvest. A total of 2 \times 10⁷ cells were suspended in 0.2 mL of low salt lysis buffer (10 mM Tris-HCl pH 7.5, 60 mM NaCl, 2 mM EDTA, 0.1% Triton-X100, 1 mM PMSF, 2 μ g/mL Aprotinin, 2 μ g/mL Leupeptin, 0.1 μ g/mL yeast total RNA, and 2 U/mL RNaseOUT) and incubated on ice for 10 min. Subsequently, 20 μ L of 0.5 μ M 2'OMe-RNA-oligo probe (SI Appendix, Table S1), 12 μ L streptavidin magnetic beads (NEB), and 150 μ L wash/binding buffer (0.5 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA) were incubated on ice for 30 min with occasional agitation. Then, 200 to 400 μ L of cell lysate was added to each tube and incubated at 4 °C for 2 h. Beads were washed twice with 100 μ L wash/binding buffer followed by one wash with 100 μ L low salt wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA), and five times with 100 μ L EDTA-free low salt buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40). After washes, 200 μ L of reaction buffer (20 mM HEPES pH 7.4, 2 mM MgCl₂, 0.1 mg/mL bovine serum albumin, 1 mM spermidine, 0.1% NP-40) with 0.5 U/ μ L RNaseOUT (Invitrogen), 0.5 μ g/ μ L yeast total RNA and 0 to 100 nM of Flag-TOE1 was added to the isolated snRNPs. Reaction tubes were incubated with gentle rotation at 30 °C for 30 min followed by incubation at 4 °C for 30 min. The beads were then isolated using a magnetic stand and resuspended in 100 μ L reaction buffer with 100 μ L formamide/2 mM EDTA and incubated at 80 °C for 5 min. Then, 1 mL of TRIzol (Thermo Fisher) was added to each sample. Two biological repeats were obtained using individual snRNP pull-down and two individual Flag-TOE1 preparations. RNA extraction, 3' end sequencing library preparation, sequencing, and data analysis were performed as described above.

In Vitro Transcription and 3' End Processing Assay. TMG-, m7G-capped, or uncapped U1 snRNAs were produced by in vitro transcription using T7 RNA polymerase (NEB). Briefly, 500 ng of DNA template, produced by PCR from a U1 snRNA expression plasmid (50) using T7_U1-Forward primer with T7_U1-Reverse_20geno or T7_U1-Reverse-10A primers (SI Appendix, Table S1), was mixed with 8 mM m7G cap analog (ARCA; NEB), TMG cap analog (Jena Bioscience) or no cap analog, 0.05 M DTT, 1 μ L RNaseOUT (2 U/ μ L; Thermo Fisher), 2 μ L NTP buffer mix (1.34 mM final concentration for each NTP, NEB), 2 μ L T7 RNA Polymerase Mix (NEB) in a total of 20 μ L. In vitro transcription was performed at 37 °C for 16 h. RNA products were purified using an RNA Clean & Concentrator kit (Zymo Research) and quantified using a Nanodrop (Thermo Fisher) per the manufacturer's recommendations. For TOE1 3' end processing assays, 10 ng of RNA was mixed with 100 nM Flag-tagged or 50 nM His-tagged TOE1 in 10 μ L reaction buffer containing 20 mM HEPES pH 7.4, 2 mM MgCl₂, 100 μ g/mL BSA, 1 mM Spermidine, 0.1% NP-40, 0.5 U/ μ L RNaseOUT, 0.5 μ g/ μ L yeast total RNA. The processing reaction was performed at 37 °C for 0 to 120 min. Reactions were terminated with 10 μ L of denaturing load buffer (95% formamide, 10 mM EDTA, 0.01% Bromophenol Blue, 0.01% Xylene Cyanol) followed by incubation at 80 °C for 10 min to denature the RNA. The reaction products were subsequently separated in a 7% acrylamide/6 M urea denaturing gel. RNA was stained with Sybr Gold (Thermo Fisher) by nutation in the dark for 30 min. RNA was imaged by a Typhoon gel imager (Amersham). Gel images were analyzed using ImageJ. The processing reactions were repeated twice using the same recombinant TOE1 preparations.

qPCR assays. AR17 (SI Appendix, Table S1)-primed cDNA was amplified using Fast SYBR Green master mix (Thermo Fisher) with primers for RNAs of interest (SI Appendix, Table S1) on a QuantStudio Real-Time PCR system (Thermo Fisher). Relative levels were quantified using the $\Delta\Delta C_t$ method (69).

Western Blotting. Western blots were performed by separating proteins in SDS-polyacrylamide gels followed by transfer to nitrocellulose membranes using standard procedures. Membranes were incubated overnight at 4 °C with rabbit polyclonal anti-CAF1Z/TOE1 (13) at 1:1000, rabbit polyclonal anti-UFP1 (70) at 1:1000, mouse monoclonal anti-SNRPB (Thermo Fisher) at 1:500, each in PBS with 0.1% Tween (PBST) and 5% nonfat milk, or with mouse monoclonal anti-U1-70 K (Synaptic System) at 1:1000 in PBST with 5% BSA. Secondary antibodies were goat anti-rabbit IRDye 680RD (LI-COR) at 1:15000, HPR donkey anti-rabbit IgG (H + L) (Thermo Fisher), or HPR goat anti-mouse rabbit IgG (H + L) (Thermo Fisher) at 1:20000 in PBST with 5% nonfat milk. Western blots were visualized using an Odyssey Fc imaging system (LI-COR).

Data, Materials, and Software Availability. RNA sequencing data have been deposited to the Gene Expression Omnibus under Accession No. GSE240774 (51).

ACKNOWLEDGMENTS. We thank Haoxin Li in the laboratory of Benjamin Cravatt (The Scripps Research Institute, La Jolla, CA) for providing His-TOE1. We thank Tim Nicholson-Shaw for sequencing analysis support. We thank Alberto Carreño, Tim Nicholson-Shaw, Cody Ocheltree, Megan Dowdle, and Stephen Sanders for useful discussion of the manuscript. We acknowledge the UC San Diego Stem Cell Genomics and Microscopy Core for use of their Illumina MiSeq instrument. This work was supported by NIH grant R35 GM118069 to J.L.-A. R.M.L. was the recipient of a National Research Service Award Postdoctoral Fellowship (NIH F32 GM106706) and was a San Diego Institutional Research and Academic Career Development Award Fellow (NIH K12 GM06852). This publication includes data generated at the UC San Diego Institute for Genomic Medicine Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a NIH Shared Instrumentation Grant (#S10 OD026929).

Author affiliations: ^aDepartment of Molecular Biology, School of Biological Sciences, University of California, San Diego, La Jolla, CA 92093

1. G. Chanfreau, P. Legrain, A. Jacquier, Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.* **284**, 975-988 (1998).
2. E. Pefalski, T. Dandekar, Y. Henry, D. Tollervy, Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell. Biol.* **18**, 1181-1189 (1998).
3. R. M. Lardelli *et al.*, Biallelic mutations in the 3' exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. *Nat. Genet.* **49**, 457-464 (2017).

4. A. Son, J.-E. Park, V. N. Kim, PARN and TOE1 constitute a 3' end maturation module for nuclear non-coding RNAs. *Cell Rep.* **23**, 888-898 (2018).
5. R. M. Lardelli, J. Lykke-Andersen, Competition between maturation and degradation drives human snRNA 3' end quality control. *Genes Dev.* **34**, 989-1001 (2020).
6. A. Fatica, Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J.* **19**, 6218-6229 (2000).

7. C. Allmang, Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399–5410 (1999).
8. A. Van Hoof, P. Lennertz, R. Parker, Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol. Cell. Biol.* **20**, 441–452 (2000).
9. T. Kawamoto, R. Yoshimoto, I. Taniguchi, M. Kitabatake, M. Ohno, ISG20 and nuclear exosome promote destabilization of nascent transcripts for spliceosomal U snRNAs and U1 variants. *Genes Cells* **26**, 18–30 (2021).
10. D. Ustianenko *et al.*, TUT-DIS3L2 is a mammalian surveillance pathway for aberrant structured non-coding RNAs. *EMBO J.* **35**, 2179–2191 (2016).
11. A. Labno *et al.*, Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs. *Nucleic Acids Res.* **44**, 10437–10453 (2016).
12. A. Roithová, Z. Feketová, Š. Vaňáčková, D. Staněk, DIS3L2 and LSM proteins are involved in the surveillance of Sm ring-deficient snRNAs. *Nucleic Acids Res.* **48**, 6184–6197 (2020).
13. E. Wagner, S. L. Clement, J. Lykke-Andersen, An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. *Mol. Cell. Biol.* **27**, 1686–1695 (2007).
14. C. G. Körner, E. Wahle, Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. *J. Biol. Chem.* **272**, 10448–10456 (1997).
15. H. Berndt *et al.*, Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* **18**, 958–972 (2012).
16. D. H. Moon *et al.*, Poly(A)-specific ribonuclease (PARN) mediates 3'-end maturation of the telomerase RNA component. *Nat. Genet.* **47**, 1482–1488 (2015).
17. S. Shukla, J. C. Schmidt, K. C. Goldfarb, T. R. Cech, R. Parker, Inhibition of telomerase RNA decay rescues telomerase deficiency caused by dyskerin or PARN defects. *Nat. Struct. Mol. Biol.* **23**, 286–292 (2016).
18. C. M. Roake *et al.*, Disruption of telomerase RNA maturation kinetics precipitates disease. *Mol. Cell* **74**, 688–700.e3 (2019).
19. T. Deng *et al.*, TOE1 acts as a 3' exonuclease for telomerase RNA and regulates telomere maintenance. *Nucleic Acids Res.* **47**, 391–405 (2019).
20. N. Izumi *et al.*, Identification and functional analysis of the pre-piRNA 3' trimmer in silkworms. *Cell* **164**, 962–973 (2016).
21. D. Anastakis *et al.*, Mammalian PNLD1 is a novel poly(A) specific exonuclease with discrete expression during early development. *Nucleic Acids Res.* **44**, 8908–8920 (2016).
22. D. Ding *et al.*, PNLD1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. *Nat. Commun.* **8**, 819 (2017).
23. S. Mroczek *et al.*, C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3' end modification. *Genes Dev.* **26**, 1911–1925 (2012).
24. V. Shchepachev, H. Wischniewski, E. Missiaglia, C. Soneson, C. M. Azzalin, Mpn1, mutated in poikiloderma with neutropenia protein 1, is a conserved 3'-to-5' RNA exonuclease processing U6 small nuclear RNA. *Cell Rep.* **2**, 855–865 (2012).
25. H.-C. Jeong *et al.*, USB1 is a miRNA deadenylase that regulates hematopoietic development. *Science* **379**, 901–907 (2023).
26. H. Tummala *et al.*, Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. *J. Clin. Invest.* **125**, 2151–2160 (2015).
27. S. Dhanraj *et al.*, Bone marrow failure and developmental delay caused by mutations in poly(A)-specific ribonuclease (PARN). *J. Med. Genet.* **52**, 738–748 (2015).
28. L. Nagiraj *et al.*, Variant PNLD1, defective piRNA processing, and azoospermia. *N. Engl. J. Med.* **385**, 707–719 (2021).
29. A. J. Walne, T. Vulliamy, R. Beswick, M. Kirwan, I. Dokal, Mutations in C16orf57 and normal-length telomeres unify a subset of patients with dyskeratosis congenita, poikiloderma with neutropenia and Rothmund-Thomson syndrome. *Hum. Mol. Genet.* **19**, 4453–4461 (2010).
30. C. L. Will, R. Luhrmann, Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
31. D. Baillat *et al.*, Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
32. M. Salditt-Georgieff, M. Harpold, S. Chen-Kiang, J. E. Darnell, The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* **19**, 69–78 (1980).
33. M. Ohno, A. Segref, A. Bachi, M. Wilm, I. W. Mattaj, PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187–198 (2000).
34. B. R. So *et al.*, A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange. *Nat. Struct. Mol. Biol.* **23**, 225–230 (2016).
35. U. Fischer, Q. Liu, G. Dreyfuss, The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* **90**, 1023–1029 (1997).
36. Q. Liu, U. Fischer, F. Wang, G. Dreyfuss, The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* **90**, 1013–1021 (1997).
37. D. Buhler, V. Raker, R. Luhrmann, U. Fischer, Essential role for the Tudor domain of SMN in spliceosomal U snRNP assembly: Implications for spinal muscular atrophy. *Hum. Mol. Genet.* **8**, 2351–2357 (1999).
38. W. J. Friesen *et al.*, The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol. Cell. Biol.* **21**, 8289–8300 (2001).
39. G. Meister, D. Bühler, R. Pillai, F. Lottspeich, U. Fischer, A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell Biol.* **3**, 945–949 (2001).
40. L. Pellizzoni, J. Yong, G. Dreyfuss, Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775–1779 (2002).
41. S. Massenet, L. Pellizzoni, S. Pauskhin, I. W. Mattaj, G. Dreyfuss, The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol. Cell. Biol.* **22**, 6533–6541 (2002).
42. J. Yong, M. Kasim, J. L. Bachorik, L. Wan, G. Dreyfuss, Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol. Cell* **38**, 551–562 (2010).
43. I. W. Mattaj, Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**, 905–911 (1986).
44. J. Mouaikel, C. Verheggen, E. Bertrand, J. Tazi, R. Bordonné, Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell* **9**, 891–901 (2002).
45. J. Mouaikel *et al.*, Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep.* **4**, 616–622 (2003).
46. J. Huber *et al.*, Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.* **17**, 4114–4126 (1998).
47. I. Palacios, Nuclear import of U snRNPs requires importin beta. *EMBO J.* **16**, 6783–6792 (1997).
48. M. T. Bohnsack, K. E. Sloan, Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol. Chem.* **399**, 1265–1276 (2018).
49. H. Yang, M. L. Moss, E. Lund, J. E. Dahlberg, Nuclear processing of the 3'-terminal nucleotides of pre-U1 RNA in *Xenopus laevis* oocytes. *Mol. Cell. Biol.* **12**, 1553–1560 (1992).
50. S. Shukla, R. Parker, Quality control of assembly-defective U1 snRNAs by decapping and 5'-to-3' exonucleolytic digestion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3277–E3286 (2014).
51. T. Ma, snRNA 3' end sequencing, NCBI Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE240774>. Deposited 14 August 2023.
52. D. E. Tsai, D. S. Harper, J. D. Keene, U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res.* **19**, 4931–4936 (1991).
53. H. Hernández *et al.*, Isoforms of U1-70k control subunit dynamics in the human spliceosomal U1 snRNP. *PLoS ONE* **4**, e7202 (2009).
54. Y. Kondo, C. Oubridge, A.-M.M. van Roon, K. Nagai, Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *eLife* **4**, e04986 (2015).
55. J. Martínez *et al.*, A 54-kDa fragment of the poly(A)-specific ribonuclease is an oligomeric, processive, and cap-interacting poly(A)-specific 3' exonuclease. *J. Biol. Chem.* **275**, 24222–24230 (2000).
56. M. Gao, D. T. Fritz, L. P. Ford, J. Wilusz, Interaction between a poly(A)-specific ribonuclease and the 5' cap influences mRNA deadenylation rates in vitro. *Mol. Cell* **5**, 479–488 (2000).
57. E. Dehlin, M. Wormington, C. G. Körner, E. Wahle, Cap-dependent deadenylation of mRNA. *EMBO J.* **19**, 1079–1086 (2000).
58. L. Cheng *et al.*, Loss of the RNA trimethylguanosine cap is compatible with nuclear accumulation of spliceosomal snRNAs but not pre-mRNA splicing or snRNA processing during animal development. *PLoS Genet.* **16**, e1009098 (2020).
59. L. Chen *et al.*, TGS1 impacts snRNA 3'-end processing, ameliorates survival motor neuron-dependent neurological phenotypes in vivo and prevents neurodegeneration. *Nucleic Acids Res.* **50**, 12400–12424 (2022).
60. O. O'Reilly *et al.*, Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.* **23**, 281–291 (2013).
61. J. W. Mabin, P. W. Lewis, D. A. Brow, H. Dvinge, Human spliceosomal snRNA sequence variants generate variant spliceosomes. *RNA* **27**, 1186–1203 (2021).
62. L. Chen *et al.*, Loss of human TGS1 hypermethylase promotes increased telomerase RNA and telomere elongation. *Cell Rep.* **30**, 1358–1372.e5 (2020).
63. J. Kufel, P. Grzechnik, Small nucleolar RNAs tell a different tale. *Trends Genet.* **35**, 104–117 (2019).
64. V. Shchepachev, H. Wischniewski, C. Soneson, A. W. Arnold, C. M. Azzalin, Human Mpn1 promotes post-transcriptional processing and stability of U6atac. *FEBS Lett.* **589**, 2417–2423 (2015).
65. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
67. H. Li *et al.*, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
68. T. Nicholson-Shaw, J. Lykke-Andersen, Tailer: A pipeline for sequencing-based analysis of non-polyadenylated RNA 3' end processing. *RNA* **28**, 645–656 (2022).
69. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods* **25**, 402–408 (2001).
70. J. Lykke-Andersen, M.-D. Shu, J. A. Steitz, Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**, 1121–1131 (2000).