Activation and inhibition of nonsense-mediated mRNA decay control the abundance of alternative polyadenylation products

Aparna Kishor¹, Sarah E. Fritz¹, Nazmul Haque¹, Zhiyun Ge¹, Ilker Tunc², Wenjing Yang³, Jun Zhu³ and J. Robert Hogg*¹

¹Biochemistry and Biophysics Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA, ²Bioinformatics and Computational Biology Laboratory, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA and ³Systems Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received February 20, 2020; Revised May 26, 2020; Editorial Decision May 27, 2020; Accepted May 29, 2020

ABSTRACT

Alternative polyadenylation (APA) produces transcript 3′ untranslated regions (3′UTRs) with distinct sequences, lengths, stabilities and functions. We show here that APA products include a class of cryptic nonsense-mediated mRNA decay (NMD) substrates with extended 3′UTRs that gene- or transcript-level analyses of NMD often fail to detect. Transcriptome-wide, the core NMD factor UPF1 preferentially recognizes long 3′UTR products of APA, leading to their systematic downregulation. Counteracting this mechanism, the multifunctional RNA-binding protein PTBP1 regulates the balance of short and long 3′UTR isoforms by inhibiting NMD, in addition to its previously described modulation of co-transcriptional polyadenylation (polyA) site choice. Further, we find that many transcripts with altered APA isoform abundance across multiple tumor types are controlled by NMD. Together, our findings reveal a widespread role for NMD in shaping the outcomes of APA.

INTRODUCTION

The 3′ untranslated regions (3′UTRs) of messenger RNAs (mRNAs) coordinate post-transcriptional regulatory mechanisms. The sequence, structure and length of 3′UTRs determine their interactions with trans-acting factors, thereby controlling transcript functions, localization and stability (1). Correspondingly, cells exploit alternative pre-mRNA cleavage and polyadenylation (APA) to generate transcript isoforms with functionally distinct 3′UTRs. A major class of APA events involves differential selection among multiple potential polyadenylation (polyA) sites within a single last exon (frequently termed tandem APA) (2). Tandem APA typically produces mRNAs with identical coding sequences but 3′UTRs of different sequence and length.

Changes in the cellular abundance of core cleavage and polyadenylation factors, often observed in cellular differentiation and transformation, result in systematic expression of mRNAs with 3′UTRs that are either short (increased use of stop codon-proximal polyA sites) or long (increased use of distal polyA sites) (3–12). APA is also modulated by RNA binding proteins (RBPs) that prevent recognition of potential polyadenylation sites or enhance recruitment of cleavage and polyadenylation factors (13–17). While relative APA isoform abundance has primarily been studied with respect to the regulation of polyA site choice in the nucleus, it may also be controlled at the level of mRNA stability. For example, long 3′ UTR products of tandem APA are frequent targets of miRNA-mediated repression, enabled by increased frequency of miRNA binding sites in distal 3′UTR segments (18).

In addition to post-transcriptional control of APA products by miRNAs, the nonsense-mediated mRNA decay pathway (NMD) can regulate the abundance of long 3′UTRs produced by APA (19). NMD is a translation-dependent mRNA decay pathway dually responsible for performing RNA quality control and for regulation of many apparently normal genes (20). A major class of targets of the regulatory activity of NMD is that of mRNAs containing long 3′UTRs (21,22). 3′UTR length-sensing by NMD is mediated through UPF1, a highly conserved RNA helicase that binds mRNA in a sequence-independent fashion due to contacts on the sugar-phosphate backbone (23,24). Increased UPF1 occupancy on an mRNA increases the chances of UPF1 phosphorylation by SMG1, the principal UPF1 kinase (25,26). UPF1 phosphorylation is an important signal that favors transcript decay, leading to
recruitment of the SMG6 endonuclease and the SMG5/7 complex, which links NMD to decapping and deadenylation enzymes (27–30). Proximal polyA site usage may produce short 3′UTRs relatively immune to NMD, while distal polyA sites may produce NMD-sensitive mRNAs. Despite the characterized ability of NMD to recognize and degrade long 3′UTRs, the role of the pathway in post-transcriptional regulation of APA events has not been extensively explored.

The NMD sensitivity of transcripts with long 3′UTRs can be modulated by two cellular proteins that act as protective factors, PTBP1 and hnRNP L (31,32). When these proteins bind near the termination codon (TC) of a mature transcript, they prevent UPF1 accumulation on the 3′UTR, inhibiting decay. In addition to their anti-NMD functions, both PTBP1 and hnRNP L have roles in regulation of splicing and polyA site selection (14,33–36). Thus, these proteins may affect the abundance of mRNA isoforms both by regulating their production in the nucleus and by modulating their stability in the cytoplasm.

In the present study, we investigate whether NMD influences the relative abundance of long and short 3′UTR transcript isoforms produced by tandem APA. We find that NMD systematically targets long 3′UTR products of APA due to direct binding of UPF1 to extended 3′UTR segments. Our analyses indicate that many APA products regulated by NMD may not be detected by transcript-level approaches, instead requiring interrogation of the relative abundances of individual transcript segments. Based on prior results, we expect that PTBP1 affects APA product abundance at two levels: control of polyA site choice in the nucleus and antagonism of NMD on long 3′UTR APA products in the cytoplasm. However, our data provide evidence that PTBP1 binds similar 3′UTR regions for both functions, meaning that direct evaluation of NMD inactivation and RNA stability is required to unambiguously determine the mechanism by which it affects the abundance of specific transcript isoforms. Together, these findings demonstrate that APA products represent a frequently undetected class of NMD targets under negative regulation by UPF1 and positive regulation by PTBP1.

MATERIALS AND METHODS

Cultured cells and siRNA

HEK-293 Tet-Off (Clontech) and Flp-In T-REx-293 cells (ThermoFisher Scientific) were maintained at 37°C in ambient oxygen and 5% CO2 in DMEM (Gibco #11965-092) supplemented with 10% fetal bovine serum (Gibco) and a 1× penicillin, streptomycin and L-glutamine mixture (Gibco). Human Flp-In T-REx-293 cells expressing 3×FLAG-PTBP1 or CLIP-UPF1 were generated following the manufacturer’s protocol (ThermoFisher Scientific). In brief, Flp-In T-REx-293 cells were transfected with pcDNA5/FRT/TO-PTBP1 or pcDNA5/FRT/TO-CLIP-UPF1 and pOG44 plasmids (ThermoFisher Scientific) with TurboFect or LipoFectamine 3000 transfection reagents according to the manufacturer’s suggestions (ThermoFisher Scientific). Cells were selected for hygromycin resistance (100 µg/ml Hygromycin B (Invitrogen) for 10–14 days). Polyclonal cells were treated with doxycycline hyclate (1 µg/ml final concentration, ThermoFisher Scientific) for 48–72 h for induction of transgene expression. Transgene expression was confirmed by western-blot analysis (anti-FLAG, Sigma; anti-UPF1, Bethyl Laboratories).

siRNAs used in this study were non-targeting siRNA: Silencer Select negative control #2 siRNA (ThermoFisher Scientific #4390846); UPF1 siRNA: GAUGCAGUUCCGCUCAUUUU; PTBP1 siRNA: CUUCGUCAUUCCGAGAAUU (ThermoFisher Scientific; (37,38)).

Plasmids

Full-length PTBP1 cDNA was amplified by reverse transcription-polymerase chain reaction (RT-PCR) from HEK-293 total RNA and cloned into the tetracycline-inducible expression vector pcDNA5/FRT/TO (ThermoFisher Scientific), which was modified to harbor an N-terminal 3×FLAG tag (39).

PTBP1 eCLIP
eCLIP was performed as previously reported, with the additional use of infrared dye-labeled 3′ adapter oligonucleotides to allow visualization of crosslinked RNA–protein species (40,41). Details of oligonucleotide labeling and eCLIP are provided in the Supplementary Methods. RNA-seq data from eCLIP IP and input samples was analyzed using the ENCODE eCLIP-seq Processing Pipeline (42). For metagene analysis, the polyA sites marking the largest fold-change in expression of adjacent 3′UTR segments consistent with TreeFar evaluation were used as ‘regulated polyA sites.’ As controls, polyA positions from non-regulated genes were selected at random. RBP-Maps was used to determine peak density at the indicated positions, using the top 50% most significant eCLIP peaks as judged by the SMI normalization functionality in the ENCODE pipeline and analyzed using the Fisher exact test (43).

TreeFar: segment-level analysis and decision tree design

Creation of the custom transcriptome for polyadenylation analysis. The analysis exploits human polyA sites in the PolyA_DB 3 repository curated by the Tian group (44). We extracted the polyA sites annotated as ‘3′UTR’ for each transcript from the database and associated a ReelSeq ID to each transcript: for each gene that has multiple mRNA isoforms, the isoform that covered the most 3′UTR polyA annotations was selected. Additionally, we excluded genes in which the termination codon was not in the last exon. While transcripts with 3′UTR introns represent potential NMD targets, their removal allowed us to focus only on the effects of 3′UTR length rather than exon–junction complexes in the 3′UTR. Genes not represented in PolyA_DB 3 were not evaluated. The finalized set contained 15 844 genes. In order to create the custom transcriptome, each transcript was sectioned based on polyA sites that appeared in greater than 30% of datasets referenced in PolyA_DB 3. The first section of each transcript was from the annotated 5′ end of the isoform to the first polyA site from the database. Each subsequent section was from the previous polyA site to the next polyA site, including polyA sites beyond the annotated 3′ end of the transcript (schematized in Figure 1A).
Pseudoalignment, quantification and decision tree design.

The coordinates from the sectioned transcripts were used for pseudoalignment of RNA-seq datasets by kallisto (45). Segment-by-segment differences in abundance between conditions, and the significance of those changes, were calculated based on kallisto TPM output. TreeFar was used to assign weights to the changes between segments, which were then converted to an overall score to determine whether the longer 3′UTR isoforms were more abundant (see Supplementary Methods for full details). TreeFar was designed using jsland and JMP 14.0.0 and converted to Python for ease of use.

Transcript-level changes in abundance

For transcript-level analyses, kallisto was used for the pseudoalignment of RNA-seq reads against the standard RefSeq transcriptome (45). Abundance changes between conditions were calculated by averaging the TPMs for each transcript across trials and then calculating the log2FC between conditions. kallisto was used for this process in order to keep the tools of this study consistent, but the results of this analysis conform to the outcome of DESeq2 analysis on the same dataset (data not shown; (46)).

Metabolic labeling

The method used for metabolic labeling has been documented in our previous work (31,47). Briefly, HEK-293TO cells were reverse-transfected with a gene-specific or non-targeting control siRNA for 72 h. At the end of the depletion, cells were treated with 0.5 mM 5-ethyluridine (5-EU) for 60 min. RNA was isolated using TRIzol. Total and nascent RNA levels in each sample was partitioned using the Click-iT Nascent RNA Capture Kit (ThermoFisher Scientific, Cat. No. C10365) following the manufacturer’s protocol. mRNA abundance was determined using qRT-PCR. This approach was used to assess whether the synthesis rate of a given transcript (as demonstrated by the captured fraction) differed among experimental conditions (48). Individual halflives were determined using the equation: \( t_{1/2} = -\frac{tL * \ln(2)}{\ln(1/R)} \), where \( tL \) is the 5-EU labeling time in minutes and \( R \) is the abundance in nascent RNA fraction/abundance in total RNA fraction (39,49). At least four independent biological replicates were performed for each experimental condition.

UPF1 RNA affinity purification

Flp-In T-REx-293 stable cell lines expressing CLIP-tagged UPF1 (above) were seeded in 6 × 15 cm plates and then treated with 200 ng/ml doxycycline hyclate (Sigma) for 48 h to induce CLIP-UPF1 expression. In parallel, a human cell line that had been stably integrated with GFP was used as a control. CLIP-tagged proteins were covalently labeled with CLIP-Biotin (New England Biolabs; 10 μM final) at 4°C and purified using Dynabeads™ MyOne™ Streptavidin T1 (Invitrogen). A total of three biological replicates from each condition were processed. Sequencing libraries were prepared from input and bound RNA using the Illumina TruSeq Stranded Total RNA Human kit and sequenced on an Illumina HiSeq 3000 instrument. Details of extract preparation, labeling, affinity purification and RNA extraction are provided in Supplementary Methods.

qRT-PCR

For steady state samples, RNA was extracted from cells using TRIzol and treated with RQ1 DNase (Promega, Madison, WI, USA). A total of 500 ng of each sample was used for cDNA synthesis using the Maxima First Strand cDNA Synthesis Kit for qRT-PCR (Thermo Scientific, Philadelphia, PA, USA). cDNAs were diluted 1:20 with water and used for qPCR with iTaq Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) on a LightCycler 96 thermocycler (Roche, Basel, Switzerland). For steady-state relative isoform abundance quantification, abundances were calculated by the \( \Delta \Delta Ct \) method normalized to the abundance of the coding sequence (CDS) primer pair and then the non-targeting condition. At least three independent biological replicates were performed for each condition, and statistical significance was assessed by two-tailed Student’s \( t \)-test. For metabolic labeling samples, reverse transcription was carried out as part of the Click-iT Nascent RNA protocol and qRT-PCR was completed as above for nascent and total RNA fractions. Abundances were calculated by the \( \Delta \Delta Ct \) method normalized to the abundance of GAPDH. The cDNA abundance in each biological replicate was assessed at least twice to minimize technical variation. Primer sequences are listed in Supplementary Table S1.

Software

Graphical representations of RNA-seq reads and eCLIP peaks were generated by the Integrative Genomics Viewer (IGV 2.3.68). Statistical tests and half-life calculations were performed using Microsoft Excel version 16.28. Prism 8.4.1 for MacOS was used for graphing as well as error analysis.

RESULTS

Analysis of the role of NMD in determining 3′UTR isoform abundance

Many human genes contain multiple potential sites for co-transcriptional cleavage and polyadenylation. Differential selection of polyA sites within the same terminal exon leads to production of mRNA isoforms with common coding sequences and partially overlapping 3′UTRs. Most transcript-level gene expression analyses include reads that map to mRNA segments shared by multiple isoforms, limiting their utility in revealing the impact of NMD on products of APA. The limitation arises because quantification of long 3′UTR isoforms may be biased by the high abundance of short 3′UTR isoforms that share common 5′UTR and CDS regions. To avoid this problem, we pursued a 3′UTR segment-level analysis that considered reads common to multiple transcript isoforms separately from reads that uniquely identify 3′UTR segments generated by differential polyA site choice (Figure 1A).

To quantify the effects of NMD on APA products, we first created a custom transcriptome defining 3′UTR segments of interest. We used annotated sites from polyA_DB
Figure 1. A segment-level RNA-seq analysis strategy reveals mRNA isoform-specific abundance changes. (A) Computational and analytical workflow. See ‘Materials and Methods’ section for details. (B) RNA-seq trace of the BID locus. Vertical red lines indicate segment boundaries. qPCR primer positions are indicated. The quantification results from segment-level and transcript-level analyses are shown for comparison to each other. The segment from pA2 to pA3 was excluded from analysis because it did not meet the TPM cutoff. Stars indicate $P \leq 0.05$. 
3, a collection of empirically determined polyA sites from multiple cell types (44). Specifically, we selected polyA sites in 3'UTRs observed in 30% or more of the cell types used to build polyA_DB 3 and divided each transcript into segments based on their locations (Figure 1A). The first segment consisted of the 5' UTR, CDS and 3' UTR upstream of the first polyA site, followed by separate segments bounded by each additional polyA site (transcripts binned by numbers of polyA sites are shown in Supplementary Figure S1A). The abundance of each segment was then compared to the abundance of the first segment to determine relative 3'UTR isoform use \(\Delta \log_2(\text{TPM}_{\text{norm}})\). To integrate the segment-by-segment analysis into an overall evaluation of changes of isoform abundance in different experimental conditions, we developed TreeFar (Decision tree for analysis of RNA APA, see ‘Materials and Methods’ section for details). To determine whether longer or shorter 3'UTR isoforms are favored, TreeFar places the most weight on relative abundance changes of the last 3' UTR segment between conditions, but also accounts for significant abundance changes of internal segments.

We first used this strategy to assess changes in relative transcript segment abundance in HEK-293 cells transfected with non-targeting siRNA or UPF1-specific siRNA to impair NMD (31). The sensitivity of the segment-level approach is illustrated by transcripts encoding the apoptosis regulator BID (Figure 1B). Segment-level analysis showed that two BID 3' UTR isoforms were significantly more abundant with siUPF1 than non-targeting siRNA transfection \(\Delta \log_2(\text{TPM}_{\text{norm}})\) of 1.09 and 1.08, respectively. In contrast, a transcript-level analysis showed no increase in the abundance of long 3'UTR-containing BID mRNAs between these two conditions because the high abundance of short 3'UTR BID mRNAs masked the change in levels of the less abundant long isoforms. Additional examples illustrating how the segment-level analysis was used to identify transcripts with an increased abundance of the long 3'UTR isoform with UPF1 knockdown (KD) are provided in Supplementary Figure S1B.

**UPF1 systematically suppresses long 3'UTR products of APA**

TreeFar revealed that UPF1 KD induced a systematic shift toward increased usage of long 3' UTR isoforms (Figure 2A; Supplementary Figure S2A and Table S2). It should be noted that the results of TreeFar will sometimes differ from the comparison of the last segment when a significant abundance change occurs at an internal segment of the 3'UTR (Figure 2A, note red points below threshold). In the interest of benchmarking the results of TreeFar, we analyzed our RNA-seq data using the QAPA pipeline (50), one of several publicly available programs designed to identify changes in polyA site use. QAPA recapitulated our finding that UPF1 KD results in increased long 3' UTR isoforms (Supplementary Figure S2B), with a high degree of overlap with transcripts identified by TreeFar (Supplementary Figure S2C).

The large number of NMD-sensitive APA targets (1456 transcripts) were observed in 30% or more of the cell types used to build polyA_DB 3 and divided each transcript into segments based on their locations (Figure 1A). The first segment consisted of the 5' UTR, CDS and 3' UTR upstream of the first polyA site, followed by separate segments bounded by each additional polyA site (transcripts binned by numbers of polyA sites are shown in Supplementary Figure S1A). The abundance of each segment was then compared to the abundance of the first segment to determine relative 3'UTR isoform use \(\Delta \log_2(\text{TPM}_{\text{norm}})\). To integrate the segment-by-segment analysis into an overall evaluation of changes of isoform abundance in different experimental conditions, we developed TreeFar (Decision tree for analysis of RNA APA, see ‘Materials and Methods’ section for details). To determine whether longer or shorter 3'UTR isoforms are favored, TreeFar places the most weight on relative abundance changes of the last 3' UTR segment between conditions, but also accounts for significant abundance changes of internal segments.

We first used this strategy to assess changes in relative transcript segment abundance in HEK-293 cells transfected with non-targeting siRNA or UPF1-specific siRNA to impair NMD (31). The sensitivity of the segment-level approach is illustrated by transcripts encoding the apoptosis regulator BID (Figure 1B). Segment-level analysis showed that two BID 3' UTR isoforms were significantly more abundant with siUPF1 than non-targeting siRNA transfection \(\Delta \log_2(\text{TPM}_{\text{norm}})\) of 1.09 and 1.08, respectively. In contrast, a transcript-level analysis showed no increase in the abundance of long 3'UTR-containing BID mRNAs between these two conditions because the high abundance of short 3'UTR BID mRNAs masked the change in levels of the less abundant long isoforms. Additional examples illustrating how the segment-level analysis was used to identify transcripts with an increased abundance of the long 3'UTR isoform with UPF1 knockdown (KD) are provided in Supplementary Figure S1B.

**UPF1 systematically suppresses long 3'UTR products of APA**

TreeFar revealed that UPF1 KD induced a systematic shift toward increased usage of long 3' UTR isoforms (Figure 2A; Supplementary Figure S2A and Table S2). It should be noted that the results of TreeFar will sometimes differ from the comparison of the last segment when a significant abundance change occurs at an internal segment of the 3'UTR (Figure 2A, note red points below threshold). In the interest of benchmarking the results of TreeFar, we analyzed our RNA-seq data using the QAPA pipeline (50), one of several publicly available programs designed to identify changes in polyA site use. QAPA recapitulated our finding that UPF1 KD results in increased long 3' UTR isoforms (Supplementary Figure S2B), with a high degree of overlap with transcripts identified by TreeFar (Supplementary Figure S2C).

The large number of NMD-sensitive APA targets (1456 of 5981 total genes analyzed; Figure 2B) demonstrated that the known role of NMD in targeting substrates with long 3'UTRs extends to preferential clearance of long APA products. The analysis of this set with traditional transcript-level quantification, however, did not reveal this effect (Figure 2C). Thus, transcriptome-wide, this class of NMD targets would be missed, consistent with the BID example shown in Figure 1B.

Because APA products arising from a single gene differ only in the length and information content of their 3'UTRs, analysis of alternative 3'UTR isoforms is an opportunity to probe the 3'UTR length-dependent effects of NMD. In order to investigate the difference between the lengths of NMD-sensitive versus -resistant 3'UTRs, we examined transcripts with two polyA sites, dividing each gene into ‘short’ and ‘long’ isoforms (Supplementary Figure S2D). Among these transcripts, those that use the longer 3'UTR segments more with UPF1 KD tend to have significantly shorter ‘short’ 3'UTRs when compared to the overall group, with a median value of 335 nucleotides (nt) compared to 422 nt overall. Additionally, NMD-sensitive transcript isoforms tend to have significantly longer 3'UTRs than non-sensitive isoforms. These findings suggest that NMD predominantly affects genes with large differences in alternative 3'UTR lengths, where active NMD will favor expression of very short, NMD-insensitive isoforms.

To validate the transcriptome-wide trend of increased long 3'UTR abundance upon NMD inactivation, we selected several genes with varying magnitudes of response to UPF1 KD (Figure 2A). For steady-state isoform abundance differences, we used qRT-PCR to compare amplicons generated from the CDS of a transcript to those from its 3'UTR extension (primer positioning illustrated in Figure 1B). Under conditions of UPF1 depletion, each of the selected transcripts exhibited increased relative abundance of the long isoform compared to the total (Figure 2D), validating the categorizations made by TreeFar. To distinguish between control of 3'UTR isoform abundance through polyA site choice versus differential decay by NMD, we used metabolic labeling with 5-ethynyluridine (5-EU). In this analysis, the long 3'UTR isoform transcripts were stabilized with UPF1 KD (Figure 2E), suggesting that the effect of UPF1 on APA product abundance is indeed via RNA decay. This conclusion is further supported by more rapid turnover of long 3'UTR isoforms than the total transcript pool derived from each gene (compare gray bars in Figure 2E to those in Supplementary Figure S2E).

It has become increasingly evident that the decay-promoting activities of UPF1 extend beyond mRNAs classically defined as NMD targets (51–53). A variety of cellular conditions and signals, including structured RNAs and glucocorticoid exposure, may require disparate accessory factors to induce mRNA decay though UPF1. For insight into the consequences of these other UPF1-dependent pathways on transcript 3'UTR isoform abundance, we analyzed additional datasets with TreeFar.

One pathway of interest is Staufen-mediated decay (SMD), which degrades specific transcripts with extended double-stranded regions (51,54–56). In myoblasts depleted of UPF1, TreeFar identified a transcriptome-wide shift to longer 3'UTRs (102 transcripts are shorter with UPF1 KD compared to 1554 that are longer), similar to HEK-293 cells (Supplementary Figure S3A). However, when STA1 was depleted (56), the transcriptome-wide trend in 3'UTR iso-
Compromise of NMD globally increases abundance of long 3′ UTR isoforms. (A) Volcano plot representing $\Delta \log_2 (TPM)_{\text{norm}}$ of last 3′ UTR segments between non-targeting and UPF1 knockdown. Horizontal line indicates the significance threshold $P \leq 0.05$ ($n = 3$). Points are color coded in accordance with the results of TreeFar as indicated in the figure key. (B) Results of TreeFar, color-coded as in (A). (C) CDF plot of $\log_2$-FC of transcripts quantified using transcript-level analysis grouped by the classes in (A). Significance was evaluated with the two-tailed Kolmogorov–Smirnov (K–S) test. (D) Relative abundance of long isoforms compared to transcript CDS in siNT and siUPF1 knockdown conditions determined using qRT-PCR. An example of primer placement is shown in Figure 1B. The ratio of relative siUPF1 amplification was normalized to that of the non-targeting condition ($n = 3$). Significance was determined by two-tailed Student’s $t$-test ($* P \leq 0.05$; error bars = 1SD). (E) 5-EU incorporation to compare the half-lives of transcript long isoforms under conditions of UPF1 knockdown or non-targeting siRNA ($n = 4$). Genes to the left of the vertical line are controls. Significance was determined by using a two-tailed Student’s $t$-test ($* P \leq 0.05$; error bars = 1SD).

form abundance changes was less extreme than with UPF1 KD (427 transcripts shortened with STAU1 KD and 678 lengthened, Supplementary Figure S3B). Comparison of the lengthened transcripts in UPF1 KD and in STAU1 KD revealed that most of the transcripts lengthened with STAU1 KD were also UPF1 targets, but UPF1 depletion altered the abundance of numerous additional mRNA 3′ UTR isoforms (Supplementary Figure S3C).

In the case of glucocorticoid receptor-mediated decay (GMD), GR activation has been described to elicit UPF1-dependent decay (52,57). Analysis of T47D A1–2 breast cancer cells treated for 8 h with dexamethasone (58) showed that GR activation had the smallest effect on relative APA isoform abundance changes of the datasets we analyzed, and the transcriptome-wide shift was also less extreme than seen in UPF1 KD (368 shorter with treatment compared to 297 longer with treatment, Supplementary Figure S3D). Still, there was partial overlap between the set of transcripts with decreased long 3′ UTR isoforms in dexamethasone-treated breast cancer and the set of transcripts lengthened...
Figure 3. Transcripts with longer 3′UTRs bind more UPF1. (A) Volcano plot representing the last 3′UTR segment $\Delta \log_2(\text{TPM})_{\text{RIP}}$ between input RNA and RNA affinity purified with CLIP-UPF1. Horizontal line indicates the significance threshold equal to $P \leq 0.05$ ($n=3$). (B) CDF plot of $\Delta \log_2(\text{TPM})_{\text{RIP}}$ from (A). Genes were divided into classes according to the TreeFar analysis of siUPF1 knockdown shown in Figure 1B. Significance was calculated by K–S test. (C) CDF plot of RIP-seq data from (A) quantified using transcript-level analysis. Transcripts were grouped according to the TreeFar analysis outcome shown in Figure 1B. Significance was calculated by K–S test.

by UPF1 depletion in HEK-293 cells (Supplementary Figure S3E). Together, these results suggest that other pathways involving UPF1 do not impact the relative abundance of 3′UTR isoforms as broadly as UPF1 but raise the possibility that UPF1 may interact with multiple additional factors to determine the stability of some mRNA products of APA.

UPF1 preferentially recognizes extended 3′UTR segments

Our metabolic labeling studies revealed that long 3′UTR isoform abundance changes in UPF1 KD were associated with increased stability (Figure 2E), suggesting that the affected isoforms are NMD targets. We next used a biochemical approach to gain further insight into whether the transcript isoform changes we observed upon UPF1 KD were directly regulated by UPF1. To analyze UPF1 association with specific mRNA isoforms, we performed RNA-seq of mRNAs bound to affinity purified UPF1 and used TreeFar to analyze the results (Supplementary Table S3). TreeFar revealed a strong preference for UPF1 recovery of mRNA isoforms containing distal 3′UTR segments, resulting in a transcriptome-wide enrichment of the last 3′UTR segment in the RIP compared to the input RNA (Figure 3A and Supplementary Figure S4). Importantly, distal transcript segments with increased abundance upon UPF1 KD exhibited significantly increased relative recovery with affinity purified UPF1 (Figure 3B). As with the analyses of mRNA abundance presented above, enrichment of UPF1-associated RNA versus input evaluated without 3′UTR segmentation did not distinguish between NMD targets and non-targets (Figure 3C). This could contribute to previous observations of a poor correlation between overall UPF1 binding and decay activities transcriptome-wide (59). Together, these data indicate that direct recognition of extended 3′UTR segments is responsible for widespread suppression of long 3′UTR isoforms and demonstrate that relative UPF1 occupancy can be used to distinguish this population of NMD substrates from NMD-insensitive mRNAs.

NMD regulates many genes identified as APA targets in cancer

In order to investigate the relationship between the NMD-sensitive isoforms identified here and documented APA events in disease, we compared our dataset to a previous analysis that demonstrated pervasive shortening of 3′UTRs in seven tumor types from the Cancer Genome Atlas (TCGA) (5,60). Of the 1062 genes that were identified as APA targets in one or more tumor types and met read
Figure 4. Many APA targets affected in cancer are potential NMD substrates. (A) The proportion of transcripts with increased long isoforms upon UPF1 knockdown in common with the set of mRNAs identified as APA targets in TCGA (5). mRNAs are binned by the number of tumor types that share the APA event. The significance of the over- or under-representation of these transcript bins in the UPF1 KD set was evaluated by a two-tailed binomial test (*P ≤ 0.05 for over-represented bins). (B) The overlapping genes from the bins of six or seven tumor types in (A). Genes confirmed by qRT-PCR in this study are underlined. (C) Overlap between genes from the OncoKB or COSMIC inventories and those with 3′ UTR isoform sensitivity to NMD identified by TreeFar in this study. For either inventory, the proportion of cancer drivers in the NMD set was compared to the proportion of cancer-relevant genes in all human protein-coding genes (63) using a two-sided Fisher’s exact test (*P ≤ 0.05 for enrichment). Gene identities are listed in Supplementary Table S4.

count criteria in the present study, we found that 384 were present in our NMD-sensitive set of 1456 genes (Figure 4A). In each of the seven tumor types, our analysis identified a consistent 43–48% of the cancer APA targets as potential NMD substrates (Supplementary Figure S5). Interestingly, the overlap between long 3′ UTR isoforms repressed by NMD and the previously identified APA events in cancers was greater for events represented in multiple tumor types. Of the 39 genes affected in six tumor types, 24 exhibited increased long 3′ UTR isoform expression in UPF1 KD and 14 of 15 genes identified in all tumor types were identified here as UPF1 targets (Figure 4A). The overlapping genes are involved in a variety of cellular processes, from protein complex assembly to apoptosis (Figure 4B). Manipulation of the SMD and GMD pathways did not have a transcriptome-wide effect on the APA isoforms of the most highly shared cancer genes (Supplementary Figure S3, volcano plots: boxed points).

We next asked whether genes with NMD-regulated APA isoforms are enriched in genes known to be cancer drivers (COSMIC gene census, currently 723 genes) or strongly associated with cancer (OncoKB cancer genes, currently 1053 genes) (61,62). Indeed, genes from both these lists were over-represented among transcripts that have NMD-sensitive long 3′ UTRs compared to their proportion in the protein-coding genes in the genome as a whole (Figure 4C) (63). Specific genes identified in each of these comparisons are listed in Supplementary Table S4. Together with a recent report of increased NMD efficiency in many cancers, these data suggest that NMD may contribute to preferential expression of short 3′ UTR isoforms recognized as a common feature of cancer transcriptomes (64).

Multiple roles of PTBP1 in determining 3′ UTR isoform abundance

The data presented thus far show that steady-state abundance of transcript isoforms, widely thought to be driven by co-transcriptional polyA site choice, can also be substantially affected by NMD. Given this effect, we wondered whether the multifunctional RBP PTBP1, which has been implicated in the regulation of pre-mRNA cleavage and polyadenylation (14,33,35) and the stabilization of potential NMD target mRNAs (32), can alter APA product sensitivity to NMD. To begin to deconvolve the role of PTBP1 in polyA site selection from its role in transcript protection, we analyzed relative 3′ UTR isoform abundance in cells depleted of PTBP1, either alone or together with UPF1 KD.

Analysis of RNA-seq data through TreeFar revealed similar numbers of genes undergoing increases (1561 genes) and decreases (1677 genes) in relative abundance of long isoforms upon PTBP1 KD compared to non-targeting siRNA (Figure 5A and Supplementary Table S5), consistent with reports that PTBP1 can affect mRNA isoform abundance through multiple mechanisms. We elected to focus on transcripts that had decreased long isoform abundance with PTBP1 KD, as these are candidates for PTBP1-mediated protection from NMD (32). In the majority of transcripts in this group, expression of the extended 3′ UTR isoforms was restored when PTBP1 and UPF1 were simultaneously
knocked down as compared to PTBP1 KD alone (n = 1050) (Figure 5A and B, TRAM2; Supplementary Figure S6A, green box). Genes that underwent a decrease in long isoform abundance with PTBP1 KD without restoration by simultaneous UPF1 KD (n = 627, Supplementary Figure S6A, purple box) demonstrated behavior consistent with either a direct effect of PTBP1 on alternative polyA site usage in the nucleus or PTBP1-mediated inhibition of other RNA decay pathways (Figure 5B, TUBB). Through the process of investigating our knockdown effects on transcript isoforms, we found that the overall pattern of 3′ UTR usage alteration in the PTBP1/UPF1 double KD was highly correlated with that in the UPF1 KD alone (ρ = 0.65, Supplementary Figure S6B). This specific comparison obscures effects that may be due to PTBP1 depletion outside of the NMD context and indicates that a substantial fraction of changes in 3′ UTR isoform abundance in the double knockdown may be attributed to reduced NMD efficiency.
We used qRT-PCR to confirm that PTBP1 KD decreased the relative amount of the long isoform of selected transcripts at steady state, which could be restored with simultaneous NMD compromise (Figure 5C; traces for other example transcripts in Supplementary Figure S6C). Further, we used metabolic labeling with 5-EU to confirm that decay of selected long 3’UTR isoforms was enhanced in the absence of PTBP1 and abrogated by UPF1 KD (Figure 5D), an effect obscured in an isoform-independent approach (Supplementary Figure S6D).

Distinct PTBP1 binding profiles in NMD inhibition and APA

Binding of PTBP1 in the vicinity of polyA sites has supported the hypothesis that it directly regulates APA predominantly by suppressing polyA site recognition (14,33,36), while PTBP1 association near TCs confers NMD inhibition (32). In order to investigate the relationship between PTBP1 binding and its potential nuclear and cytoplasmic roles in determining APA outcomes, we performed eCLIP on FLAG-tagged PTBP1 stably expressed in HEK-293 cells (40). We first identified the polyA sites at which the greatest abundance changes between adjacent 3’UTR segments consistent with the TreeFar verdict were observed (e.g. sites marked ‘regulated polyA’ in Figure 5B and Supplementary Figure S6C). Importantly, 50% of the regulated polyA sites were within ~500 nt of the termination codon (Figure 6A). Metagene plots of PTBP1 eCLIP peaks revealed that the class of transcripts with decreased long isoform abundance in the absence of PTBP1 had elevated PTBP1 binding over a broad region downstream of the TC (Figure 6B). Given that many regulated polyA sites are close to the TC, binding through this region is consistent with both the NMD-protective effects of PTBP1 as well as its suppression of polyA site use (Figure 6A).

Consistent with previous findings (36), we observed elevated PTBP1 binding both upstream and downstream of regulated polyA sites associated with decreased long 3’UTR isoform abundance in PTBP1 KD (Figure 6C, red trace). In contrast, the class of transcripts with increased long 3’UTR use upon PTBP1 KD did not show evidence of increased PTBP1 binding relative to those unaffected by PTBP1 KD (Figure 6C, blue trace). We next asked whether eCLIP peak enrichment could be used to distinguish transcripts regulated by PTBP1 at the level of polyA site choice from those protected from NMD by PTBP1. When the population of transcripts shortened by PTBP1 knockdown was subdivided by response to the double KD in an effort to differentiate binding sites relevant for APA from those relevant for NMD (Figure 6D green trace compared to purple trace), no altered binding pattern emerged transcriptome-wide. Together, these findings likely reflect the close proximity and potential dual use of PTBP1 binding sites responsible for APA and NMD regulation.

DISCUSSION

Here, we uncover a widespread role for NMD in regulating the stability of extended 3’UTRs generated by APA. Coupled APA-NMD predominantly regulates the relative abundance of short (typically < 500 nt), NMD-insensitive 3’UTR isoforms and long (median length ~2400 nt), NMD-sensitive isoforms (Supplementary Figure S2D). Supporting a direct role for UPF1 in this process, we show that long 3’UTR APA products of selected genes are destabilized in the presence of UPF1 (Figure 2) and that mRNAs containing extended 3’UTR isoforms are preferentially recovered in UPF1 RIP-seq (Figure 3). The activity of NMD in suppressing expression of long 3’UTR products of APA is counteracted by PTBP1, a multifunctional RBP previously implicated in direct regulation of both NMD and pA site selection (14,32–33,35–36) (Figure 5). Using RNA stability analyses, RNA-seq of cells undergoing concurrent PTBP1 and UPF1 depletion, and eCLIP, we provide evidence that PTBP1 binding near polyA sites can affect cleavage site selection, NMD susceptibility, or both (Figures 6 and 7).

As studies of UPF1 have revealed more about its activities in the cell, the scope of its targets has broadened beyond mRNAs with aberrantly positioned termination codons (in vertebrates, operationally defined as a termination codon more than 50 nt upstream of an exon-exon junction). The expanded catalog of NMD substrates includes transcripts with upstream open reading frames, selenocysteine codons and long 3’UTRs, all of which have historically been referred to as NMD targets despite lacking nonsense mutations. In addition to its role in the turnover of this broadly defined set of NMD targets, specialized mRNA decay pathways requiring UPF1 have subsequently been described, including SMD, GMD and histone mRNA decay (65). While we find some genes with altered 3’UTR usage responsive to STAU1, and to a lesser extent GR activation, most of the events studied here do not appear to be affected by either specialized decay pathway (Supplementary Figure S3). Similarly, disruption of STAU1 or induction of GR did not lead to a systematic change in transcripts with frequently altered 3’UTR isoform expression in cancer. For simplicity, we refer to the class of UPF1-dependent mRNAs as NMD targets, consistent with the nomenclature widely used to describe long 3’UTR-containing UPF1-dependent decay targets. Further work will be required to elucidate the precise mechanistic details for how UPF1 engagement results in mRNA decay in these varied contexts.

One challenge in identification of NMD target isoforms is that they are often expressed at lower levels than non-NMD target isoforms from the same locus, meaning that non-target mRNAs can obscure the effects of NMD if gene- or transcript-level analyses are used. It has long been recognized that examination of the use of individual exons is necessary to understand the interplay between alternative splicing and NMD (66). Our data show that a similar approach is needed to uncover the role of NMD in determining the outcomes of APA (Figure 2A and C; Supplementary Figure S2A). This approach may be especially important when considering results generated from standard laboratory lines such as HEK-293 or other transformed cells since these cells tend to have enhanced use of proximal polyA sites, resulting in a low basal level of long 3’UTR isoform expression (18). Thus, any NMD-induced changes in the abundance of the long isoforms will be hidden unless analysis pipelines are used that are sensitive to this reservoir of cryptic NMD targets (Figure 2C).

Our discovery of over a thousand genes that show evidence of UPF1-mediated control of APA product abun-
Figure 6. Transcriptome-wide PTBP1 binding near polyA sites is associated with decreased long isoform abundance. (A) Histogram of the distance between the regulated polyA sites and the termination codon. Gray shaded region represents 50% of the area under the curve. (B) eCLIP metagene plot illustrating PTBP1 binding downstream of termination codons. Transcripts are divided into classes based on TreeFar. Shaded region around each trace indicates the standard error computed by RBP-Maps software. (C) eCLIP metagene plot for PTBP1 binding sites near regulated polyA sites. Transcripts are divided into classes based on TreeFar results. Shaded region around each trace indicates the standard error computed by RBP-Maps software. (D) eCLIP metagene plot for PTBP1 binding near regulated polyA sites in the class of transcripts with decreased long isoform abundance with PTBP1 knockdown subdivided by whether long isoform abundance was restored with simultaneous UPF1 knockdown. Shaded region around each trace indicates the standard error computed by RBP-Maps software.
Figure 7. NMD regulates APA transcript isoforms. Three different transcript classes are schematized with short and long isoforms generated in the nucleus and resulting cytoplasmic NMD sensitivity of each. Narrow regions of each transcript are 5′ and 3′ UTRs, wider regions are CDS. Vertical black lines in the CDS represent exon junctions. Zigzags in 3′ UTRs are possible polyA cleavage sites. For the class of transcripts that may have APA regulated by PTBP1 in the nucleus but is also protected from NMD by PTBP1 in the cytoplasm (orange), nuclear PTBP1 is represented in lighter purple to indicate possible dual use. Isoforms with shorter 3′ UTRs or protection are resistant to NMD, while isoforms with long 3′ UTRs lacking protection are NMD-sensitive.

dance illustrates that prior approaches substantially underestimate the scope of long 3′ UTR turnover by NMD. Among the putative NMD-targeted APA products, we validated several genes involved in translation and decay as undergoing UPF1-dependent changes in 3′ UTR isoform abundance (Figure 2D). In addition, we found that the long 3′ UTR isoform of MECP2 is repressed by UPF1 (Supplementary Figure S1B). MECP2 is mutated in the neurodevelopmental disorder Rett syndrome (67), and alteration of MECP2 APA by NUDT21 copy number variations has been implicated in neuropsychiatric disease (68).

Our findings highlight that evidence of binding near a polyA site and altered 3′ UTR isoform abundance upon protein depletion are not sufficient evidence to infer the mechanism by which a particular RBP binding site influences transcript fates (Figure 6). PTBP1 binding near polyA sites is used as evidence of direct regulation of APA (33,36), and previous biochemical studies strongly suggest that PTBP1 does fulfill this role on some APA substrates (14,35). However, our data show that further experiments, such as additional decay factor knockdowns and RNA stability assays, are needed to deconvolve the contributions of multifunctional RBPs, such as PTBP1, at multiple stages in the mRNA lifecycle. Several factors contribute to the difficulty in inferring function from PTBP1 binding sites. Most importantly, the predominant architecture of APA products regulated by NMD means that PTBP1 binding near the polyA site is potentially competent to regulate both APA and NMD (Figure 7). Experimental uncertainty may also play a role, as eCLIP provides limited spatial resolution of binding sites and RNA-seq data relying on protein depletion by RNAi often leads to partial effects on RNA isoform abundance.

Finally, we find a surprising degree of overlap between UPF1 targets and genes previously found to undergo 3′ UTR ‘shortening’ in cancer, several of which are cancer drivers (Figure 4). Multiple lines of evidence suggest that preferential use of proximal polyA sites in cancer is associated with increased expression of core cleavage and polyadenylation proteins (69), and our data indicate that outcomes of nuclear control of 3′ UTR isoform expression in cancer may be bolstered by cytoplasmic NMD sensitivity. These findings raise two possibilities that merit further investigation: first, NMD efficiency may contribute to 3′ UTR isoform expression changes in cancer and second, APA may be in part used by cancer cells to evade NMD. In support of the first hypothesis, recent analyses of the TCGA Pan-Cancer database have identified frequent concurrent amplification of NMD genes in cancer (70), accompanied by evidence of enhanced NMD efficiency. These findings suggest that increased NMD efficiency may help to drive systematic use of short 3′ UTR isoforms in many tumor types. Thus, co-
transcriptional selection of polyA sites is critical for generating APA isoforms, but the abundance of those isoforms is also subject to regulation at the level of RNA stability.

DATA AVAILABILITY
TreeFar has been developed and tested on Linux and is available at https://github.com/NHLBI-BCB/TreeFar under MIT license. The RNA-seq datasets used in this study reflecting the transcriptomes of cells with UPF1 KD, PTBP1 KD and dual UPF1+PTBP1 KD have been documented in our previous studies (31,32). Sequencing data are available from the NCBI GEO database: accessions GSE109143 (three trials for siNS and three trials for siUPF1) (31), GSE59884 (two trials each for four PTBP1 experimental conditions) (32), GSE138736 (two trials each, FLAG-PTBP1 eCLIP and size-matched input controls), GSE134059 (three trials each of CLIP-UPF1 RIP-seq and input samples), GSE141834 (three trials each control treatment and 8-h dexamethasone treatment) (58) and GSE89588 (three trials each UPF1 KD, STAU1 KD and control siRNA) (56).

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank Joseph Chapman and Clara Wang for critical evaluation of the manuscript, Fayaz Seifuddin and Mehdi Pirrooznia in the NHLBI Bioinformatics Core for advice on development and implementation of TreeFar, Thomas Baird for metabolic labeling samples and the NHLBI DNA Sequencing and Genomics Core for performing RNA-seq. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov).

Author contributions: A.K.: experimental framework, steady state qRT-PCR, metabolic labeling and related qRT-PCR, RNA-seq data analysis and polyA analysis, TreeFar development, manuscript; S.E.F.: RIP-seq; NH: PTBP1 eCLIP; ZG: samples for the RNA-seq library for the KD experiments; I.T.: Adaptation of TreeFar to Python; W.Y.: RNA-seq analysis; J.Z.: experimental framework; J.R.H.: experimental framework, RNA-seq data analysis and polyA analysis, TreeFar development, eCLIP analysis, manuscript.

FUNDING
National Institutes of Health (NIH), Intramural Research Program; National Heart, Lung and Blood Institute (NHLBI). Funding for open access charge: NIH, Intramural Research Program; NHLBI.

Conflict of interest statement. None declared.

REFERENCES
1. Mayr C. (2016) Evolution and biological roles of alternative 3’UTRs. Trends Cell Biol., 26, 227–237.
2. Edwalds-Gilbert G., Veraldi K.L. and Milcarek C. (1997) Alternative poly (A) site selection in complex transcription units: means to an end? Nucleic Acids Res., 25, 2547–2561.
3. Takagaki Y., Seipel, R.L., Peterson, M.L. and Manley, J.L. (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell, 87, 941–952.
4. Takagaki Y. and Manley J.L. (1998) Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. Mol. Cell, 2, 761–771.
5. Xia Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J. and Li, W. (2014) Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3’UTR landscape across seven tumour types. Nat. Commun., 5, 5274.
6. Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.-A., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M. et al. (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. EMBO J., 33, 878–889.
7. Kim Guisbert, K.S., Li, H. and Guthrie, C. (2006) Alternative 3’ Pre-mRNA processing in saccharomyces cerevisiae is modulated by Nab4/Herp1 in vivo. PLoS Biol., 5, e6.
8. Martin, G., Gruber, A.R., Keller, W. and Zavolan, M. (2012) Genome-wide analysis of pre-mRNA 3’ end processing reveals a decisive role of human cleavage factor I in the regulation of 3’ UTR length. Cell Rep., 1, 753–763.
9. Gruber, A.R., Martin, G., Keller, W. and Zavolan, M. (2014) Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. Wiley Interdiscip. Rev. RNA, 5, 183–196.
10. Yaffe C., Biesinger, J., Wan, J., Weng, L., Xing, Y., Xie, X. and Shi, Y. (2012) Transcription-wide analyses of CstF64- RNA interactions in global regulation of mRNA alternative polyadenylation. Proc. Natl. Acad. Sci. U.S.A., 109, 18773–18778.
11. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. and Tian, B. (2009) Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl. Acad. Sci. U.S.A., 106, 7028–7033.
12. 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.
13. Nazim, M., Masuda, A., Rahman, M.A., Nasrin, F., Takeda, J.-I., Oh, E., Ohkawara, B., Ito, M. and Ohno, K. (2017) Competitive regulation of alternative splicing and alternative polyadenylation by hnRNPH and CstF64 determines actinylcholinesterase isoforms. Nucleic Acids Res., 45, 1445–1468.
14. Castelo-Branco, P., Forurger, A., Wollerton, M., Smith, C., Moreira, A. and Proudfoot, N. (2004) Polypyrimidine tract binding protein modulates efficiency of polyadenylation. Mol. Cell. Biol., 24, 4174–4183.
15. Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D.J., Goodwin, M., Zhang, C., Sobczak, K., Thornton, C.A. et al. (2014) Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. Mol. Cell, 56, 311–322.
16. Kyburz, A., Friedlein, A., Langen, H. and Keller, W. (2006) Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3’ end processing and splicing. Mol. Cell, 23, 195–205.
17. Millevoi, S., Decorsiere, A., Loulegue, C., Iaconovii, J., Bernat, S., Antoniou, M. and Vagner, S. (2009) A physical and functional link between splicing factors promotes pre-mRNA 3’ end processing. Nucelic Acids Res., 37, 4672–4683.
18. Mayr, C. and Bartel, D.P. (2009) Widespread shortening of 3’UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell, 138, 673–684.
19. Bao, J., Vitting-Seerup, K., Waage, J., Tang, C., Ge, Y., Porse, B.T. and Yan, W. (2016) UPF2-dependent nonsense-mediated mRNA decay pathway is essential for spermatogenesis by selectively eliminating longer 3’UTR transcripts. PLoS Genet., 12, e1005863.
20. Kishor, A., Fritz, S.E. and Hogg, J.R. (2019) Nonsense-mediated mRNA decay: the challenge of telling right from wrong in a complex transcriptome. Wiley Interdiscip. Rev. RNA, 16, e1548.
21. Singh, G., Rebbapragada, J. and Lykke-Andersen, J. (2008) A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol., 6, e111.
22. Eberle, A.B., Stelder, L., Mathys, H., Orozco, R.Z. and Mühlemann, O. (2008) Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. *PLoS Biol.*, 6, e92.

23. Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H. and Conit, E. (2011) Molecular mechanisms for the RNA-Dependent ATPase activity of Upf1 and its regulation by Upf2. *mol. cell.*, 41, 693–703.

24. Hogg,J.R. and Goff,S.P. (2010) Upf1 senses 3′ UTR length to potentiate mRNA decay. *Cell*, 143, 379–389.

25. Durand, S., Franks, T.M. and Lykke-Andersen, J. (2016) Hyperphosphorylation amplifies Upf1 activity to resolve stalls in nonsense-mediated mRNA decay. *Nat. Commun.*, 7, 2434.

26. Hoshino,S., Ohno,M., Dreyfuss,G. and Ohno,S. (2006) Binding of a novel SMG-1-Urf1-RfI-Rf3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.*, 20, 355–367.

27. Nicholas, P., Gikratou, A., Josi, C., Colombo, M. and Mühlemann, O. (2018) Dissecting the functions of SMG5, SMG7, and NPRC2 in nonsense-mediated mRNA decay of human cells. *RNA*, 24, 557–573.

28. Eberle, A.B., Lykke-Andersen, S., Mühlemann, O. and Jensen, T.H. (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.*, 16, 49–55.

29. Huntzinger, E., Kaszma, I., Fauser, M., Saulère, J. and Izaurralde, E. (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*, 14, 2689–2697.

30. Koh, B., Jonas, S. and Izaurralde, E. (2013) The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev.*, 27, 2125–2138.

31. Kishor, A., Ge, Z. and Hogg, J.R. (2019) hnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *EMBO J.*, 38, e99128.

32. Ge, Z., Quek, B.L., Beemon, K.L. and Hogg, J.R. (2016) Polypyrimidine tract binding protein I protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *Elife*, 5, e11155.

33. Attig, J., Agostini, F., Gooding, C., Chakrabarti, A.M., Singh, A., Haberman, N., Zagalak, J.A., Emmett, W., Smith, C.W.J., Luscombe, N.M. et al. (2018) Heteromeric RNP assembly at LINEs controls Lineage-Specific RNA processing. *Cell*, 174, 1067–1081.

34. Huang, Y., Li, W., Yao, X., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Khavari, P.A. (2016) OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.*, 2017, doi:10.1200/JCO.2017.74.81.

35. Peroe, M., Shumate, A., Perte, G., Varabou, A., Breitwieser, F.P., Chang, Y.-C., Madugundu, A.K., Pandey, A. and Salzberg, S.L. (2018) CHESS: a human gene catalog curated from thousands of

36. 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.

37. Park, J., Seo, J.-W., Ahn, N., Park, S., Hwang, J. and Nam, J.-W. (2019) SMG7-dependent microRNA-mediated gene regulation. *Nat. Commun.*, 10, 4181.

38. Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y. and Maquat, L.E. (2009) Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.*, 21, 90.

39. Yee, B.A., Pratt, G.A., Gravel, B.R., Van Nostrand, E.L. and Yeo, G.W. (2019) RBP-Maps enables robust generation of splicing regulatory maps. *RNA*, 25, 193–204.

40. 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.

41. Kurosaki, T., Li, W., Hoque, M., Popp, M. W., L. Ermolchenko, D. N., Tien, B. and Maquat, L.E. (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3′ UTRs so as to elicit mRNA decay. *Cell*, 120, 195–208.

42. Cho, H., Park, O.H., Park, J., Ryu, I., Kim, J., Ko, J. and Kim, Y.K. (2015) Glucocorticoid receptor interacts with NPRC2 in a ligand-dependent manner to recruit UPF1 for rapid mRNA degradation. *Proc. Natl. Acad. Sci. U.S.A.*, 112, E1540–E1549.

43. Park, J., Seo, J.-W., Ahn, N., Park, S., Hwang, J. and Nam, J.-W. (2019) UPF1/SMG7-dependent microRNA-mediated gene regulation. *Nat. Commun.*, 10, 4181.

44. Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y. and Maquat, L.E. (2009) SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. *Genes Dev.*, 23, 54–66.

45. Gowravaram, M., Schwarz, J., Khilji, S.K., Urlaub, H. and Chakrabarti, S. (2019) Insights into the assembly and architecture of a Staufen-mediated mRNA decay (SMD)-competent mRNPs. *Nat. Commun.*, 10, 5054.

46. Lu, B., Lavi, E., Shiue, L., Cho, H., Katzmann, S., Miyoshi, K., Sioni, M.C., Carmel, L., Ares, M. and Maquat, L.E. (2018) Evidence for convergent evolution of SINE-directed Staufen-mediated mRNA decay. *Proc. Natl. Acad. Sci. U.S.A.*, 115, 968–973.

47. Park, O.H., Park, J., Yu, M., An, H.-T., Ko, J. and Kim, Y.K. (2016) Identification and molecular characterization of cellular factors required for glucocorticoid receptor-mediated mRNA decay. *Genes Dev.*, 30, 2093–2105.

48. Hoffman, J.A., Papas, B.N., Trotter, K.W. and Archer, T.K. (2020) Single-cell RNA sequencing reveals a heterogeneous response to glucocorticoids in breast cancer cells. *Commun. Biol.*, 3, 126.

49. Kuroskii, T., Li, W., Hoque, M., Popp, M. W., L. Ermolchenko, D. N., Tien, B. and Maquat, L.E. (2014) A post-translational regulatory switch on UPF1 controls targeted mRNA degradation. *Genes Dev.*, 28, 1900–1916.

50. Hu, Z., Yao, C. and Ahmed, A.A. (2017) A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. *Nat. Commun.*, 8, 15943.

51. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselas, H., Cole, C.G., Creatore, C., Dawson, E. et al. (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, 47, D941–D947.

52. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H. et al. (2017) OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.*, 2017, doi:10.1200/JCO.2017.74.0011.
large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.*, **19**, 208.

64. Zhao, B. and Pritchard, J.R. (2019) Evolution of the nonsense-mediated decay pathway is associated with decreased cytolytic immune infiltration. *PLoS Comput. Biol.*, **15**, e1007467.

65. Kim, Y.K. and Maquat, L.E. (2019) UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *RNA*, **25**, 407–422.

66. McGlincy, N.J. and Smith, C.W.J. (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.*, **33**, 385–393.

67. Lyst, M.J. and Bird, A. (2015) Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.*, **16**, 261–275.

68. Gennarino, V.A., Alcott, C.E., Chen, C.-A., Chaudhury, A., Gillentine, M.A., Rosenfeld, J.A., Parikh, S., Wheless, J.W., Roeder, E.R., Horovitz, D.D.G. *et al.* (2015) NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife*, **4**, e10782.

69. Turner, R.E., Pattison, A.D. and Beilharz, T.H. (2018) Alternative polyadenylation in the regulation and dysregulation of gene expression. *Semin. Cell Dev. Biol.*, **75**, 61–69.

70. Zhao, B. and Pritchard, J.R. (2019) Evolution of the nonsense-mediated decay pathway is associated with decreased cytolytic immune infiltration. *PLoS Comput. Biol.*, **15**, e1007467.