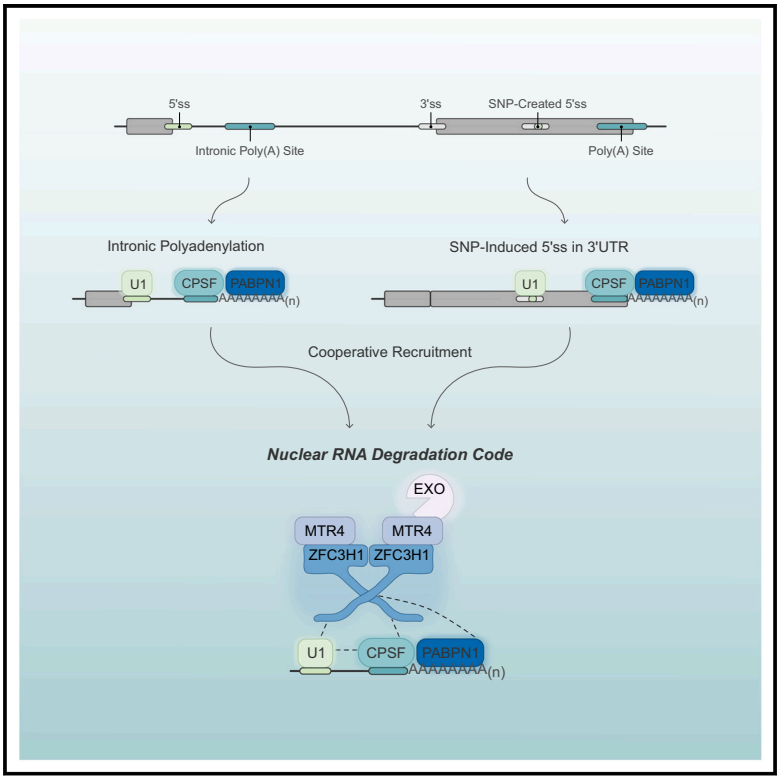
Molecular Cell

Article

A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance

Graphical abstract



Highlights

- Intronically polyadenylated RNAs are major targets of exosome-mediated degradation
- The combination of a 5′ splice site and a poly(A) junction targets RNA for decay
- U1 snRNP and CPA factors cooperatively recruit PAXT and the exosome to RNAs
- The nuclear RNA degradation code contributes to human diseases

Authors

Lindsey V. Soles, Liang Liu, Xudong Zou, ..., Georg Seelig, Lei Li, Yongsheng Shi

Correspondence

yongshes@uci.edu

In brief

Soles et al. demonstrate that the combination of a 5′ splice site and a poly(A) junction constitutes a nuclear RNA degradation code that targets RNAs for degradation by the PAXT adaptor and the RNA exosome. This study provides fundamental insight into how the RNA exosome specifically recognizes misprocessed RNAs.

Soles et al., 2025, Molecular Cell *85*, 1–14
April 17, 2025 © 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
https://doi.org/10.1016/j.molcel.2025.03.010

CellPress

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

Molecular Cell

CellPress

Article

A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance

Lindsey V. Soles,[1] Liang Liu,[1] Xudong Zou,[2] Yoseop Yoon,[1] Shuangyu Li,[1] Lusong Tian,[1] Marielle Valdez,[1] Angela M Yu,[3] Hong Yin,[4] Wei Li,[5,6] Fangyuan Ding,[4,6] Georg Seelig,[7] Lei Li,[2] and Yongsheng Shi[1,6,8,*]

[1]Department of Microbiology and Molecular Genetics, School of Medicine, University of California, Irvine, Irvine, CA 92617, USA
[2]Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518107, China
[3]Therapeutic Innovation Center (THINC), and Verna and Marrs McLean Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, TX, USA
[4]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92617, USA
[5]Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92617, USA
[6]The Center for RNA Science and Therapeutics, University of California, Irvine, Irvine, CA 92697, USA
[7]Department of Electrical and Computer Engineering, University of Washington, Seattle, Seattle, WA 98195, USA
[8]Lead contact
*Correspondence: yongshes@uci.edu
https://doi.org/10.1016/j.molcel.2025.03.010

SUMMARY

The RNA exosome plays critical roles in eukaryotic RNA degradation, but how it specifically recognizes its targets remains unclear. The poly(A) tail exosome targeting (PAXT) connection is a nuclear adaptor that recruits the exosome to polyadenylated RNAs, especially transcripts polyadenylated at intronic poly(A) sites. Here, we show that PAXT-mediated RNA degradation is induced by the combination of a 5′ splice site (ss) and a poly(A) junction (PAJ) but not by either sequence alone. These sequences are bound by U1 small nuclear ribonucleoprotein particle (snRNP) and cleavage/polyadenylation factors, which, in turn, cooperatively recruit PAXT. As the 5′ ss-PAJ combination is typically absent on correctly processed RNAs, it functions as a "nuclear RNA degradation code" (NRDC). Importantly, disease-associated single nucleotide polymorphisms that create novel 5′ ss in 3′ untranslated regions can induce aberrant mRNA degradation via the NRDC mechanism. Together, our study identified the first NRDC, revealed its recognition mechanism, and characterized its role in human diseases.

INTRODUCTION

A large portion of eukaryotic transcripts are degraded in the nucleus, including promoter upstream transcripts (PROMPTs) and misprocessed mRNAs.[1] The RNA exosome is an essential RNA degradation machine that employs 3′–5′ exoribonucleolytic and endoribonucleolytic enzymes to process and degrade RNAs.[2] The RNA exosome lacks specificity and requires adaptor complexes to recognize its target RNAs.[1,3,4] Among the exosome adaptors in the nucleoplasm of mammalian cells, the nuclear exosome targeting (NEXT) complex primarily targets non-polyadenylated RNAs,[5] and the poly(A) tail exosome targeting (PAXT) connection generally targets polyadenylated RNAs.[6–8] The core of the PAXT connection is a stable dimer formed by the helicase MTR4 and the zinc-finger protein ZFC3H1.[7,8] The nuclear poly(A) binding protein PABPN1 associates with the PAXT core in a partially RNA-dependent manner and has been suggested to direct PAXT to polyadenylated RNAs.[7] However, as most mRNAs and many lncRNAs have poly(A) tails, additional RNA features are necessary to specifically recruit PAXT and the RNA exosome for degradation.

The 3′ ends of nearly all eukaryotic mRNAs are formed by cleavage and polyadenylation (CPA).[9,10] Mammalian poly(A) sites (PASs) typically consist of an A(A/U)UAAA hexamer, called the poly(A) signal, a downstream U/GU-rich element, and other auxiliary sequences. These sequences are recognized by several CPA factors, including CPSF and CstF, to assemble the CPA complex in which the two chemical reactions take place.[11,12] The majority of human genes produce multiple transcript isoforms by using alternative PASs throug a mechanism called alternative polyadenylation.[13,14] Approximately 30%–40% of annotated PASs are within introns, and transcripts produced by intronic polyadenylation (IPA) often encode truncated and non-functional proteins.[15–19] Elevated levels of IPA transcripts are detected in leukemia cells and have been proposed to inactivate tumor suppressor genes.[17] Recently, an IPA transcript of the *TP53* gene was shown to be oncogenic.[18] As such, the accumulation of IPA transcripts is generally deleterious and thus must be repressed. Previous studies suggested that the splicing factor U1 small nuclear ribonucleoprotein particle (snRNP) inhibits CPA at intronic PASs through a splicing-independent mechanism called telescripting.[20,21] More recent

All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

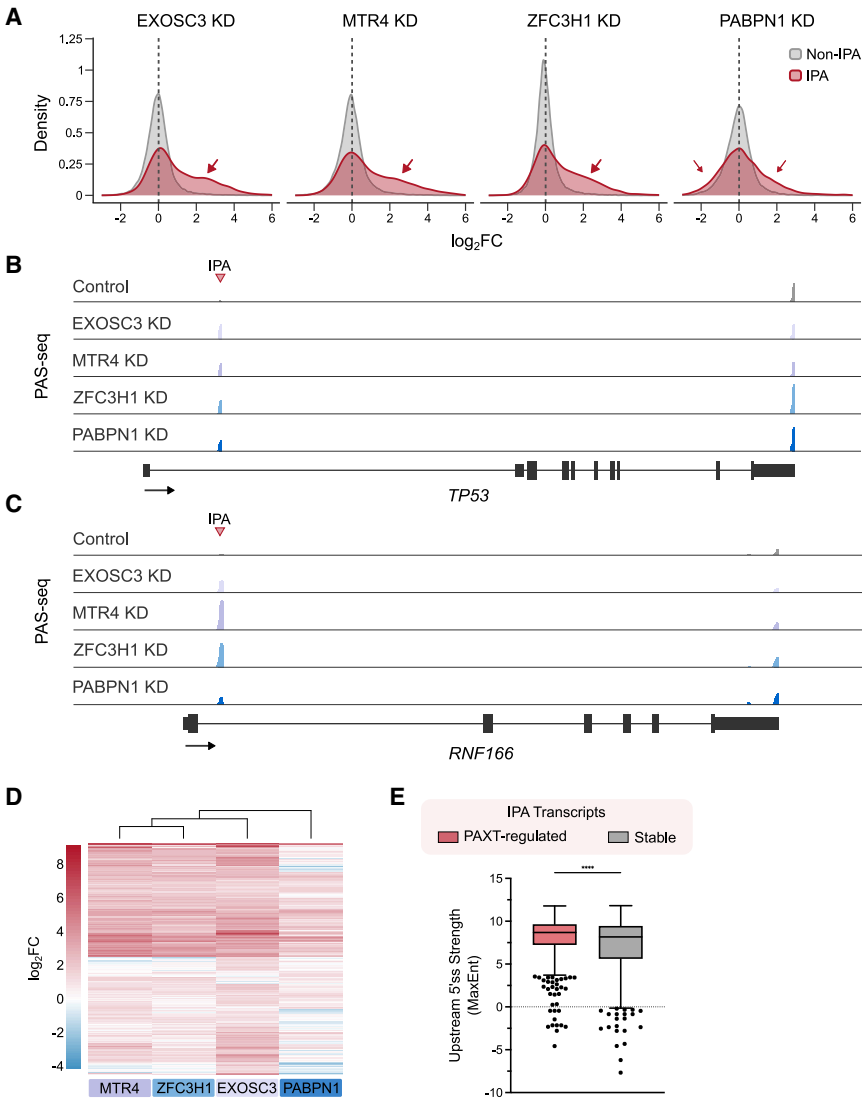


Figure 1. IPA transcripts are major targets of PAXT-mediated degradation

(A) Density plots of the log_2FC for IPA or non-IPA transcripts following the knockdown (KD) of EXOSC3, MTR4, ZFC3H1, or PABPN1. IPA transcripts are depicted in red and non-IPA in gray.

(B and C) PAS-seq data tracks for the genes *TP53* (B) and *RNF166* (C). All tracts are group auto scaled and are on the same scale.

(D) Heatmap of the log_2FC measured for all IPA transcripts that were upregulated in at least one knockdown condition ($log_2FC > 1$, FDR ≤ 0.05).

(E) Boxplot depicting the strength of the nearest upstream 5′ ss, as measured by MaxEnt. PAXT-regulated: significantly upregulated IPA transcripts ($log_2FC > 1$, FDR ≤ 0.05, $N = 507$) following depletion of EXOSC3, MTR4, and ZFC3H1. Stable IPA transcripts: IPA transcripts whose expression did not significantly change following depletion of EXOSC3, MTR4, and ZFC3H1 (counts per million [CPM] in control cells > 1, $log_2FC < 1$ and $log_2FC > -1$, FDR > 0.05, $N = 687$). Statistical analysis was calculated by Mann-Whitney test. ****p value ≤ 0.0001.

studies have provided evidence that at least some IPA transcripts are generated but are targeted for exosome-mediated degradation by PAXT.[8,22] However, similar to other unstable transcripts, it remains unknown which and how IPA transcripts are specifically recognized for degradation.

In this study, we set out to define the mechanism by which PAXT recognizes specific RNA targets. Our results revealed an nuclear RNA degradation code (NRDC) mechanism whereby a specific combination of RNA features target RNAs for degradation by PAXT/exosome and demonstrated the disease relevance of this mechanism.

RESULTS

IPA transcripts are major targets of PAXT- and exosome-mediated degradation

To comprehensively identify the polyadenylated transcripts that are targeted by PAXT and the RNA exosome, we individually depleted the exosome subunit EXOSC3/RRP40, the PAXT core components MTR4 and ZFC3H1, and PABPN1 in HEK293T cells by RNAi (Figure S1A) and performed PAS-seq analyses. PAS-seq is a 3′-end RNA sequencing method that not only maps PASs but also measures the relative levels of polyadenylated RNAs.[23,24] As previous studies have characterized PAXT-targeted PROMPTs in great detail,[6,7] we have focused on the sense-strand polyadenylated transcripts of annotated human genes. Our results revealed that depletion of EXOSC3, MTR4, or ZFC3H1 led to significantly higher levels of 2,371, 2,874, and 2,076 transcripts, respectively (log_2 fold change [FC] > 1, false discovery rate [FDR] < 0.05). We categorized these transcripts into multiple groups, including those polyadenylated in last exons, upstream internal exons, alternative last exons, or introns (Figure S1B). Among the transcripts significantly upregulated in EXOSC3-, MTR4-, and ZFC3H1-depleted cells, IPA transcripts accounted for 43%, 50.7%, and 55.8%, respectively (Figures S1B and S1C), representing the largest group in all samples. In comparison, PABPN1 depletion caused significant accumulation of 1,810 transcripts, which included more transcripts polyadenylated in the last exon (51.4%) than IPA transcripts (31.5%) (Figure S1C), indicating that the effect of PABPN1 depletion was distinct from that of EXOSC3, MTR4, or ZFC3H1 depletion. As IPA transcripts seemed to be the major group regulated by the exosome and ZFC3H1/MTR4, we compared the change in RNA levels between IPA and non-IPA transcripts, including those polyadenylated in last exons, upstream internal exons, or alternative last exons (Figure S1B). As shown in Figure 1A, a much greater proportion of IPA transcripts accumulated to higher levels in EXOSC3-, MTR4-, and ZFC3H1-depleted cells

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

compared with non-IPA transcripts. In contrast to these largely unidirectional changes, depletion of PABPN1 caused significant changes in IPA transcript abundance in both directions (Figure 1A). Two example genes, *TP53* and *RNF166*, are shown to illustrate these trends (Figures 1B and 1C). Hierarchical clustering of all IPA transcripts that were significantly upregulated in at least one condition revealed that ZFC3H1- and MTR4-depletion samples clustered together (Figure 1D). This cluster was closely related to EXOSC3-depletion but was more distinct from the PABPN1-depletion sample (Figure 1D). These results are consistent with reports that ZFC3H1 and MTR4 form a tight complex that weakly associates with PABPN1.[5] Together, our PAS-seq results suggest that, among all polyadenylated transcripts of annotated human genes, IPA transcripts are the major targets of PAXT/exosome-mediated degradation.

Although a large portion of IPA transcripts accumulated to significantly higher levels in EXOSC3-, MTR4-, or ZFC3H1-depleted cells (Figure 1A, marked by arrows), some remain unchanged. To characterize the molecular basis for such differences, we compared the 5′ splice site (ss) strength, as measured by the MaxEnt score,[25] of PAXT-regulated unstable IPA transcripts ($\log_2$FC > 1, FDR ≤ 0.05, in EXOSC3-, MTR4-, and ZFC3H1-depleted cells, N = 507) and stable IPA transcripts (counts per million [CPM] in control cells > 1, $-1 < \log_2$FC < 1, FDR > 0.05, N = 687). We found that the 5′ ss was modestly, albeit significantly, stronger for PAXT-regulated IPA transcripts (median: 8.69) than for stable IPA transcripts (median: 8.17) ($p < 0.0001$, Mann-Whitney test) (Figure 1E). We next investigated whether the distance between the 5′ ss and the IPA site may impact PAXT-dependent destabilization. Notably, for IPA transcripts in which the upstream 5′ ss was located within 1 kilobase (kb) of the IPA site, the increased 5′ ss strength of PAXT-regulated IPA transcripts (median: 8.56) compared with stable IPA transcripts (median: 7.58) was even more pronounced ($p < 0.0001$, Mann-Whitney test) (Figure S1D). By contrast, there was no significant difference in 5′ ss strength between PAXT-regulated and stable IPA transcripts when the 5′ ss was located more than 1 kb upstream of the PAS (Figure S1E). This indicates that a strong 5′ ss located near a PAS on IPA transcripts correlates with more efficient PAXT/exosome-mediated degradation.

Sequence determinants of PAXT-mediated degradation of IPA transcripts

How are IPA transcripts specifically recognized by PAXT? As all IPA transcripts contain a 5′ ss and a downstream poly(A) tail, we tested their roles in degradation using reporter assays. To mimic IPA transcripts, we inserted a 5′ ss sequence from the *NXF1* gene[26] into the 3′ untranslated region (UTR) of an EGFP reporter (Figure 2A). To test the role of the 5′ ss, we used the wild-type (WT) or a non-functional mutant (Mut) form of the 5′ ss. To test the impact of the poly(A) tail, we included in these reporters either a PAS from the bovine growth hormone (*bGH*) gene or the adenovirus major late transcript (L3) or a sequence that conferred a non-polyadenylated 3′ end (the 3′ end sequence of the nuclear lncRNA *MALAT1* or the replication-dependent histone gene *H2AC18*) (Figure 2A).[27,28] These reporters were transfected into HEK293T cells, and their mRNA levels relative to those of a co-transfected control reporter were measured by quantitative reverse-transcription PCR (RT-qPCR). Strikingly, we observed that the presence of a 5′ ss in the 3′ UTR of a polyadenylated reporter caused significant decreases in the mRNA levels (4.0-fold for the *bGH* PAS-containing reporter and 7.6-fold for the L3 PAS-containing reporter, $p < 0.0001$, unpaired t test) (Figure 2B). The 5′ ss induced only a modest decrease (1.6-fold, $p = 0.01$, unpaired t test) in the reporter with the histone *H2AC18* 3′ end, consistent with a previous study.[29] By contrast, a 5′ ss induced a mild increase in the mRNA levels in the reporter with the *MALAT1* 3′ end (1.5-fold, $p = 0.0007$, unpaired t test, Figure 2B). These results suggest that both a 5′ ss and a poly(A) tail are required to efficiently repress reporter mRNA levels.

The 5′ ss and PAS could repress reporter mRNA levels by inhibiting RNA synthesis and/or by promoting RNA degradation. To distinguish between these possibilities, we performed reporter assays in control or exosome-depleted cells. To deplete the exosome, we generated a HEK293T cell line in which a $FKBP12^{F36V}$ degron was fused to the C terminus of EXOSC3 via CRISPR-Cas9.[30] Treatment of this cell line with the dTAG molecule for 8 h led to near complete depletion of EXOSC3 protein (Figure S2A). We transfected reporters that contained a WT or a Mut 5′ ss and the *bGH* PAS or the *MALAT1* 3′ end sequence into the control (DMSO) or dTAG-treated EXOSC3 degron cell line and measured the reporter mRNA levels. Compared with DMSO-treated cells, EXOSC3 depletion led to a significant increase in the mRNA level of the 5′ ss-*bGH* PAS reporter (3.4-fold, $p = 0.004$, unpaired t test) but not for other reporters (Figure 2C). In fact, the mRNA level of the 5′ ss-*bGH* PAS reporter in EXOSC3-depleted cells was comparable with that of the Mut 5′ ss-containing reporter in untreated cells (compare Figures 2C and 2B). These results suggest that a 5′ ss and poly(A) tail repressed the reporter mRNA levels primarily by promoting exosome-mediated degradation. To determine whether such degradation was mediated by PAXT, we fused a $FKBP12^{F36V}$ degron to the N terminus of ZFC3H1 (Figure S2B). dTAG-induced depletion of ZFC3H1 caused a significant increase in the mRNA levels of the 5′ ss-PAS reporter (2.4-fold, $p = 0.004$, unpaired t test), but not for other reporters (Figure 2D), suggesting that the 5′ ss and poly(A) tail-induced exosomal degradation was indeed mediated by PAXT. We noted that the increase in mRNA levels in ZFC3H1-depleted cells was lower than that observed in EXOSC3-depleted cells (2.4- vs. 3.4-fold, Figures 2C and 2D). One likely reason for this discrepancy is that ZFC3H1 has multiple isoforms and the dTAG treatment primarily depleted the largest isoform (Figure S2B).

Next, we artificially recruited U1 snRNP to a 3′ UTR sequence without a 5′ ss and monitored its effect on RNA stability. To this end, we overexpressed a Mut U1 small nuclear RNA (snRNA) that can hybridize to the Mut 5′ ss in the *bGH* PAS-containing reporter (Figure S2C).[31] Compared with the control non-targeting U1 snRNA, overexpression of the Mut U1 snRNA led to a 1.8-fold decrease in the Mut 5′ ss-containing mRNA level ($p = 0.002$, unpaired t test) (Figure S2D), suggesting that targeting U1 snRNPs to the 3′ UTR is sufficient to induce RNA degradation. To further test the effect of U1 snRNP binding on RNA stability, we tested a series of reporters that contained a 5′ ss sequence of different strengths followed by the L3 PAS and observed a strong inverse correlation between the 5′ ss strength

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

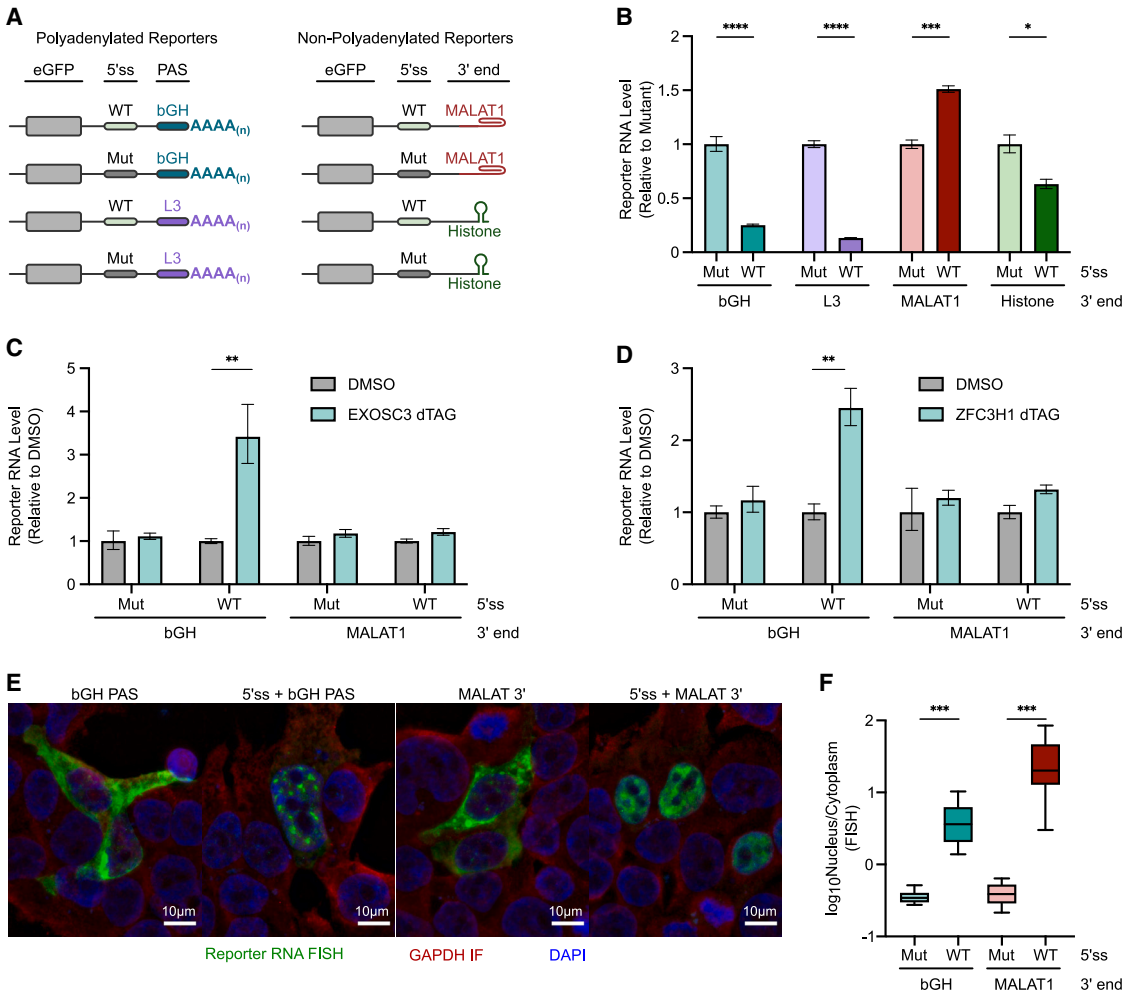


Figure 2. The combination of a 5′ splice site and a poly(A) tail triggers PAXT-dependent exosomal degradation

(A) Schematic of the EGFP reporter mRNAs. The reporters contain the coding sequence for EGFP followed by either a wild-type (WT) or mutant (Mut) 5′ ss in the 3′ UTR. The 3′ end of the reporter includes a poly(A) site (PAS), the 3′ end sequence from *MALAT1* (MALAT1), or *H2AC18* (histone).

(B) RT-qPCR analysis of reporter mRNA levels (normalized to the mRNA levels of a co-transfected control plasmid and to that the Mut 5′ ss reporter). Data are presented as mean ± SEM ($n = 3$). *p value ≤ 0.05, **p value ≤ 0.01, ***p value ≤ 0.001, ****: p-value ≤ 0.0001.

(C and D) RT-qPCR analysis of reporter mRNA levels with and without dTAG-induced depletion of EXOSC3 (C) or ZFC3H1 (D). The reporter mRNA levels were normalized to that of a co-transfected control plasmid and to that in DMSO-treated cells. Data are presented as mean ± SEM ($n = 3$). *p value ≤ 0.05, **p value ≤ 0.01, ***p value ≤ 0.001, ****p value ≤ 0.0001.

(E) Combined reporter RNA-FISH and GAPDH immunofluorescence (IF) in HEK293T cells. The RNA-FISH signal is shown in green and GAPDH IF signal in red. Nuclei were stained using DAPI (blue).

(F) Boxplot displaying the quantification of the reporter RNA-FISH signal in the nucleus vs. cytoplasm in (E) (see STAR Methods). For each 3′ end, the nucleus/cytoplasm FISH signal ratio of the WT 5′ ss reporter was normalized to that of the Mut. Statistical analysis from $n = 10$ (5′ ss-*bGH*, 5′ss Mut *bGH*), $n = 12$ (5′ss-*MALAT1* 3′ end), and $n = 11$ (5′ss Mut *MALAT1* 3′ end) images per reporter was performed using unpaired t tests. *p value ≤ 0.05, **p value ≤ 0.01, ***p value ≤ 0.001, ****p value ≤ 0.0001.

(MaxEnt score) and the mRNA levels (Figure S2E). Together, these results suggest that U1 snRNP binding near a poly(A) tail is sufficient to trigger RNA degradation.

To understand where the 5′ ss- and poly(A) tail-containing reporter mRNAs are degraded, we performed RNA fluorescence *in situ* hybridization (RNA-FISH). Without a 5′ ss in their 3′ UTR, the reporter mRNAs that contained either a poly(A) tail or the *MALAT1* 3′ end primarily localized to the cytoplasm (Figures 2E and 2F). By contrast, the mRNAs of both 5′ ss-containing reporters were retained in the nucleus (Figures 2E and 2F). This is consistent with previous studies showing that a 5′ ss promotes nuclear retention of both mRNAs and non-coding RNAs.[22,26,32] In line with the nuclear sequestration of their mRNAs, the EGFP protein levels of all 5′ ss-containing reporters were significantly lower compared with their counterparts without a 5′ ss (Figures S2G–S2I). Interestingly, although the mRNAs of both 5′ ss-PAS and 5′ ss-*MALAT1* 3′ end reporters were retained in the nucleus, their distribution patterns differed. The mRNAs of the 5′

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

ss-*MALAT1* 3′ end reporter were uniformly distributed in the nucleoplasm (Figure 2E, fourth panel). By contrast, the 5′ ss-PAS reporter mRNAs were highly concentrated in a number of large nuclear puncta (Figure 2E, second panel), which partially overlapped with nuclear speckles (Figure S2F). Together, these results suggest that although a 5′ ss alone induces RNA nuclear retention, the combination of a 5′ ss and a poly(A) tail promotes nuclear retention and PAXT-mediated exosomal degradation. Furthermore, we found that changing the location of the 5′ ss from the 3′ UTR to the coding sequence had little, if any, effect on the reporter mRNA levels (Figures S2J and S2K), consistent with a nuclear decay mechanism.

U1 snRNP and CPA factors cooperatively recruit PAXT and the exosome

To investigate how the combination of a 5′ ss and a poly(A) tail promotes PAXT-dependent exosomal RNA degradation, we performed *in vitro* RNA pull-down assays (Figure 3A). RNAs with a WT or Mut 5′ ss sequence upstream of the L3 PAS were synthesized by *in vitro* transcription. Three copies of the bacteriophage MS2 hairpin were fused to the 5′ end of the RNAs to allow for pull-down using a fusion protein comprised of the MS2 coat protein (MCP) and the maltose-binding protein (MBP).[33] The 3×MS2-fused RNAs were first incubated with the MCP-MBP fusion protein and nuclear extract from HeLa cells to allow for protein-RNA interactions and the CPA reactions to occur. As a negative control, a Mut L3 PAS was used in which the poly(A) signal was mutated from AAUAAA to AAGAAA. As shown in Figure 3B, CPA occurred for RNAs that contained the WT L3 PAS but not the Mut PAS (Figure 3B, compare lanes 1–6 with lanes 7–12), and the presence of an upstream 5′ ss did not have a significant effect on the CPA reaction efficiency (Figure 3B, compare lanes 8 and 11). Interestingly, upon longer incubation (150 min), a distinct higher band appeared for the 5′ ss-containing RNA (Figure 3B, compare lanes 12 and 9), suggesting that the 5′ ss induced hyperadenylation. Although the functional significance of such hyperadenylation is currently unclear, our results are consistent with previous reports that RNAs targeted by PAXT and PABPN1-dependent degradation are hyperadenylated.[8,34]

We next pulled down these RNAs using amylose beads and examined the associated proteins by western blotting. As expected, the WT L3 PAS-containing RNAs specifically pulled down CPA factors, including CPSF30 and PABPN1 (Figure 3C, lanes 4 and 5, and Figure S3C, lanes 4 and 5), and the U1 snRNP component U1-70K specifically associated with WT 5′ ss-containing RNAs (Figure 3C, lanes 3 and 5, and Figure S3C, lanes 3 and 5), demonstrating the specificity of this assay. Interestingly, both WT 5′ ss-Mut PAS and Mut 5′ ss-WT PAS RNAs were weakly associated with PAXT core subunits, ZFC3H1 and MTR4 (Figure 3C, lanes 3 and 4, and Figure S3C, lanes 3 and 4). Strikingly, however, dramatically higher amounts of ZFC3H1 and MTR4 were precipitated with the RNAs that contained both a WT 5′ ss and a WT PAS (Figure 3C, compare lane 5 with lanes 3–4, and Figure S3C, compare lane 5 with lanes 3–4). Analysis of these RNA pull-down assays by quantitative western blotting revealed that ZFC3H1 and MTR4 were enriched 8.5- and 3.8-fold, respectively, on RNAs that contained both a WT 5′ ss and a WT PAS compared with a WT PAS alone (Figure S3C, compare lane 5 to lane 4). In contrast, ZCCHC8, a component of the NEXT adaptor complex, was not significantly precipitated (Figures 3C and S3C). These results strongly suggest that both U1 snRNP and CPA factors bind weakly to PAXT individually, but, when bound to the same RNA molecule, they can synergistically recruit PAXT.

To further test whether PAXT, U1 snRNP, and CPA factors form a complex on the same RNAs, we resolved the complexes assembled on the RNAs that contained both a WT 5′ ss and a WT PAS by glycerol gradient sedimentation and monitored their protein composition by western blotting (Figure 3D). We found that there was a peak centered around fractions 14–15 that contained PAXT (ZFC3H1, MTR4, and PABPN1), U1 snRNP components (U1A and U1-70K), and CPA factors (CPSF30). Together, these results strongly suggest that the PAXT core components ZFC3H1 and MTR4 are cooperatively recruited to RNA in a U1 snRNP- and CPA-factor-dependent manner.

To test whether the CPA reaction itself was required for PAXT recruitment, we synthesized pre-polyadenylated versions of the same RNA substrates and performed RNA pull-down assays (Figure 3A). To do so, we synthesized shorter versions of the PAS-containing RNAs that ended at the natural cleavage sites by *in vitro* transcription and polyadenylated them using *E.coli* poly(A) polymerase. Each pre-polyadenylated RNA substrate contained a poly(A) junction (PAJ), which includes the upstream PAS region and the poly(A) tail (Figure 3A, right). To test the role of the poly(A) signal itself, we generated pre-polyadenylated RNAs with a WT (AAUAAA) or Mut (AAGAAA) hexamer. We confirmed that the pre-polyadenylated RNA substrates were of comparable length and that the poly(A) tails were not further extended after incubation in nuclear extract (Figures S3A and S3B). These RNAs were incubated with HeLa nuclear extract and subsequently pulled down using amylose beads. As all of the RNAs were pre-polyadenylated, PABPN1 was precipitated with all of them (Figures 3E and S3D). For the RNAs that contained a Mut 5′ ss and a Mut poly(A) signal, neither ZFC3H1 nor MTR4 was precipitated (Figure 3E, lane 2, and Figure S3D, lane 2), demonstrating that PABPN1 alone was not sufficient to recruit PAXT. As expected, the WT 5′ ss-containing RNAs pulled down the U1 snRNP component U1-70K and the WT poly(A)-signal-containing RNAs precipitated CPA factors, such as CPSF30 (Figures 3E and S3D). Interestingly, the RNAs containing a Mut 5′ ss-WT poly(A) signal or those containing a WT 5′ ss-Mut poly(A) signal only pulled down low amounts of ZFC3H1 and MTR4 (Figure 3E, lanes 3 and 4, and Figure S3D, lanes 3 and 4), similar to the corresponding full-length pre-mRNA substrates (Figure 3C, lanes 3 and 4, and Figure S3C, lanes 3 and 4). By contrast, the polyadenylated RNAs that contained both a WT 5′ ss and a WT poly(A) signal pulled down much higher amounts of ZFC3H1, MTR4, and EXOSC3 (Figure 3E, compare lane 5 to lanes 3–4, and Figure S3D, compare lane 5 to lanes 3–4). Specifically, ZFC3H1 and MTR4 were enriched 14.5- and 3-fold, respectively, on pre-polyadenylated RNAs that contained both a WT 5′ ss-WT poly(A) signal compared with those containing a Mut 5′ ss-WT poly(A) signal (Figure S3C, compare lane 5 to lane 3). As pre-polyadenylated RNAs containing a WT 5′ ss and a WT poly(A) signal can pull down PAXT, this indicates that PAXT recruitment does not

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

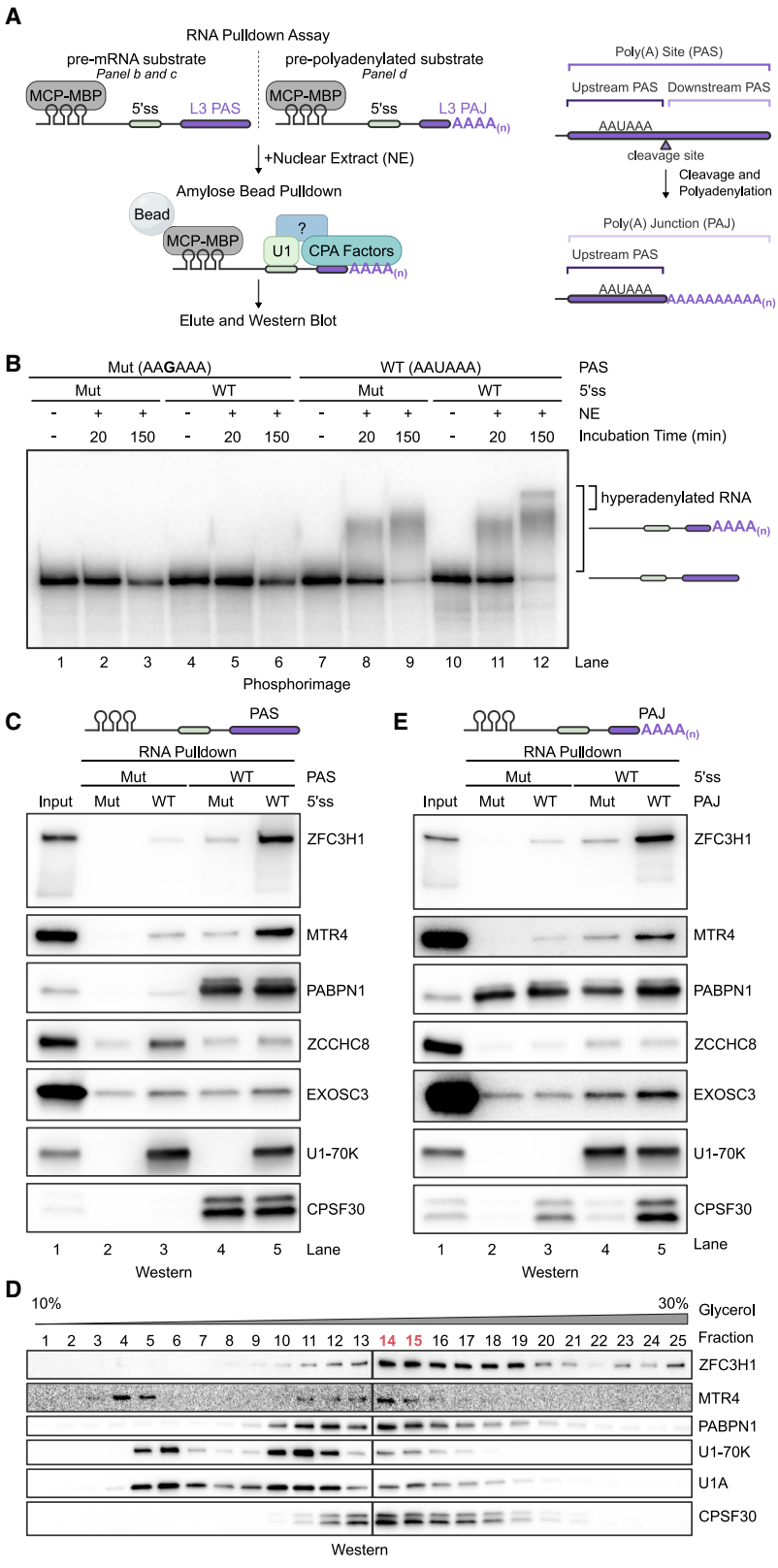


Figure 3. U1 snRNPs and cleavage and polyadenylation factors cooperatively recruit PAXT and the RNA exosome

(A) Left: schematic of the RNA substrates and the RNA pull-down assays. The depicted pre-mRNA substrate was used in (B) and (C) and the pre-polyadenylated substrate was used in (D). Right: schematic depicting terms used to describe RNA regions.

(B) Cleavage/polyadenylation assay using radiolabeled pre-mRNA substrates. Pre-mRNAs and polyadenylated and hyperadenylated RNAs were marked.

(C) Western blotting analyses of proteins associated with the four RNAs.

(D) Western blotting analysis of protein complexes assembled on a 5′ ss- and PAS-containing pre-mRNA substrate resolved on a 10%–30% glycerol gradient.

(E) Western blotting analyses of proteins associated with the four pre-polyadenylated substrates.

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

require the CPA reaction. On the other hand, when we normalized the amounts of PAXT to the amount of polyadenylated RNAs (see STAR Methods for details), we observed that ZFC3H1 was recruited 11-fold more efficiently to 5′ ss-containing polyadenylated RNA that underwent the CPA reaction (Figures 3C and S3C) as compared with pre-polyadenylated RNA (Figures 3E and S3D) ($p = 0.03$, unpaired t test) (Figure S3E), indicating that the CPA reaction promotes PAXT recruitment.

Together, our RNA pull-down results demonstrate that both 5′ ss-bound U1 snRNP or PAJ-bound CPA factors displayed weak interactions with PAXT. By contrast, when U1 snRNP and CPA factors bind to their cognate sequences on the same RNA, they synergistically recruit PAXT. Furthermore, in addition to the poly(A) tail, our results demonstrated an essential role for upstream PAS sequences in PAXT recruitment.

ZFC3H1 interacts with U1 snRNP and CPA factors

Our RNA pull-down results suggest that both U1 snRNP and CPA factors can weakly associate with PAXT (Figures 3C and 3E). Consistently, U1-70K and CPA factors, such as CPSF subunits, were detected in the interactome of ZFC3H1,[7,22] but not in that of MTR4.[7] ZFC3H1 contains four major regions/domains: an acidic short linear motif at the N terminus called the short acid-rich linear motif (SLiM) region, five coiled-coil (CC) domains, a zinc-finger domain (ZnF), and tetratricopeptide repeats (TPRs) near its C terminus (Figure 4A). We next sought to characterize the interactions between ZFC3H1 and U1 snRNP or CPA factors in detail. To this end, we expressed FLAG-tagged ZFC3H1 in HEK293T cells and performed immunoprecipitations using anti-FLAG antibodies (FLAG-IP) (Figure 4B). Consistent with previous reports,[7,35] PABPN1, MTR4, and the cap-binding complex interacting protein ARS2 were co-immunoprecipitated with ZFC3H1 (Figure 4B, lanes 6 and 7). ZFC3H1 interacted with MTR4 and ARS2 in an RNase A-resistant manner and, consistent with previous studies,[7] its association with PABPN1 was severely reduced upon RNase treatment (Figure 4B, lanes 6 and 7). Low levels of U1 snRNP subunits, including U1-70K, U1A, and U1C, as well as CPSF components, including CPSF100, CPSF73, and CPSF30, were co-immunoprecipitated with ZFC3H1 in a partially RNA-dependent manner (Figure 4B, lanes 6 and 7). These interactions were further validated by reciprocal FLAG-IPs (Figures S4A and S4B), suggesting that ZFC3H1 can indeed interact with both U1 snRNPs and CPSF.

To map the regions in ZFC3H1 that mediate interactions with U1 snRNPs and CPSF, we performed FLAG-IPs using ZFC3H1 mutants in which a specific region was deleted (Figure 4A). To broadly define the regions in ZFC3H1 that mediate interactions with U1 snRNPs and CPSF, we overexpressed FLAG-tagged N-terminal (residues 1–990) or C-terminal (residues 991–1,989) regions of ZFC3H1 and performed IPs with anti-FLAG antibodies. As previously reported,[35] ARS2 and PABPN1 were co-precipitated with the N-terminal region, whereas MTR4 specifically associated with the C-terminal region of ZFC3H1 (Figure 4B, lanes 8 and 9, lanes 10 and 11).[35] Both U1 snRNPs and CPSF associated specifically with the N-terminal region of ZFC3H1 (Figure 4B, lanes 8–11), and its interaction with U1 snRNPs appeared to depend more on RNA than that with CPSF (Figure 4B, compare lanes 8 and 9).

We next further dissected the N-terminal region of ZFC3H1 and characterized the role of the SLiM region and the CC domains in mediating interactions with U1 snRNPs and CPSF. We first mutated the SLiM region present within residues 12–33 of ZFC3H1 (abbreviated SLiM Mut, Figure 4A). The SLiM region contains a highly acidic motif (EEGELEDGEI) and, consistent with previous studies,[36,37] we observed that substituting four of the acidic residues in the SLiM motif with alanine (EAGALEAGAI) abolished the interaction between ZFC3H1 and ARS2 (Figure 4C, lanes 6–7, compare with Figure 4B, lanes 6–7). By contrast, these mutations in the SLiM region did not affect ZFC3H1 interactions with U1 snRNP, CPSF, or PABPN1 (Figure 4C, lanes 6 and 7), suggesting that the SLiM motif is not necessary for ZFC3H1 to interact with these factors.

We then focused on the N-terminal region of ZFC3H1 that contains five CC domains (residues 50–990). We expressed two FLAG-tagged sub-regions that contained three and two CC domains (residues 50–599 and 600–990, respectively) and performed FLAG-IPs (Figure 4A). Subunits of both U1 snRNP and CPSF components interacted most strongly with residues 50–599 of ZFC3H1 (Figure 4C, compare lanes 8 and 9 with 10 and 11). The interactions between residues 50–599 of ZFC3H1 and U1 snRNP components were sensitive to RNase A treatment, whereas the interactions with CPSF components were partially resistant (Figure 4C, compare lanes 8 and 9). In contrast, PABPN1 interacted most strongly with residues 600–990 of ZFC3H1 (Figure 4C, compare lanes 8 and 9 with lanes 10 and 11), consistent with previous reports.[35] These interactions were again validated by reciprocal FLAG-IPs (Figures S4C and S4D). Summarized in Figure 4D, these interaction domain mapping results demonstrate that both U1 snRNPs and CPSF associate with the same region in ZFC3H1.

Given that U1 snRNPs and CPSF bind to a similar region in ZFC3H1 (Figures 4B and 4C), we hypothesized that the formation of a U1-CPA-PAXT-RNA complex (Figure 3D) would require multiple copies of ZFC3H1 in this complex. Indeed, ZFC3H1 was predicted to form a dimer.[36] To test whether ZFC3H1 forms dimers, we co-expressed a FLAG-tagged and a Strep-tagged ZFC3H1 and performed a FLAG-IP (Figure S4E). The Strep-tagged version of ZFC3H1 was co-immunoprecipitated with FLAG-ZFC3H1 in an RNA-independent manner, consistent with self-interaction (Figure S4E, lanes 7 and 8). ZFC3H1 dimerization is predicted to be mediated by residues 813–932,[36] which span the entire fourth CC domain (Figure 4A). We next attempted to disrupt self-interaction by mutating 13 leucine residues to arginine within the fourth CC domain (CC4 Mut), deleting the fourth CC domain (ΔCC4), or deleting all residues predicted to support dimerization (Δ813-932). Interestingly, however, self-interaction persisted in all three ZFC3H1 mutants (Figure S4E). This led us to conclude that ZFC3H1 self-interaction is extensive and involves multiple regions. Consistently, while this study was under review, a study was published that reported that ZFC3H1 self-interacts and that this self-interaction is supported by at least two discrete regions within ZFC3H1, including the CC4 region.[38]

The NRDC and human diseases

Our results demonstrate that the combination of a 5′ ss and a PAJ, which we refer to as an NRDC, targets RNAs for

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

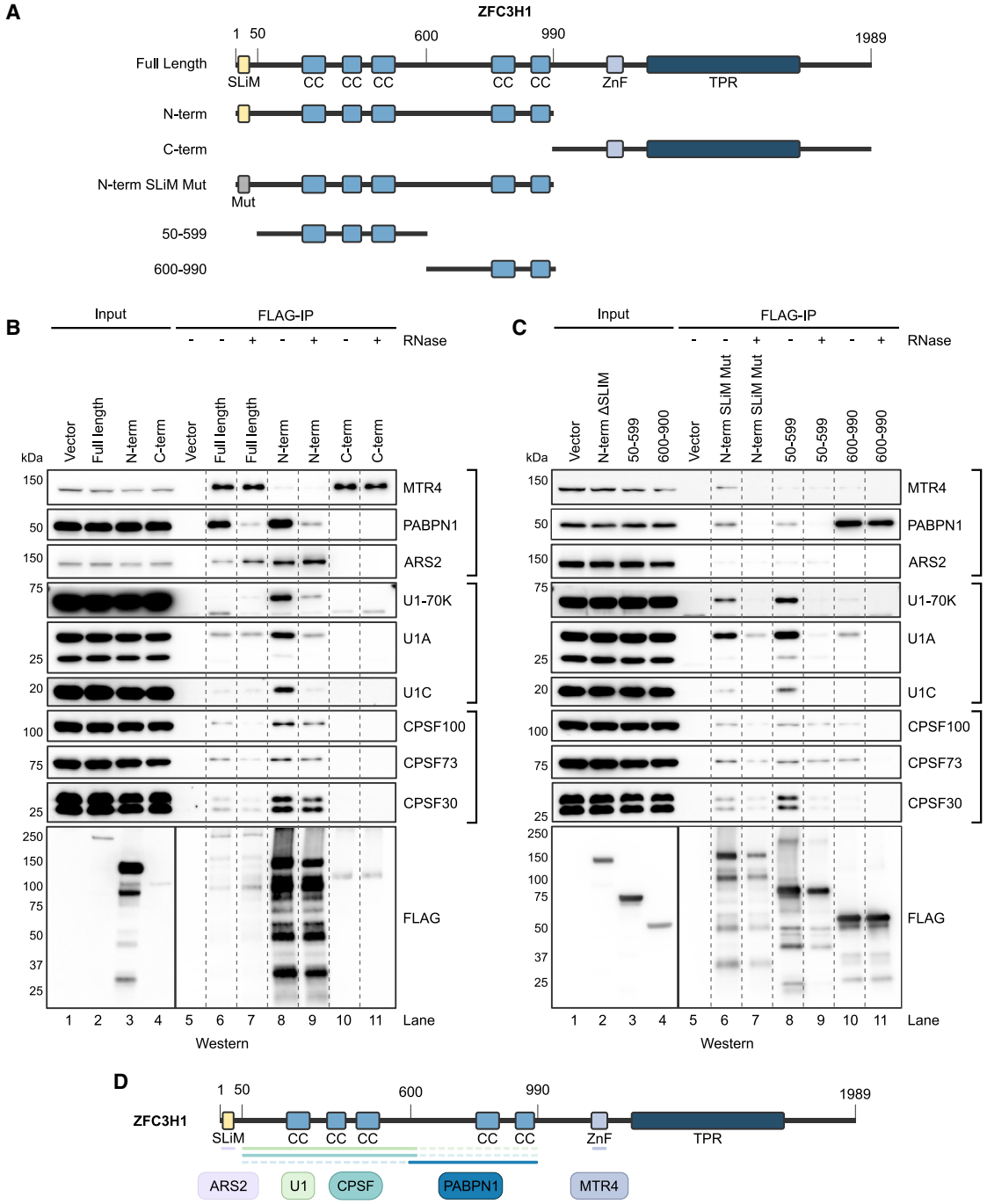


Figure 4. ZFC3H1 interacts with U1 snRNPs and cleavage and polyadenylation factors

(A) Schematic of the predicted domains within full-length and truncated ZFC3H1 used for FLAG-IP in (B) and (C). Relevant residues are indicated as numbers above full-length ZFC3H1. SLiM, short acid-rich linear motif; CC, coiled-coil; ZnF, zinc-finger domain; TPR, tetratricopeptide repeat.

(B and C) Western blotting analyses of the indicated FLAG-IP. The anti-FLAG western blots contain 1% input and 10% eluted IP sample. All other western blots contain 0.5% input and 30% eluted IP sample.

(D) Schematic depicting regions in ZFC3H1 that mediate interactions with factors shown on the bottom.

PAXT-mediated exosomal degradation. We next explored how this mechanism may contribute to human diseases. Previous studies identified an SNP in the 3′ UTR of the human gene *Factor IX* (*F9*)—which encodes a blood coagulation factor—that caused significant downregulation of its protein levels and, in turn, severe hemophilia B.[39] Similarly, an SNP in the 3′ UTR of the

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

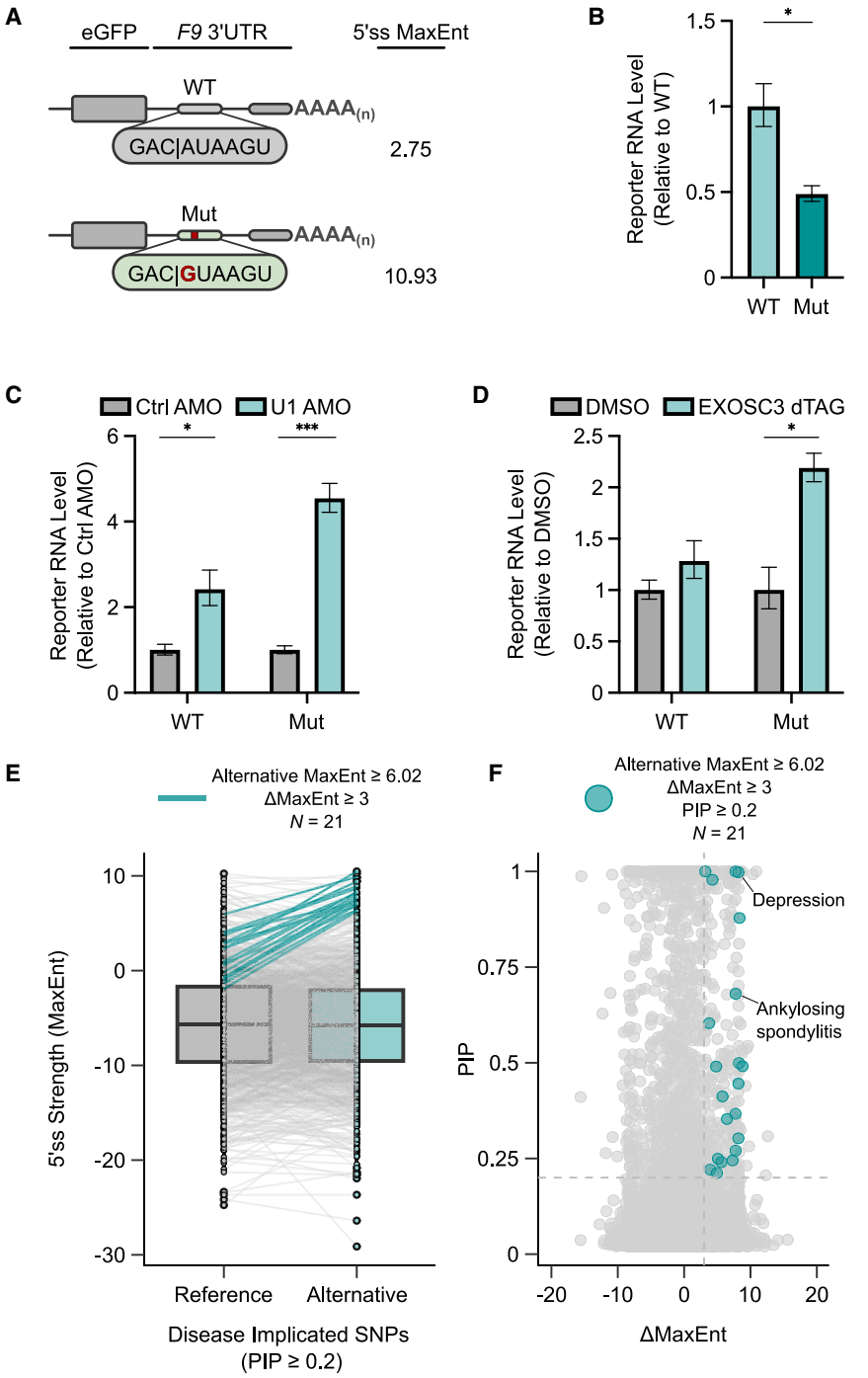


Figure 5. SNP-activated U1 snRNP-dependent mRNA decay in human diseases

(A) Schematic depicting the EGFP-F9 reporters used in (B)–(D). Each reporter contains the EGFP coding sequence followed by the WT or patient mutant (Mut) 3′ UTR from the *F9* gene. The 9 nucleotides shown were used to compute the 5′ ss strength by MaxEnt. The A > G SNP is shown in red.

(B) RT-qPCR analysis of reporter mRNA levels (normalized to those of a co-transfected control reporter and to the *F9* WT reporter). Statistical analysis from $n = 3$ independent samples was calculated using an unpaired t test. Data are presented as mean ± SEM. **p* value ≤ 0.05.

(C and D) RT-qPCR analysis of reporter mRNA levels in Ctrl or U1 AMO-treated cells (C) or in the DMSO- or dTAG-treated EXOSC3 degron cell line (D). The reporter mRNA levels were normalized to those of a co-transfected reporter and to Ctrl AMO-treated (C) or DMSO-treated cells (D). Statistical analysis from $n = 3$ independent samples was calculated using unpaired t tests. Data are presented as mean ± SEM. **p* value ≤ 0.05, ***p* value ≤ 0.01, ****p* value ≤ 0.001.

(E) Boxplot with paired points depicting the 5′ ss strength of all reference and alternative allele variants with a PIP score ≥ 0.2. Lines connect each reference and variant SNP MaxEnt score. Variants marked with a teal line met our threshold for NRDC-activating SNPs.

(F) Scatterplot depicting the change in MaxEnt score (ΔMaxEnt) and the PIP score for all analyzed variants. Variants marked with a teal circle have a PIP score ≥ 0.2 and increase the reference allele 5′ ss strength by at least 3 MaxEnt units to an alternative allele 5′ ss strength ≥ 6.02 (1 IQR below the median 5′ ss strength across all 5′ splice sites). The variant-associated disease or trait is marked for two SNPs.

p14/*LAMTOR2* gene leads to downregulation of *p14* mRNAs and causes p14 deficiency, a primary immunodeficiency syndrome.[40,41] In both cases, the SNP created a novel 5′ ss that was suggested to cause mRNA downregulation via U1 snRNP-mediated repression of pre-mRNA 3′ end processing,[39,42] but direct evidence has been lacking. By contrast, our model predicts that these SNPs should induce mRNA degradation by the exosome through the NRDC mechanism. To test this, we created reporters by inserting the WT or SNP-containing *F9* or *p14* 3′ UTR sequences, including the PAS, downstream of the EGFP or *p14* coding sequence, respectively (see STAR Methods) (Figures 5A and S5A). When we expressed these reporters in HEK293T cells, we observed a 2.0-fold ($p = 0.01$, unpaired t test) and a 6.3-fold ($p < 0.0001$, unpaired t test) decrease in the mRNA level of the SNP-containing Mut compared with that of the WT for *F9* and *p14*, respectively (Figures 5B and S5A). Using the *F9* reporter as an example, we showed that the SNP-induced decrease in mRNA levels was completely reversed by U1 antisense morpholino oligo (AMO) (4.5-fold increase, $p = 0.0002$, unpaired t test), which blocks U1 snRNP-RNA interactions (Figure 5C), consistent with a previous report.[42] This result strongly suggests that the SNP-induced reduction in *F9* mRNA levels is mediated by U1 snRNP. To test whether the RNA exosome was involved, we expressed both *F9* and *p14* reporters in the control (DMSO) or dTAG-treated EXOSC3 degron cell line. Strikingly, dTAG-induced

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

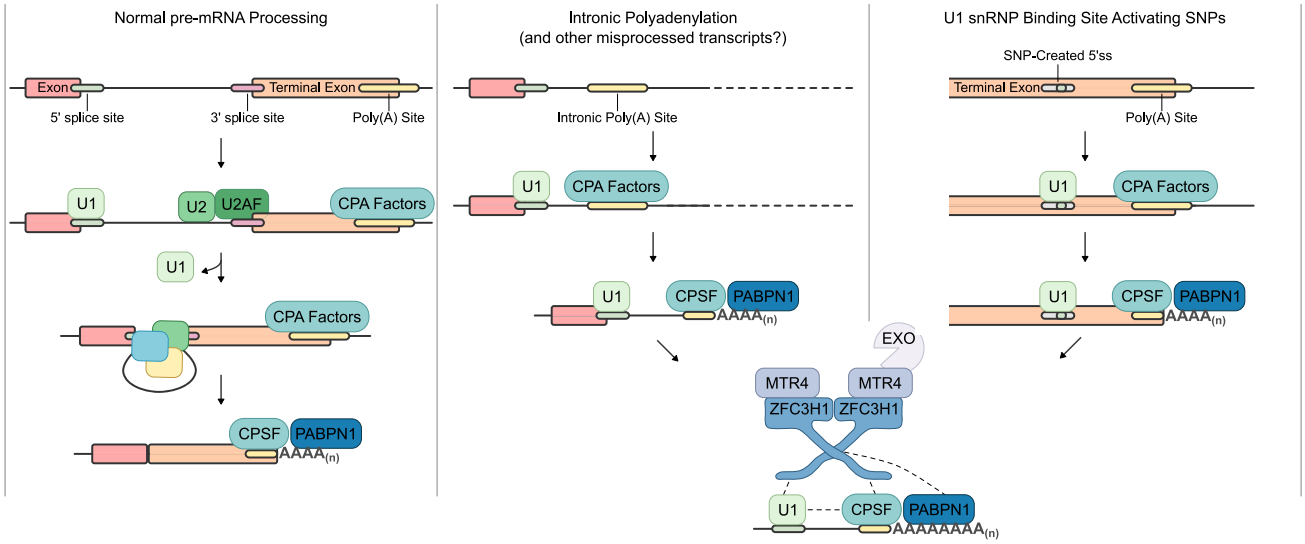


Figure 6. Nuclear RNA degradation code

This diagram illustrates normal pre-mRNA processing (left), IPA processing (middle), and the 3′ processing of RNAs containing SNP-created novel 5′ ss in their 3′ UTRs (right) and how the latter two RNAs recruit PAXT and the RNA exosome. Please see discussion for details.

EXOSC3 depletion restored the SNP-containing reporter mRNA levels to that of the WT reporter (*F9*: 2.2-fold increase, $p = 0.02$; *p14*: 3.4-fold increase, $p < 0.0001$, unpaired t test) but had little effect on the WT reporters (Figures 5D and S5B). These results strongly suggest that the SNP-created 5′ ss in the *F9* and *p14* 3′ UTR suppresses their mRNA levels by promoting exosome-mediated mRNA degradation. We will refer to this type of SNP as an NRDC-activating SNP. Furthermore, our data provided evidence that exosome inhibition and antisense oligos that block an SNP-created 5′ ss may be suitable strategies for treating diseases caused by NRDC-activating SNPs.

To identify additional NRDC-activating SNPs that are associated with human diseases, we extracted 7,848 fine-mapped SNPs within the annotated 3′ UTRs of human genes from the CAUSALdb[43] and UK Biobank[44] databases and identified those that create novel 5′ ss. We selected 106 SNPs that increased the 5′ ss MaxEnt scores of the surrounding sequence by at least 3 units to 6.02 or higher, which is 1 interquartile range (IQR) below the median 5′ ss strength across all annotated introns (mean: 8.66, IQR: 2.64). Next, we filtered these SNPs based on their posterior inclusion probability (PIP) scores, which is a measurement of the probability that a variant is a causal factor for a disease or trait.[45] We selected those SNPs whose PIP score was at least 0.2, which has been used in previous studies to select causal variants.[46] Using these parameters, we identified 21 disease-associated SNPs that created novel 5′ ss (Figure 5E). These SNPs are associated with a wide range of diseases/disorders, including ankylosing spondylitis and depression (Figure 5F; Table S1). Similar to the aforementioned SNPs in *F9* and *p14*, we propose that these SNPs could contribute to diseases by promoting aberrant mRNA degradation via the NRDC mechanism. Conversely, we also identified 8 potential NRDC-inactivating SNPs. These SNPs are found in naturally occurring 5′ ss in 3′ UTRs (MaxEnt $\geq$ 6.02), indicating that these mRNAs may be naturally unstable due to the NRDC mechanism. These SNPs reduced the 5′ ss strength scores of the surrounding sequences by at least 3 units (Figures S5C and S5D; Table S1), which we predict would lead to stabilization of these mRNAs. These NRDC-inactivating SNPs included a variant strongly associated with schizophrenia (Figure S5D; Table S1). Taken together, our results provided evidence that genetic variations can contribute to human diseases by causing aberrant RNA degradation via the NRDC pathway.

DISCUSSION

A central question in the RNA degradation field has been how the substrate specificity of the RNA degradation enzymes/machinery is determined. Previous studies have identified several individual RNA sequence motifs that stimulate RNA degradation, such as AU-rich elements in 3′ UTRs,[47] a "determinant of selective removal" (DSR) sequence in meiotic transcripts,[48] and short sequence motifs in cryptic unstable transcripts (CUTs).[2] However, such destabilizing sequence features have not been identified for most unstable RNAs. In this study, we demonstrate that the combination of a 5′ ss and a PAJ, but not either sequence alone, drives exosome-mediated RNA degradation (Figure 6). This mechanism is fundamentally distinct from the single-motif-induced degradation mechanisms in that it requires a specific combination of multiple RNA features. 5′ ss are typically found in unspliced precursor RNAs, whereas PAJs are found in processed RNAs. The unusual combination of these sequences on the same RNA molecule, as occurs within IPA transcripts and misprocessed RNAs, serves as the degradation mark (Figure 6). Thus, this mechanism allows for cells to monitor the RNA processing status by using the processing sequences and machinery themselves. We speculate that the concept of the RNA degradation code, i.e., a combination of multiple RNA

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

features serving as the degradation mark, may be broadly applicable for other groups of exosome targets.

Given the NRDC mechanism, how do some IPA transcripts remain stable? We have observed an inverse correlation between the 5′ ss strength and RNA stability (Figure S2E) and that stable IPA transcripts tend to have weaker 5′ ss (Figures 1E and S1D). Thus, weaker U1 snRNP binding may allow these IPA transcripts to escape degradation. Further, it is well known that there are stable polyadenylated mRNAs that have retained introns.[49,50] Previous studies have provided evidence that, although these retained introns are not spliced, spliceosomes do assemble on them.[51] Therefore, the U1 snRNPs on these transcripts are most likely engaged with U2AF and other spliceosome components and, thus, cannot associate with CPA factors to synergistically recruit PAXT. Additionally, we provided evidence that the distance between the 5′ ss and the PAJ is also involved in modulating the degradation efficiency (Figures S1D and S1E).

In addition to its role in transcriptome surveillance, the NRDC may also play an important role in gene evolution. For example, many lncRNAs reside in the nucleus, and previous studies provided evidence that their nuclear retention is mediated by their interaction with U1 snRNPs.[26,32] Some of these nuclear lncRNAs are highly abundant and stable, including *MALAT1* and *NEAT1*. Interestingly, both *MALAT1* and the long isoform of *NEAT1*, *NEAT1_2*, have non-polyadenylated 3′ ends.[27] Our NRDC model predicts that both *MALAT1* and *NEAT1* would be targeted for exosomal degradation if they had poly(A) tails. Indeed, the 3′ end of the short isoform of *NEAT1*, *NEAT1_1*, is formed through the canonical CPA process,[52] and our data showed that it is targeted by PAXT-mediated exosomal degradation (Figures S6A–S6C). Thus, it is likely that both *MALAT1* and *NEAT1_2* have evolved non-polyadenylated 3′ ends to evade degradation via the NRDC mechanism.

Limitations of the study

Due to the extensive interactions involved in ZFC3H1 dimerization, we were not able to directly test whether such dimerization is required for PAXT-mediated RNA degradation. Additionally, our initial search for NRDC-activating SNPs serves as a proof-of-principle study and is by no means comprehensive. For example, none of the databases that we searched contained the disease-causing SNPs reported for *F9* and *p14*. Thus, more extensive analyses are needed in the future to uncover NRDC-activating genetic variations at the genome-wide level.

RESOURCE AVAILABILITY

Lead contact

Requests for further information or reagents should be directed to the lead contact, Yongsheng Shi (yongshes@uci.edu).

Materials availability

All reagents generated in this study are available upon request from the lead contact with a completed materials transfer agreement.

Data and code availability

- PAS-seq datasets have been deposited into the GEO (GSE267857). These data are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We would like to thank Drs. Shalini Sharma, Doug Black, Bert Semler, and François Bachand for providing reagents. We wish to acknowledge the support of the Chao Family Comprehensive Cancer Center Shared Resource Genomics High-Throughput Facility, supported by the National Cancer Institute of the National Institutes of Health, under award number P30CA062203. This study was supported by the following grants: R35GM149294 (Y.S.), R01CA297834 (Y.S. and G.S.), and NIH Director's New Innovator Award, DP2GM149554 (H.Y. and F.D.). L.T. is supported by the Hewitt Foundation Postdoctoral Fellowship. A.M.Y. was supported by the Washington Research Foundation postdoctoral fellowship.

AUTHOR CONTRIBUTIONS

Conceptualization, L.V.S. and Y.S.; investigation, L.V.S., L. Liu, X.Z., and Y.Y.; formal analysis, L.V.S., X.Z., A.M.Y., W.L., and Y.S.; methodology, L.V.S., H.Y., and L.T.; resources, H.Y.; validation, L.V.S., S.L., L. Liu, L.T., and M.V.; writing – original draft, L.V.S.; writing – review and editing, L.V.S. and Y.S.; funding acquisition, Y.S.; supervision, Y.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Generation of FLAG-dTAG Edited HEK293T Cell Lines
  - RNAi
  - PAS-seq
  - PAS-seq Data Processing
  - Differential Isoform Expression Analysis
  - Molecular Cloning
  - Reporter Assays
  - p14 Reporters
  - 5′ ss strength reporters
  - 5′ ss within coding sequence reporters
  - U1 snRNA Targeting of 5′ ss mutant reporter
  - Inhibition of U1 snRNP by AMO
  - RT-qPCR
  - RNA FISH and Immunofluorescence
  - *In vitro* Transcription
  - *In vitro* Polyadenylation Assay
  - MS2-MBP RNA Pulldown and Western Blotting
  - MS2-MBP RNA Pulldown and Glycerol Gradient Sedimentation
  - FLAG-Immunoprecipitation
  - SNP Analyses
  - Prediction of Nuclear RNA Degradation Code Activating and Inactivating SNPs
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - General Statistical Analysis
  - Reporter Assay Protein Level Quantification
  - *In vitro* Polyadenylation Assay Quantification
  - RNA Pulldown Assay Protein Quantification

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2025.03.010.

Received: July 5, 2024
Revised: January 9, 2025
Accepted: March 7, 2025
Published: April 17, 2025

REFERENCES

1. Kilchert, C., Wittmann, S., and Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. Nat. Rev. Mol. Cell Biol. *17*, 227–239. https://doi.org/10.1038/nrm.2015.15.
2. Schmid, M., and Jensen, T.H. (2018). Controlling nuclear RNA levels. Nat. Rev. Genet. *19*, 518–529. https://doi.org/10.1038/s41576-018-0013-2.
3. Dziembowski, A., Lorentzen, E., Conti, E., and Séraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat. Struct. Mol. Biol. *14*, 15–22. https://doi.org/10.1038/nsmb1184.
4. Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome. Cell *127*, 1223–1237. https://doi.org/10.1016/j.cell.2006.10.037.
5. Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. Mol. Cell *43*, 624–637. https://doi.org/10.1016/j.molcel.2011.06.028.
6. Wu, G., Schmid, M., Rib, L., Polak, P., Meola, N., Sandelin, A., and Jensen, T.H. (2020). A Two-Layered Targeting Mechanism Underlies Nuclear RNA Sorting by the Human Exosome. Cell Rep. *30*, 2387–2401.e5. https://doi.org/10.1016/j.celrep.2020.01.068.
7. Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., et al. (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. Mol. Cell *64*, 520–533. https://doi.org/10.1016/j.molcel.2016.09.025.
8. Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., Tian, B., and Manley, J.L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. Genes Dev. *31*, 1257–1271. https://doi.org/10.1101/gad.302604.117.
9. Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev. *11*, 2755–2766. https://doi.org/10.1101/gad.11.21.2755.
10. Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev. *25*, 1770–1782. https://doi.org/10.1101/gad.17268411.
11. Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., 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. https://doi.org/10.1101/gad.250993.114.
12. Chan, S., Choi, E.-A., and Shi, Y. (2011). Pre-mRNA 3′-end processing complex assembly and function: Pre-mRNA 3′-end processing complex assembly. Wiley Interdiscip. Rev. RNA *2*, 321–335. https://doi.org/10.1002/wrna.54.
13. Soles, L.V., and Shi, Y. (2021). Crosstalk between mRNA 3′-end processing and epigenetics. Front. Genet. *12*, 637705. https://doi.org/10.3389/fgene.2021.637705.
14. Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nat. Rev. Mol. Cell Biol. *18*, 18–30. https://doi.org/10.1038/nrm.2016.116.
15. Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. *17*, 156–165. https://doi.org/10.1101/gr.5532707.
16. Singh, I., Lee, S.-H., Sperling, A.S., Samur, M.K., Tai, Y.-T., Fulciniti, M., Munshi, N.C., Mayr, C., and Leslie, C.S. (2018). Widespread intronic polyadenylation diversifies immune cell transcriptomes. Nat. Commun. *9*, 1716. https://doi.org/10.1038/s41467-018-04112-z.
17. Lee, S.-H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C.S., and Mayr, C. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature *561*, 127–131. https://doi.org/10.1038/s41586-018-0465-8.
18. Insco, M.L., Abraham, B.J., Dubbury, S.J., Kaltheuner, I.H., Dust, S., Wu, C., Chen, K.Y., Liu, D., Bellaousov, S., Cox, A.M., et al. (2023). Oncogenic *CDK13* mutations impede nuclear RNA surveillance. Science *380*, eabn7625. https://doi.org/10.1126/science.abn7625.
19. Wang, R., Nambiar, R., Zheng, D., and Tian, B. (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. Nucleic Acids Res. *46*, D315–D319. https://doi.org/10.1093/nar/gkx1000.
20. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664–668. https://doi.org/10.1038/nature09479.
21. Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., et al. (2012). U1 snRNP Determines mRNA Length and Regulates Isoform Expression. Cell *150*, 53–64. https://doi.org/10.1016/j.cell.2012.05.029.
22. Lee, E.S., Smith, H.W., Wolf, E.J., Guvenek, A., Wang, Y.E., Emili, A., Tian, B., and Palazzo, A.F. (2022). ZFC3H1 and U1-70K promote the nuclear retention of mRNAs with 5′ splice site motifs within nuclear speckles. RNA *28*, 878–894. https://doi.org/10.1261/rna.079104.122.
23. Yoon, Y., Soles, L.V., and Shi, Y. (2021). PAS-seq 2: A fast and sensitive method for global profiling of polyadenylated RNAs. Methods Enzymol *655*, 25–35. https://doi.org/10.1016/bs.mie.2021.03.013.
24. Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA *17*, 761–772. https://doi.org/10.1261/rna.2581711.
25. Yeo, G., and Burge, C.B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. J. Comput. Biol. *11*, 377–394. https://doi.org/10.1089/1066527041410418.
26. Yin, Y., Lu, J.Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B., et al. (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. Nature *580*, 147–150. https://doi.org/10.1038/s41586-020-2105-3.
27. Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.-D., Joshua-Tor, L., and Sharp, P.A. (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. Genes Dev. *26*, 2392–2407. https://doi.org/10.1101/gad.204438.112.
28. Tan, D., Marzluff, W.F., Dominski, Z., and Tong, L. (2013). Structure of Histone mRNA Stem-Loop, Human Stem-Loop Binding Protein, and 3′hExo Ternary Complex. Science *339*, 318–321. https://doi.org/10.1126/science.1228705.
29. Fortes, P., Cuevas, Y., Guan, F., Liu, P., Pentlicky, S., Jung, S.P., Martínez-Chantar, M.L., Prieto, J., Rowe, D., and Gunderson, S.I. (2003). Inhibiting expression of specific genes in mammalian cells with 5′ end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA. Proc. Natl. Acad. Sci. USA *100*, 8264–8269. https://doi.org/10.1073/pnas.1332669100.
30. Nabet, B., Ferguson, F.M., Seong, B.K.A., Kuljanin, M., Leggett, A.L., Mohardt, M.L., Robichaud, A., Conway, A.S., Buckley, D.L., Mancias, J.D., et al. (2020). Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. Nat. Commun. *11*, 4687. https://doi.org/10.1038/s41467-020-18377-w.
31. Martelly, W., Fellows, B., Kang, P., Vashisht, A., Wohlschlegel, J.A., and Sharma, S. (2021). Synergistic roles for human U1 snRNA stem-loops in

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

pre-mRNA splicing. RNA Biol. *18*, 2576–2593. https://doi.org/10.1080/15476286.2021.1932360.

32. Azam, S., Hou, S., Zhu, B., Wang, W., Hao, T., Bu, X., Khan, M., and Lei, H. (2019). Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. RNA Biol. *16*, 1001–1009. https://doi.org/10.1080/15476286.2019.1620061.
33. Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. Mol. Cell *33*, 365–376. https://doi.org/10.1016/j.molcel.2008.12.028.
34. Bresson, S.M., and Conrad, N.K. (2013). The Human Nuclear Poly(A)-Binding Protein Promotes RNA Hyperadenylation and Decay. PLoS Genet. *9*, e1003893. https://doi.org/10.1371/journal.pgen.1003893.
35. Wang, Y., Fan, J., Wang, J., Zhu, Y., Xu, L., Tong, D., and Cheng, H. (2021). ZFC3H1 prevents RNA trafficking into nuclear speckles through condensation. Nucleic Acids Res. *49*, 10630–10643. https://doi.org/10.1093/nar/gkab774.
36. Foucher, A.-E., Touat-Todeschini, L., Juarez-Martinez, A.B., Rakitch, A., Laroussi, H., Karczewski, C., Acajjaoui, S., Soler-López, M., Cusack, S., Mackereth, C.D., et al. (2022). Structural analysis of Red1 as a conserved scaffold of the RNA-targeting MTREC/PAXT complex. Nat. Commun. *13*, 4969. https://doi.org/10.1038/s41467-022-32542-3.
37. Polák, P., Garland, W., Rathore, O., Schmid, M., Salerno-Kochan, A., Jakobsen, L., Gockert, M., Gerlach, P., Silla, T., Andersen, J.S., et al. (2023). Dual agonistic and antagonistic roles of ZC3H18 provide for co-activation of distinct nuclear RNA decay pathways. Cell Rep. *42*, 113325. https://doi.org/10.1016/j.celrep.2023.113325.
38. Fan, J., Wang, Y., Wen, M., Tong, D., Wu, K., Yan, K., Jia, P., Zhu, Y., Liu, Q., Zou, H., et al. (2024). Dual modes of ZFC3H1 confer selectivity in nuclear RNA sorting. Mol. Cell *84*, 4297–4313.e7. https://doi.org/10.1016/j.molcel.2024.09.032.
39. Vielhaber, E., Jacobson, D.P., Ketterling, R.P., Liu, J.Z., and Sommer, S.S. (1993). A mutation in the 3′ untranslated region of the factor IX gene in four families with hemophilia B. Hum. Mol. Genet. *2*, 1309–1310. https://doi.org/10.1093/hmg/2.8.1309.
40. Langemeier, J., Schrom, E.-M., Rabner, A., Radtke, M., Zychlinski, D., Saborowski, A., Bohn, G., Mandel-Gutfreund, Y., Bodem, J., Klein, C., et al. (2012). A complex immunodeficiency is based on U1 snRNP-mediated poly(A) site suppression. EMBO J. *31*, 4035–4044. https://doi.org/10.1038/emboj.2012.252.
41. Bohn, G., Allroth, A., Brandes, G., Thiel, J., Glocker, E., Schäffer, A.A., Rathinam, C., Taub, N., Teis, D., Zeidler, C., et al. (2007). A novel human primary immunodeficiency syndrome caused by deficiency of the endosomal adaptor protein p14. Nat. Med. *13*, 38–45. https://doi.org/10.1038/nm1528.
42. Krooss, S., Werwitzke, S., Kopp, J., Rovai, A., Varnholt, D., Wachs, A.S., Goyenvalle, A., Aarstma-Rus, A., Ott, M., Tiede, A., et al. (2020). Pathological mechanism and antisense oligonucleotide-mediated rescue of a non-coding variant suppressing factor 9 RNA biogenesis leading to hemophilia B. PLoS Genet. *16*, e1008690. https://doi.org/10.1371/journal.pgen.1008690.
43. Wang, J., Huang, D., Zhou, Y., Yao, H., Liu, H., Zhai, S., Wu, C., Zheng, Z., Zhao, K., Wang, Z., et al. (2020). CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. Nucleic Acids Res. *48*, D807–D816. https://doi.org/10.1093/nar/gkz1026.
44. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.
45. Kanai, M., Ulirsch, J.C., Karjalainen, J., Kurki, M., Karczewski, K.J., Fauman, E., Wang, Q.S., Jacobs, H., Aguet, F., Ardlie, K.G., et al. (2021). Insights from complex trait fine-mapping across diverse populations.. Preprint at medRxiv. https://doi.org/10.1101/2021.09.03.21262975.
46. Mitra, S., Malik, R., Wong, W., Rahman, A., Hartemink, A.J., Pritykin, Y., Dey, K.K., and Leslie, C.S. (2024). Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. Nat. Genet. *56*, 627–636. https://doi.org/10.1038/s41588-024-01689-8.
47. Shaw, G., and Kamen, R. (1986). A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell *46*, 659–667. https://doi.org/10.1016/0092-8674(86)90341-7.
48. Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A., and Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature *442*, 45–50. https://doi.org/10.1038/nature04881.
49. Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes Dev. *29*, 63–80. https://doi.org/10.1101/gad.247361.114.
50. Li, Y., Bor, Y.-C., Misawa, Y., Xue, Y., Rekosh, D., and Hammarskjöld, M.-L. (2006). An intron with a constitutive transport element is retained in a Tap messenger RNA. Nature *443*, 234–237. https://doi.org/10.1038/nature05107.
51. Parra, M., Booth, B.W., Weiszmann, R., Yee, B., Yeo, G.W., Brown, J.B., Celniker, S.E., and Conboy, J.G. (2018). An important class of intron retention events in human erythroblasts is regulated by cryptic exons proposed to function as splicing decoys. RNA *24*, 1255–1265. https://doi.org/10.1261/rna.066951.118.
52. Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. *22*, 96–118. https://doi.org/10.1038/s41580-020-00315-9.
53. Ding, F., and Elowitz, M.B. (2019). Constitutive splicing and economies of scale in gene expression. Nat. Struct. Mol. Biol. *26*, 424–432. https://doi.org/10.1038/s41594-019-0226-x.
54. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10. https://doi.org/10.14806/ej.17.1.200.
55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
56. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
57. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.
58. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.
59. Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. *47*, W171–W174. https://doi.org/10.1093/nar/gkz365.
60. Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.T., and Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat. Protoc. *11*, 118–133. https://doi.org/10.1038/nprot.2015.140.
61. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441. https://doi.org/10.1038/s41589-018-0021-8.

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

62. Wang, X., Liu, L., Whisnant, A.W., Hennig, T., Djakovic, L., Haque, N., Bach, C., Sandri-Goldin, R.M., Erhard, F., Friedel, C.C., et al. (2021). Mechanism and consequences of herpes simplex virus 1-mediated regulation of host mRNA alternative polyadenylation. PLoS Genet. *17*, e1009263. https://doi.org/10.1371/journal.pgen.1009263.

63. Taylor, S.C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., and Fenrich, J. (2019). The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. Trends Biotechnol. *37*, 761–774. https://doi.org/10.1016/j.tibtech.2018.12.002.

64. Linder, J., Koplik, S.E., Kundaje, A., and Seelig, G. (2022). Deciphering the impact of genetic variation on human polyadenylation using APARENT2. Genome Biol. *23*, 232. https://doi.org/10.1186/s13059-022-02799-4.

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-FLAG M2	Sigma-Aldrich	Cat#F1804-200UG; RRID: AB_262044
Rabbit polyclonal anti-beta Tubulin	Invitrogen	Cat#PA1-16947; RRID: AB_795659
Mouse monoclonal anti-GAPDH	Santa Cruz	Cat#sc-365062; RRID: AB_10847862
Rabbit polyclonal anti-CCDC131 (ZFC3H1)	Bethyl	Cat#A301-456A; RRID: AB_960958
Rabbit polyclonal anti-MTR4	Abcam	Cat#AB70551; RRID: AB_1270701
Rabbit polyclonal anti-PABPN1	Bethyl	Cat#A303-524A; RRID: AB_10949200
Rabbit polyclonal anti-ZCCHC8	Bethyl	Cat#A301-806A; RRID: AB_1233063
Rabbit polyclonal anti-EXOSC3	Proteintech	Cat#15062-1-AP; RRID: AB_2278183
Mouse monoclonal anti-U1-70K, clone 9C4.1	Millipore	Cat#05-1588; RRID: AB_10805959
Mouse monoclonal anti-U1A	Santa Cruz	Cat#sc-101149; RRID: AB_2193721
Rat monoclonal anti-U1C	Sigma	Cat#SAB4200188-200UL; RRID: AB_10640155
Rabbit polyclonal anti-ARS2	Bethyl	Cat#A304-550A; RRID: AB_2620745
Rabbit polyclonal anti-CPSF100	Bethyl	Cat#A301-581A; RRID: AB_1078861
Rabbit polyclonal anti-CPSF73	Bethyl	Cat#A301-091A; RRID: AB_2084528
Rabbit polyclonal anti-CPSF30	Bethyl	Cat#A301-585A; RRID: AB_1078868
Rabbit monoclonal anti-FLAG	Cell Signaling	Cat#14793S; RRID: AB_2572291
Mouse monoclonal anti-SC-35 (SRRM2)	Abcam	Cat#ab11826; RRID: AB_298608
IRDye 680RD Goat Anti-Rabbit IgG	Licor	Cat#926-68071; RRID: AB_10956166
IRDye 800CW Goat Anti-Mouse IgG	Licor	Cat#926-32210; RRID: AB_621842
Goat Anti-Rabbit IgG HRP-conjugate	Millipore	Cat#12-348; RRID: AB_390191
Goat Anti-Mouse IgG HRP-conjugate	Millipore	Cat#12-349; RRID: AB_390192
Rabbit Anti-Rat IgG HRP-conjugate	Invitrogen	Cat#PA1-28573; RRID: AB_10980086
Chemicals, peptides, and recombinant proteins		
DMEM	Gibco	Cat#11995065
Fetal Bovine Serum	Sigma	Cat#F0926-100ML
Lipofectamine 3000	Invitrogen	Cat#L3000015
In-Fusion Snap Assembly Master Mix	Takara	Cat#638949
Blasticidin	Invivogen	Cat#ant-bl-05
Puromycin	Invivogen	Cat#ant-pr-1
dTAG^{V}-1	Tocris	Cat#6914
Polybrene	Sigma	Cat#TR-1003
Trizol	Invitrogen	Cat#15596026
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen	Cat#13778100
NEBNext Magnesium RNA fragmentation module	NEB	Cat#E6150S
Phusion High-Fidelity PCR Master Mix with HF Buffer	NEB	Cat#M0531L
Lonza SF Cell Line 4D-Nucleofector X Kit	Fisher Scientific	Cat#NC0281111
RQ1 RNase-Free DNase	Promega	Cat#M6101
Superscript III Reverse Transcriptase	ThermoFisher Scientific	Cat#18080044
Random hexamers	ThermoFisher Scientific	Cat#FERSO142

(*Continued on next page*)

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PowerUp™ SYBR™ Green Master Mix	Applied Biosystems	Cat#A25742
Poly-D-Lysine	Gibco	Cat#A3890401
Paraformaldehyde	ThermoFisher Scientific	Cat#AA47377-9L
2X SSC	Invitrogen	Cat#15557-036
Formamide (Deionized)	Ambion	Cat#AM9342
Tween-20	ThermoFisher	Cat#AAJ20605AP
Dextran Sulfate	Fisher	Cat#BP1585-100
E.coli tRNA	Roche	Cat#10 109 541 001
Vanadyl ribonucleoside complex	NEB	Cat#S1402S
DAPI	Invitrogen	Cat#D1306
Fluoromount G	Invitrogen	Cat#00495802
T7 RNA Polymerase	NEB	Cat#M0251S
Amersham MicroSpin G-50 Columns	Cytiva	Cat#27533001
[α-32P]-UTP	PerkinElmer	Cat#BLU007H250UC
E.coli Poly(A) Polymerase	NEB	Cat#M0276S
Yeast tRNA	Invitrogen	Cat#AM7119
Amylose Resin	NEB	Cat#E8021L
Halt™ Protease Inhibitor Cocktail	ThermoFisher Scientific	Cat#78429
ANTI-FLAG M2 Affinity Gel	Millipore	Cat#A2220
FLAG peptide	Sigma-Aldrich	Cat#F3290
RNase A	ThermoFisher Scientific	Cat#EN0531
0.45 μm nitrocellulose membrane	Bio-Rad	Cat#1620115
Radiance Q Chemiluminescent Substrate	Azure Biosystems	Cat#10147-296
Critical commercial assays		
MycoAlert Mycoplasma Detection Kit	Lonza	Cat#LT07-318
qPCR Lentivirus Titer Kit	Applied Biological Materials	Cat#LV900
4-20% pre-cast SDS-PAGE gels	BioRad	Cat#4568096
Deposited data		
PAS-seq	This study	GEO: GSE267857
Experimental models: Cell lines		
HEK293T	ATCC	Cat#CRL-3216
EXOSC3 dTAG HEK293T	This study	N/A
ZFC3H1 dTAG HEK293T	This study	N/A
Oligonucleotides		
RT-qPCR primers, shRNA sequences, sgRNA sequence, siRNA sequences (see Table S2)	This paper	N/A
eGFP FISH probes	Ding and Elowitz[53]	N/A
MISSION siRNA universal control	MilliporeSigma	SIC001
Recombinant DNA		
Plasmid: pLKO: scramble shRNA	This study	N/A
Plasmid: pLKO: ZFC3H1 shRNA 1	This study	N/A
Plasmid: pLKO: ZFC3H1 shRNA 2	This study	N/A
Plasmid: pLKO: MTR4 shRNA	This study	N/A
Plasmid: pLKO: PABPN1 shRNA	This study	N/A
Plasmid: pCRIS-PITCHv2-dTAG-BSD (EXOSC3)	This study	N/A
Plasmid: pCRIS-PITCHv2-BSD-dTAG (ZFC3H1)	This study	N/A

(*Continued on next page*)

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pX330A-S-EXOSC3-sgRNA	This study	N/A
Plasmid: pX330A-S-ZFC3H1-sgRNA	This study	N/A
Plasmid: pGL3-promoter	Promega	Cat#E1751
Plasmid: pRL-TK	Promega	Cat#E2241
Plasmid: pCDNA5: FLAG-eGFP-5′ss WT-*bGH* PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss Mut-*bGH* PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss WT-MALAT 3′	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss Mut-MALAT 3′	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss WT-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss Mut-*L3* PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss WT-Histone 3′	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss Mut-Histone 3′	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss WT in coding sequence-*bGH* PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss Mut in coding sequence-*bGH* PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 10.86-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 10.28-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 9.88-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 8.84-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 8.66-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 6.12-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength 3.05-L3 PAS	This study	N/A
Plasmid: pCDNA5: FLAG-eGFP-5′ss strength (5′ss Mut)-*L3* PAS	This study	N/A
Plasmid: pBluescript: 3XMS2-5′ss WT-L3 PAS WT	This study	N/A
Plasmid: pBluescript: 3XMS2-5′ss Mut-*L3* PAS WT	This study	N/A
Plasmid: pBluescript: 3XMS2-5′ss WT-L3 PAS Mut	This study	N/A
Plasmid: pBluescript: 3XMS2-5′ss Mut-*L3* PAS Mut	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 Full length	This study	N/A
Plasmid: pCDNA3: Strep-ZFC3H1 Full length	This study	N/A

(*Continued on next page*)

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pCDNA3: FLAG-ZFC3H1 N-term c-myc NLS	This study	N/A
Plasmid: pCDNA3: Strep-ZFC3H1 N-term c-myc NLS	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 C-term c-myc NLS	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 N-term SLiM Mut c-myc NLS	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 50-599 c-myc NLS	This study	N/A
Plasmid: pCDNA3: Strep-ZFC3H1 50-599 c-myc NLS	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 600-990 c-myc NLS	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 CC4 Mut	This study	N/A
Plasmid: pCDNA3: Strep-ZFC3H1 CC4 Mut	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 ΔCC4	This study	N/A
Plasmid: pCDNA3: Strep-ZFC3H1 ΔCC4	This study	N/A
Plasmid: pCDNA3: FLAG-ZFC3H1 Δ813-932	This study	N/A
Plasmid: pCDNA3: Strep-ZFC3H1 Δ813-932	This study	N/A
Plasmid: pCDNA3: FLAG-U1A	This study	N/A
Plasmid: pCDNA3: FLAG-CPSF30	This study	N/A
Plasmid: pNS6: U1 snRNA targeting KPNB1 5′ ss Mut (non-targeting control)	This study	N/A
Plasmid: pNS6: U1 snRNA targeting 5′ss Mut-*bGH* reporter	This study	N/A
Software and algorithms		
Cutadapt v2.10	Martin[54]	N/A
STAR v2.7.3a	Dobin et al.[55]	N/A
Bedtools v2.29.2	Quinlan and Hall[56]	N/A
edgeR v3.40.1	Robinson et al.[57]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
ImageJ	Schneider et al.[58]	https://imagej.net/ij/
GraphPad Prism	GraphPad Software	https://www.graphpad.com
CHOPCHOP	Labun et al.[59]	https://chopchop.cbu.uib.no/
ImageQuant TL	ImageQuant™ TL analysis software	https://info.cytivalifesciences.com/image-analysis-software.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell Lines

Wild-type and edited HEK293T cell lines were maintained in DMEM (Gibco) supplemented with 10% FBS (Sigma). Cells were grown at 37C with 5% CO_2. Cell lines were routinely monitored for mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza).

METHOD DETAILS

Generation of FLAG-dTAG Edited HEK293T Cell Lines

The C-terminal dTAG-FLAG knockin of EXOSC3 and the N-terminal FLAG-dTAG knockin of ZFC3H1 in HEK293T cells were achieved using CRISPR/Cas9 and the microhomology end joining (MMEJ) approach as described in,[60] except that we used Lipofectamine 3000 to perform the transfections and used blasticidin (10 μg/mL) during antibiotic selection. To construct the donor vector, we

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

modified pCRIS-PITCHv2-dTAG-BSD (BRD4) and pCRIS-PITCHv2-BSD-dTAG (BRD4) to replace the HA tag with a 3X-FLAG tag and then replaced the microhomology sequences for EXOSC3 or ZFC3H1 using In Fusion cloning (Takara). pCRIS-PITCHv2-dTAG-BSD (BRD4) and pCRIS-PITCHv2-BSD-dTAG (BRD4) were gifts from James Bradner & Behnam Nabet (Addgene plasmid # 91795; https://www.addgene.org/91795/ and Addgene plasmid # 91792; https://www.addgene.org/91792/).[61] We used the online tool CHOPCHOP[59] to select a specific and efficient sgRNA. We screened and validated single colonies using genomic DNA PCR followed by Sanger Sequencing where possible and by western blotting for all cell lines. Homozygous edited colonies were selected and used in experiments. For ZFC3H1, only the largest isoform exhibited dTAG-dependent degradation.

RNAi

Lentiviral Knockdown

Lentiviral particles were prepared by co-transfecting HEK293T cells with the following plasmids: pLKO: shRNA vector, PMD2.G, and psPAX2. The virus-containing cell culture media was collected 24 and 48 hours after transfection and passed through a 45 μM filter. The titer was measured using a qPCR-based kit (Applied Biological Materials). HEK293T cells were transduced at an MOI of 10 with lentivirus expressing one or two different shRNAs targeting MTR4, ZFC3H1, PABPN1 or a control scramble sequence. Polybrene (Sigma) was included during transduction (8 μg/mL). 24 hours after transduction, puromycin was added to the media at a concentration of 1.25 μg/mL. 4 days after transduction, cells were harvested in Trizol for RNA purification or directly lysed in 1X SDS loading dye for western blotting. pMD2.G was a gift from Didier Trono (Addgene plasmid # 12259). psPAX2 was a gift from Didier Trono (Addgene plasmid # 12260).

siRNA Knockdown

An EXOSC3-targeting siRNA (IDT) or a pool of control siRNAs (Sigma) was transfected into cells using Lipofectamine RNAiMAX at a final siRNA concentration of 20 nM. 48 hours later, the transfection was repeated. 24 hours later, cells were harvested in Trizol for RNA purification or directly lysed in 1X SDS loading dye for western blotting.

PAS-seq

Library Preparation

PAS-seq libraries were prepared as previously described[23] with the following minimal modifications. Purified mRNA was fragmented at 94°C for 3 minutes using the NEBNext Magnesium RNA fragmentation module. In addition, the libraries were resolved on a 2.5% agarose gel and the region between 200-300 basepairs was purified by gel extraction and sequenced on the Illumina NovaSeq6000 platform. For all samples, a small amount of mouse RNA was spiked in but was not used in data processing or during differential gene expression calculations. For the spike-ins, an extra control or treatment well was prepared and the cells in this extra well were counted and recorded. 1 μg total RNA from each sample was mixed with up to 10% by mass (100 ng) mouse total RNA (purified from mouse brain tissue) prior to library preparation. The spike-in RNA mass was calculated relative to the measured cell count for each sample. For example, if there were 1 x 10^6 control cells and 0.8 x 10^6 treated cells, then 100 ng mouse spike-in RNA would be added to 1 μg of control total RNA and 80 ng mouse spike-in RNA would be added to 1 μg of treated total RNA. As such, the level of spike-in RNA could be used to normalize sequencing reads to the original cell number.

PAS-seq Data Processing

PAS-seq data processing was conducted as previously described with few modifications.[62] Briefly, we used Cutadapt (v2.10)[54] to produce trimmed PAS-seq reads by: 1) trimming the 6 nucleotide linker sequence, 2) trimming the poly(A) tail sequence, and 3) removing any untrimmed reads. Trimmed reads were then mapped to a concatenated hg19 and mm9 genome using STAR (v2.7.3a)[55] with the –alignEndsType EndToEnd parameter. The mapped reads were converted from a bam file to a bed file using bedtools (v2.29.2).[56] Using a custom python script, reads were removed as potential internal priming events if they mapped to a genomic region where 6 consecutive A's or 7 A's out of 10 nucleotides were observed in the 10 nucleotides downstream of the read. The remaining reads were converted to a bam file and used to generate bigwig files using bedtools. The 3' ends of the reads were extracted from the bed files using bedtools flank and the initial read counts for each poly(A) site were calculated using bedtools coverage with a master file of all annotated poly(A) sites.

Differential Isoform Expression Analysis

To detect differentially expressed transcripts using PAS-Seq data, we treated counts mapping to each poly(A) site region as a "gene" and analyzed differential gene expression using the edgeR[57] (v3.40.1) topTags function. We considered a transcript to be differentially expressed if it displayed an FDR $\leq$ 0.05 and a $\log_2$FC > 1 or $\log_2$FC < -1, unless indicated otherwise.

Molecular Cloning

eGFP-5′ ss Reporters

The eGFP-5′ ss reporters contained the FLAG-eGFP coding sequence followed by a 3′ UTR in the pCDNA5 expression vector. In the 5′ ss-containing reporters, the 3′ UTR included a 28 nucleotide sequence from the gene *NXF1* that was previously shown to contain one strong and one weaker 5′ ss.[26] In the 5′ ss mutant reporters, both 5′ splice sites were mutated by PCR mutagenesis using Phusion. The polyadenylated reporters contained either the bovine growth hormone (bGH) poly(A) site or the adenovirus major late (L3)

poly(A) site downstream of the 5′ ss. The non-polyadenylated reporters contained either: 1) a 164-nucleotide sequence from the 3′ end of *MALAT1* or, 2) a 129-nucleotide region including part of the *H2AC18* 3′UTR and 56 nucleotides of downstream sequence to elicit histone 3′ end processing.

eGFP-5′ ss strength Reporters

The eGFP-5′ ss strength reporters were produced by cloning a single 5′ ss sequence followed by the L3 poly(A) site sequence downstream of the FLAG-eGFP coding sequence. The single 5′ ss was mutated to produce reporters with varying 5′ ss strengths. The utilized 5′ss sequences followed by their MaxEnt score, where applicable, were: CAG|GTAAGT (10.86); CAG|GTAGGT (10.28); CAG|GTAAGC (9.88); TGG|GTAAGC (8.84); CAG|GTAACC (8.66); AAG|GTTAGA (6.12); CTG|GTCTGT (3.05); AGA|AGCCATA (Mut 5′ ss).

RNA Pulldown Constructs

The RNA pulldown constructs contained 3 copies of the MS2 hairpin sequence followed by the 5' ss sequence 5'-CAG|GTAAGT-3′ and the L3 poly(A) site in the pBluescript vector. The distance between the end of the 5′ ss and the L3 AAUAAA poly(A) signal was 68 nucleotides. In the 5′ ss mutant constructs, the 5′ ss was mutated to 5′-AGA|AGCCAT-3′. In the poly(A) site mutant constructs, the L3 poly(A) signal was mutated from 5′-AAUAAA-3′ to 5′-AA**G**AAA-3′.

FLAG-ZFC3H1 Expression Constructs

The FLAG-ZFC3H1 expression constructs were cloned into pCDNA3. The truncations and mutations were produced by PCR mutagenesis using Phusion. The ZFC3H1 SLiM mutant contained the following mutations: 18-EEGELEDGEI-27 to 18-E**A**G**A**LE**A**G**A**I-27. All constructs contained an N-terminal FLAG tag, and the c-myc nuclear localization signal was added to all expression constructs except the full length.

eGFP-F9 Reporters

For the eGFP-F9 reporters, the entire *F9* 3′UTR sequence and 597 nucleotides of downstream genomic sequence was cloned downstream of FLAG-eGFP in pCDNA5. The A>G patient mutation 1157 basepairs downstream of the stop codon was introduced by PCR mutagenesis using Phusion.

p14 Reporters

For the *p14* reporters, the *p14* coding sequence, terminal intron, 3′ UTR, and 317 nucleotides of downstream genomic sequence was cloned into the pGL3: promoter backbone by replacing firefly luciferase. The C>A patient mutation 23 basepairs downstream of the stop codon was introduced by PCR mutagenesis.

Reporter Assays

For baseline eGFP-5′ss reporter mRNA and protein level measurements, eGFP-5′ ss reporters were co-transfected with the renilla luciferase expressing plasmid pRL-TK at a 10:1 ratio by mass into HEK293T cells using Lipofectamine 3000 (i.e., 250 ng pCDNA5 and 25 ng pRL-TK). 48 hours after transfection, cells were harvested in Trizol for RNA purification or directly lysed in 1X SDS loading buffer for western blotting.

For baseline eGFP-F9 reporter mRNA level measurements, eGFP-F9 reporters were co-transfected with the firefly luciferase expressing plasmid pGL3: promoter at a 10:1 ratio by mass into HEK293T cells. 24 hours after transfection, cells were harvested in Trizol for RNA purification.

To measure the mRNA level of reporters after dTAG-induced depletion of EXOSC3 or ZFC3H1, or U1 inhibition by U1 AMO, eGFP-5′ ss or eGFP-F9 reporters were co-transfected with the firefly luciferase expressing plasmid pGL3: promoter at a 10:1 ratio by mass into HEK293T cells. To deplete EXOSC3 or ZFC3H1, 16 hours after transfection, dTAG^{V}-1 or an equivalent volume of DMSO was added to cell culture media at a final concentration of 500 nM, and cells were harvested 8 hours later.[30] To inhibit U1 snRNP using U1 AMO, 16 hours after transfection 25 μM U1 AMO or Control AMO was nucleofected into transfected cells and cells were harvested 8 hours later. Cells were harvested in Trizol for RNA purification or directly lysed in 1X SDS loading buffer for western blotting.

p14 Reporters

For baseline *p14* reporter mRNA level measurements, *p14* reporters were co-transfected with the firefly luciferase expressing plasmid pGL3: promoter at a 10:1 ratio by mass into HEK293T cells. 24 hours after transfection, cells were harvested in Trizol for RNA purification. RT-qPCR was performed as described below using cDNA reverse transcribed with oligo d(T)$_{20}$ normalized to firefly luciferase expression to monitor the mRNA levels.

To measure the mRNA level of *p14* reporters after dTAG-induced depletion of EXOSC3, *p14* reporters were co-transfected with the firefly luciferase expressing plasmid pGL3: promoter at a 10:1 ratio by mass into HEK293T cells. To deplete EXOSC3, 16 hours after transfection dTAG^{V}-1 or an equivalent volume of DMSO was added to cell culture media at a final concentration of 500 nM and cells were harvested 8 hours later.[30] Cells were harvested in Trizol for RNA purification or directly lysed in 1X SDS loading buffer for western blotting. RT-qPCR was performed as described below using cDNA reverse transcribed with oligo d(T)$_{20}$ normalized to firefly luciferase expression to monitor the mRNA levels.

5′ ss strength reporters

For eGFP-*L3* with varying 5′ ss strength reporter mRNA level measurements, eGFP-5′ ss strength reporters were co-transfected with the firefly luciferase expressing plasmid pGL3:promoter at a 10:1 ratio by mass into HEK293T cells using Lipofectamine 3000 (i.e., 250 ng pCDNA5 and 25 ng pGL3:promoter). 48 hours after transfection, cells were harvested in Trizol for RNA purification.

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

5′ ss within coding sequence reporters

For eGFP-*bGH* with 5′ ss in the coding sequence reporter mRNA level measurements, eGFP-5′ ss coding sequence reporters were co-transfected with the firefly luciferase expressing plasmid pGL3:promoter at a 10:1 ratio by mass into HEK293T cells using Lipofectamine 3000 (i.e. 250 ng pCDNA5 and 25 ng pGL3:promoter). 48 hours after transfection, cells were harvested in Trizol for RNA purification.

To measure the mRNA level of eGFP-*bGH* reporters with a 5′ ss in the coding sequence after dTAG-induced depletion of EXOSC3, the reporters were co-transfected with the firefly luciferase expressing plasmid pGL3: promoter at a 10:1 ratio by mass into HEK293T cells. To deplete EXOSC3, 16 hours after transfection dTAG^{V}-1 or an equivalent volume of DMSO was added to cell culture media at a final concentration of 500 nM and cells were harvested in Trizol 8 hours later.

U1 snRNA Targeting of 5′ ss mutant reporter

To direct U1 snRNP to bind the 5′ ss mutant *bGH* poly(A) site reporter, HEK293T cells were co-transfected in a 12-well plate with 250 ng eGFP-5′ss mutant *bGH* poly(A) site reporter, 250 ng pNS6: U1 snRNA complementary to the 5′ ss mutant or a control sequence, and 25 ng pGL3: promoter as a co-transfection control. 48 hours after transfection, cells were harvested in Trizol for RNA purification. RT-qPCR was performed as described above using cDNA reverse transcribed with oligo d(T)$_{20}$ normalized to firefly luciferase expression to monitor the mRNA levels.

Inhibition of U1 snRNP by AMO

To inhibit U1 snRNP, 25 μM control or U1 AMO (GeneTools, LLC) was delivered into 1x10^{6} cells by nucleofection using the Lonza SF Cell Line 4D-Nucleofector™ X Kit using the DH-135 program. Cells were harvested 8 hours later in Trizol for RNA purification.

RT-qPCR

Total RNA was purified following the Trizol manufacturer instructions. 500 ng-1 μg total RNA was DNase-treated using RQ1 DNase (Promega) followed by addition of RQ1 DNase stop solution and heat inactivation. For experiments using eGFP-5′ ss reporters, the DNase-treated RNA was reverse transcribed using Superscript III (ThermoFisher) and random hexamers or oligo d(T)$_{20}$ where indicated. For experiments using eGFP-F9 and *p14* reporters, the DNase-treated RNA was reverse transcribed using Superscript III and oligo d(T)$_{20}$. The resulting cDNA was diluted and used for qPCR using the PowerUp™ SYBR™ Green Master Mix (Applied Biosystems). All qPCR analyses used the co-transfected plasmid to normalize reporter mRNA measurements (renilla luciferase from pRL:TK or firefly luciferase from pGL3: promoter). The delta-delta Ct method was used to analyze qPCR data as described previously.[63] Normalized expression data was plotted on the y-axis and SEM was used to plot error bars.

RNA FISH and Immunofluorescence

To monitor RNA localization of eGFP-5′ ss reporters, we seeded 56,250 HEK293T cells per well (225,000 cells/mL) of 8-well glass bottom chamber slides (ibidi) coated with 0.1 mg/mL Poly-D-Lysine (Gibco). The next morning, we transfected 125 ng of each eGFP-5′ ss reporter using Lipofectamine 3000. 12 hours later, we fixed cells with 4% paraformaldehyde for 10 minutes. Cells were then washed with 1X PBS and permeabilized with 70% ethanol at -20C overnight. RNA-FISH probes targeting the eGFP coding sequence were synthesized as described previously.[53] To hybridize probes to the reporter RNAs, cells were washed once with wash buffer (2X SSC (Invitrogen), 20% deionized formamide (Ambion), and 0.1% Tween-20 (ThermoFisher)) and then incubated overnight at 30C with 125 ng of RNA-FISH probes in 100 μL hybridization buffer (0.1 g/mL dextran sulfate (Fisher), 0.1 mg/mL E.coli tRNA (Roche), 2 mM Vanadyl ribonucleoside complex (NEB), 2X SSC, 20% deionized formamide, 0.1 % Tween-20). The next day, cells were washed twice with 30C wash buffer and then incubated for 30 minutes in wash buffer. Cells were post-fixed with 4% paraformaldehyde for 10 minutes and incubated with immunofluorescence primary antibodies for one hour at room temperature in PBS-Tween 20 (PBST). Cells were washed once with PBST and then incubated with immunofluorescence secondary antibodies for 30 minutes at room temperature in PBST. Cells were washed with PBST and stained for DAPI (Invitrogen) using a 1:20,000 dilution for 5 minutes. Finally, cells were mounted with Fluoromount G (Invitrogen) before imaging. Imaging was performed using the Leica DMi8 THUNDER. Ten to twelve images per reporter were taken. The nuclear versus cytoplasmic RNA FISH signal was quantified using ImageJ.[58] All transfected cells were manually selected and the total nuclear FISH intensity was divided by the total cytoplasmic FISH intensity per image.

In vitro Transcription

For RNA pulldowns and polyadenylation assays, 3XMS2-RNA was *in vitro* transcribed using T7 RNA Polymerase (NEB) and free nucleotides were removed using a column (Cytiva). For radiolabeled RNA substrates, [α-^{32}P]-UTP (PerkinElmer) was included in the transcription reaction. The RNA was purified by phenol/chloroform extraction followed by ethanol precipitation. For radiolabeled RNAs only, the RNA was further subjected to Urea-PAGE purification and ethanol precipitated. To produce the pre-polyadenylated substrates, the DNA template for transcription was prepared by PCR linearizing the pBluescript constructs at the cleavage site using Phusion and used for *in vitro* transcription. The resulting "pre-cleaved" RNA was polyadenylated by *E. coli* Poly(A) Polymerase (NEB) and passed through a column (Cytiva) to remove free nucleotides and purified by phenol/chloroform extraction followed by ethanol precipitation.

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

In vitro Polyadenylation Assay

1 pmol (20 cps) radiolabeled RNA was mixed with 1 μL 0.1 M ATP, 2 μL 1 M creatine phosphate, 1 μL 10 μg/μL yeast tRNA, 4.4 μL HeLa nuclear extract, and H_2O to a final volume of 10 μL. The reaction was incubated at 30℃ for 20 minutes or 150 minutes as indicated. The proteins in the reaction were then digested using Proteinase K and the RNA was purified by phenol/chloroform extraction followed by ethanol precipitation. The RNA was then resolved on an 8% Urea-PAGE gel. The gel was transferred to filter paper, vacuum dried, and used to expose a phosphor screen overnight. The phosphor screen was imaged using a GE Amersham™ Typhoon™.

MS2-MBP RNA Pulldown and Western Blotting

11.25 pmol RNA and 112.5 pmol of MBP-MS2 fusion protein was brought to 50 μL with Buffer D-100 (20 mM HEPES-NaOH pH 7.9, 100 mM NaCl, 1 mM MgCl2, 0.2 mM EDTA) and incubated on ice for 30 minutes. The following was then added: 7.5 μL 0.1 M ATP, 15 μL 1 M creatine phosphate, 7.5 μL 10 μg/μL yeast tRNA, 300 μL HeLa nuclear extract, and H_2O to a final volume of 750 μL. The reaction was incubated at 30°C for 20 minutes. The reaction was then chilled on ice, centrifuged for 1 minute at 14,000 x rpm (>16,000 x g) at 4℃. The supernatant was mixed with pre-washed amylose beads (55 μL slurry) and rotated for 1 hour at 4℃. The beads were washed three times for 10 minutes per wash with 1 mL wash buffer (20 mM HEPES-NaOH pH 7.9, 100 mM KCl, 4% glycerol, 1 mM DTT) supplemented with detergent (0.05% IGEPAL CA-630). The beads were washed once more for 10 minutes per wash with 1 mL wash buffer without detergent. Bound proteins were eluted twice using 125 μL of wash buffer without detergent supplemented with 20 mM maltose for 10 minutes per elution. The eluted proteins were acetone precipitated overnight at -20°C. The next day, samples were centrifuged for 20 minutes at 14,000 x rpm, pellets were resuspended in 1X SDS loading buffer, heated to 95℃ for 7 minutes, and subjected to standard western blotting procedures. All pulldown protein samples were resolved on 4-20% pre-cast SDS-PAGE gels (BioRad) and subjected to a "standard" eBlot transfer.

MS2-MBP RNA Pulldown and Glycerol Gradient Sedimentation

The RNA pulldowns were performed the same as above through elution with maltose except that all components were scaled up to a 1 mL reaction. The eluted RNA-protein complexes were then loaded directly onto a freshly prepared 10-30% glycerol gradient and centrifuged at 30,000 x rpm for 15 hours at 4℃. The next morning, 500 μL fractions were manually prepared and acetone precipitated overnight at -20℃. The next day, samples were centrifuged for 20 minutes at 14,000 x rpm, pellets were resuspended in 1X SDS loading buffer, heated to 95℃ for 7 minutes, and subjected to standard western blotting procedures. All pulldown protein samples were resolved on 4-20% pre-cast SDS-PAGE gels (BioRad) and subjected to a "standard" eBlot transfer.

FLAG-Immunoprecipitation

$2.2x10^6$ HEK293T cells were seeded per 10-cm dish and transfected the next day. For the ZFC3H1 IPs, the transfected DNA amount was titrated in an effort to achieve approximately equal levels of each construct. Accordingly, we transfected the following amounts: 8 μg empty vector, 8 μg full-length ZFC3H1, 10 μg N-term, 24 μg C-term, 8 μg N-term SLiM Mut, 8 μg 50-599, 16 μg 600-990. For the U1A FLAG-IPs, we transfected 8 μg FLAG-U1A, or 4 μg FLAG-U1A and 4 μg Strep-tagged ZFC3H1 N-term or ZFC3H1 50-599. For the CPSF30 FLAG-IPs, we transfected 8 μg FLAG-CPSF30, or 4 μg FLAG-CPSF30 and 4 μg Strep-tagged ZFC3H1 N-term or ZFC3H1 50-599. For the ZFC3H1 self-interactions FLAG-IPs, we transfected 4 μg each of FLAG-tagged and Strep-tagged ZFC3H1 Full-length, CC4 Mut, ΔCC4, or Δ813-932. Two days after transfection, cells were scraped into 5 mL cold 1X PBS, centrifuged at 100 x g for 3 mins, and the cell pellets were flash frozen on dry ice and stored at -80℃ overnight. The next day, cells were resuspended in 1 mL cold lysis buffer per plate (20 mM HEPES-NaOH pH 7.9, 150 mM NaCl, 2 mM EDTA, 0.1% Triton X-100, 1 mM PMSF added fresh, 1X Thermo Scientific Halt™ Protease Inhibitor Cocktail added fresh). While on ice, each sample was sonicated with a microtip at amplitude 1, for 3 seconds on, 10 seconds off, repeated 6 times total. The lysate was then cleared by centrifugation at 14,000 x rpm for 30 minutes at 4℃ and an aliquot was removed for input. During this time, 40 μL of Anti-FLAG M2 agarose beads slurry (Sigma) was washed 3 times with 1 mL lysis buffer per wash. For + RNase A samples, 2 μL RNase A (20 μg RNase A/IP) was added to the supernatant and mixed by pipetting. The supernatant was then added to the pre-washed beads and rotated at 4℃ for 2 hours. The beads were washed 4 times with 1 mL lysis buffer for 10 minutes per wash. The bound proteins were then eluted three times with 100 μL lysis buffer supplemented with 3XFLAG peptide, 10 minutes per elution. The eluted proteins were acetone precipitated overnight at -20℃. The next day, the samples were centrifuged at 14,000 x rpm for 20 minutes and the pellets were resuspended in 1X SDS loading buffer and heated to 95℃ for 7 minutes prior to western blotting. For anti-FLAG western blot from all ZFC3H1 FLAG-IPs, the samples were resolved on a 4-20% pre-cast gel (BioRad) and subjected to a "long" eBlot transfer. For all other proteins, the samples were resolved on a hand-poured 10% SDS-PAGE gel or a 4-20% pre-cast gel and subjected to a "standard" eBlot transfer.

SNP Analyses

Selection of SNPs

To analyze all potentially causal variants, all fine-mapped, single-nucleotide variants within 3′ UTRs with a PIP score of at least 0.02 were extracted from the CAUSALdb and UK Biobank databases. SNPs that overlapped with 3′ UTRs were identified and downloaded as described in a previous study.[64]

Please cite this article in press as: Soles et al., A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance, Molecular Cell (2025), https://doi.org/10.1016/j.molcel.2025.03.010

Prediction of Nuclear RNA Degradation Code Activating and Inactivating SNPs

To compute the 5′ ss strength of the reference sequence and the variant sequence, we extracted 25 nucleotides of flanking genomic sequence upstream and downstream of the variant position using bedtools slop and bedtools getfasta. We then combined the upstream and downstream genomic sequence with the reference nucleotide or the variant nucleotide in the middle position. We then extracted 9 nucleotide "5′ ss" sequences, wherein the reference or variant nucleotide was located in every possible position (9 permutations for reference and variant). We calculated the 5′ ss strength (MaxEnt) for all sequences.[25] Finally, we selected the maximum 5′ ss strength score calculated for each reference and variant and calculated the difference between these values. We identified potentially disease-causing, nuclear RNA degradation code-activating variants by filtering for variants with a PIP score of at least 0.2 that increased the reference maximum 5′ ss strength by at least 3 MaxEnt units to 6.02 or higher (1 IQR below the median 5′ ss strength: 8.66). We identified potentially disease-causing, nuclear RNA degradation code-inactivating variants by filtering for variants with a PIP score of at least 0.2 that decreased the reference maximum 5′ ss strength of at least 6.02 (1 IQR below the median 5′ ss strength: 8.66) by at least 3 MaxEnt units.

QUANTIFICATION AND STATISTICAL ANALYSIS

General Statistical Analysis

All statistics were performed using the GraphPad Prism 10.2.1 software. Statistical tests used are listed in the figure legends. Where applicable, unlabeled or ns indicates p-value > 0.05, * indicates p-value $\leq$ 0.05, ** indicates p-value $\leq$ 0.01, *** indicates p-value $\leq$ 0.001, and **** indicates p-value $\leq$ 0.0001. For all boxplots, the Tukey method was used for error bars. All unpaired t-tests and Mann-Whitney tests were two-tailed. For all bar graphs, data are presented as mean ± SEM or SD, as indicated.

Reporter Assay Protein Level Quantification

To quantify differences in eGFP-5′ ss reporter protein levels (*bGH* or *MALAT1* 3′ end), protein lysates were subjected to western blotting using IR dye-conjugated secondary antibodies (Licor). The membranes were imaged using a GE Amersham™ Typhoon™ and quantified using ImageQuant TL software.

In vitro Polyadenylation Assay Quantification

The average proportion of polyadenylated RNA detected in each polyadenylation reaction condition after 20 minutes incubation was quantified using the ImageQuant TL software (pre-mRNA: 20.3% polyadenylated at 20 minutes, pre-polyadenylated: 100% polyadenylated at 20 minutes, n = 4 biological replicates).

RNA Pulldown Assay Protein Quantification

To quantify the eluted protein levels from the RNA pulldown assays, protein lysates were subjected to western blotting using IR dye-conjugated secondary antibodies (Licor). The membranes were imaged using a GE Amersham™ Typhoon™ and quantified using ImageQuant TL software. The normalized protein levels were calculated by dividing the indicated protein level by the MCP-MBP protein level. The level of ZFC3H1 on polyadenylated RNA was calculated by dividing the normalized ZFC3H1 protein level detected in the pre-mRNA or pre-polyadenylated RNA pulldowns by the average proportion of polyadenylated RNA detected in each polyadenylation reaction condition after 20 minutes incubation (pre-mRNA: 20.3% polyadenylated at 20 minutes, pre-polyadenylated: 100% polyadenylated at 20 minutes).