Polysome fractionation analysis reveals features important for human nonsense-mediated mRNA decay

James P. B. Lloyd[1,2#a], Courtney E. French[3,#b], and Steven E. Brenner[1,2,3,*]

[1] Plant and Microbial Biology Department, University of California, Berkeley, California 94720, USA

[2] Center for RNA Systems Biology, University of California, Berkeley, California 94720, USA

[3] Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA

[#a]Current address: ARC Centre of Excellence in Plant Energy Biology, University of Western Australia, Perth, Australia

[#b]Current address: Department of Paediatrics, School of Clinical Medicine, University of Cambridge, CB2 0QQ, UK

* Corresponding author

E-mail: brenner@compbio.berkeley.edu

1

Abstract

Nonsense-mediated mRNA decay (NMD) is a translation-dependent mRNA surveillance pathway that eliminates transcripts with premature termination codons. Several studies have tried defined the features governing which transcripts are targeted to NMD. However, these approaches often rely on inhibiting core NMD factors, which often have roles in non-NMD processes within the cell. Based on reports that NMD-targeted transcripts are often bound by a single ribosome, we analyzed RNA-Seq data from a polysome fractionation experiment (TrIP-Seq) to characterize the features of NMD-targeted transcripts in human cells. This approach alleviates the need to inhibit the NMD pathway. We found that the exon junction complex (EJC) model, wherein an exon-exon junction located ≥50 nucleotides downstream of a stop codon is predicted to elicit NMD, was a powerful predictor of transcripts with high abundance in the monosome fraction (bound by a single ribosome). This was also true for the presence of an upstream open reading frame. In contrast, as 3' UTR lengths increase, the proportion of transcripts that are most abundant in the monosome fraction does not increase. This suggests that either longer 3' UTRs do not consistently act as potent triggers of NMD or that the degradation of these transcripts is mechanistically different to other NMD-targeted transcripts. Of the ribosome-associated transcripts annotated as "non-coding", we find that a majority are bound by a single ribosome. Many of these transcripts increase in response to NMD inhibition, including the oncogenic SHNG15, suggesting many might be NMD targets. Finally, we found that retained intron transcripts without a premature termination codon are over-represented in the monosome fraction, suggesting an alternative mechanism is responsible for the low level of translation of these transcripts. In summary, our analysis finds that the EJC model is a powerful predictor of NMD-targeted transcripts, while the presence of a long 3' UTR is not.

Introduction

Gene expression is a complex set of interconnected processes including transcription, alternative splicing, RNA stability and translation. Nonsense-mediated mRNA decay (NMD) is a translation-dependent RNA decay mechanism that targets transcripts with an early stop codon (Peltz et al., 1993; Ishigaki et al., 2001). Several models have been proposed to explain what marks a stop codon as authentic, where termination of translation occurs normally, versus a premature termination codon (PTC), which leads to NMD in animals, yeast and plants (Nagy & Maquat, 1998; Amrani et al., 2004; Bühler et al., 2006; Behm-Ansmant et al., 2007; Kerényi et al., 2008; Singh et al., 2008; Hogg & Goff, 2010; Drechsel et al., 2013; Lloyd et al., 2018). Alternative splicing has a well known role in altering the coding potential of transcripts, but can also alter the translation efficiency of the transcript (Floor & Doudna, 2016; Weatheritt et al., 2016; Blair et al., 2017; Fagg et al., 2017; Reixachs-Solé et al., 2019) or change the stability of transcripts by targeting them to NMD (Jumaa & Nielsen, 1997; Hilleren & Parker, 1999; Lejeune et al., 2001; Lewis et al., 2003; Lareau et al., 2007; Ni et al., 2007; Floor & Doudna, 2016; Reixachs-Solé et al., 2019).

The EJC model, also known as the 50nt rule, has been proposed to describe how a stop codon is recognized as a PTC in mammals, which leads to degradation by NMD. Reporter constructs expressed in mammalian cells have determined that exon-exon junctions located ≥50-55 nt downstream of a stop codon would trigger NMD (Nagy & Maquat, 1998; Zhang et al., 1998a; Zhang et al., 1998b). Such a stop codon was termed a Premature Termination Codon; we designate this as PTC_{50nt}. Exon-exon junctions are associated with exon junction complexes (EJCs), which are deposited on a transcript during splicing, where the EJC marks the location of the spliced intron (Le Hir et al., 2000a; Le Hir et al., 2000b). Normally, the ribosome removes

3

EJCs during the first round of translation (Gehring et al., 2009). However, if the EJC is associated with an exon-exon junction located ≥50-55 nt downstream of a stop codon, the ribosome does not remove the EJC and the EJC signals that the stop codon is premature (Le Hir et al., 2001; Lykke-Andersen et al., 2001; Kashima et al., 2006; Lopez-Perrote et al., 2016). Alternative models suggest that transcripts can be targeted to NMD due to the presence of a long 3' UTR (Amrani et al., 2004; Bühler et al., 2006; Behm-Ansmant et al., 2007; Singh et al., 2008; Hogg & Goff, 2010; Bao et al., 2016; Fanourgakis et al., 2016; Ge et al., 2016; Kishor et al., 2019). Several intronless transcripts and transcripts with no intron in their 3' UTR have been found to be targeted to NMD, and the long lengths of their 3' UTRs have been reported to be responsible (Bühler et al., 2006; Kerényi et al., 2008; Yepiskoposyan et al., 2011; Toma et al., 2015; Balagopal & Beemon, 2017; Nyiko et al., 2017).

It has previously been reported that some NMD targets in *C. elegans* are mostly bound by a single ribosome (Barberan-Soler et al., 2009). By separating the cell's RNA out on a mass gradient, fractions of RNA corresponding to transcript bound by a single ribosome (the monosome fraction) or transcripts bound by multiple ribosomes (the polysome fractions) can be isolated. In subsequent work, NMD targets are observed at greater levels in the monosome fraction than in the polysome fractions in yeast (Heyer & Moore, 2016). Additionally, predicted NMD-targeted transcripts were depleted from the polysome fraction, relative to the total cytoplasm or monosome fraction, in human cells (Sterne-Weiler et al., 2013; Floor & Doudna, 2016; Kim et al., 2017; Mencke et al., 2017). These findings are consistent with the evidence that many NMD targets are recognized in the pioneer round of translation (Ishigaki et al., 2001; Chiu et al., 2004). If the termination of translation triggers NMD, the first translating ribosome could trigger NMD (Ishigaki et al., 2001; Chiu et al., 2004; Barberan-Soler et al., 2009). It has been shown that the nuclear cap complex promotes NMD in mammalian cells (Hwang et al., 2010), further suggesting that NMD would occur before many ribosomes have loaded on the

transcript, and often when only a single NMD-eliciting ribosome is bound. Therefore, transcripts that are most abundant in the monosome fraction are likely enriched for NMD targets.

Previous studies characterizing the features of transcripts that lead to NMD have often relied upon inhibition of NMD via direct inhibition of an NMD factor, or by inhibiting translation with drugs such as cycloheximide or emetine. All of these approaches will lead to changes in the transcriptome unrelated to NMD; for example, because NMD factors often have extensive roles in non-NMD processes (Azzalin & Lingner, 2006; Isken & Maquat, 2008; Lloyd, 2018).

In this study, we used the relative high abundance of NMD targets in the monosome fraction compared to the polysome fraction to test features previously linked to NMD: the EJC model, longer 3' UTRs, and upstream open reading frames (uORFs). By characterizing the abundance of transcripts in different polysome fractions, we are able to avoid many of the secondary effects of long term NMD inhibition and find NMD targets in physiologically normal cells. We used an RNA-Seq dataset generated from polysome fractions (Transcript Isoforms in Polysomes-Seq; TrIP-Seq) by Floor and Doudna (2016). Floor and Doudna performed polysome fractionation of a human cell line (HEK293) on a sucrose gradient and then sequenced RNA in the monosome fraction, in polysome fractions representing two through seven ribosomes binding to a single transcript and in a polysome fraction representing eight or more ribosomes bound to a single transcript. We then tested which NMD linked features were associated with transcript isoforms most abundant in the monosome fraction. We found support for the EJC model acting as a potent trigger of NMD, as transcripts with an exon-exon junction ≥50nt downstream of a stop codon (designated PTC_{50nt}^{+}) are much more relatively abundant in the monosome fraction than transcripts without a PTC_{50nt} (designated PTC_{50nt}^{-}). In contrast, we do not see a correlation between 3' UTR length and relative abundance in the monosome fraction. In the monosome

fraction, we also found enrichment of transcripts containing uORFs with a consensus Kozak sequence and transcripts annotated as non-coding RNAs, suggesting NMD could play an important role in the expression of these transcripts. Finally, we find that many retained intron transcripts are most abundant in the monosome fraction even when no premature termination codon is present. This suggests that a novel mechanism controls the sequestration of these transcripts in the monosome fraction, similar to what has been previously reported for retention in the nucleus of retained intron transcripts (Gohring et al., 2014; Boutz et al., 2015; Braun et al., 2017).

Results

Relatively high abundance of transcripts with a PTC_{50nt} in the monosome fraction supports the EJC model of NMD

NMD targets have previously been shown to be most abundant in the monosome fraction (Barberan-Soler et al., 2009; Sterne-Weiler et al., 2013; Floor & Doudna, 2016; Heyer & Moore, 2016). Motivated by this quality of NMD targets, we characterized transcript abundance across the polysome fractions for features associated with NMD. We re-analyzed polysome fractionated RNA-Seq data (Floor & Doudna, 2016) by quantifying a transcriptome assembled using transcript data after NMD inhibition, therefore capturing many more NMD targeted transcripts than in a reference transcriptome (French, 2016) (Supp Table 1). By the nature of this experiment, it is enriched for non-NMD targets because NMD targets are degraded and therefore low abundance. However, not introducing a global treatment to inhibit NMD allows us to examine potential NMD features without artefacts arising from these treatments. We examined in depth SRSF6 and hnRNP DL, which each have alternative splicing, leading to production of distinct transcript isoforms that are differentially sensitive to NMD (Lareau et al., 2007; Ni et al., 2007; French, 2016; Kemmerer et al., 2018). For the major SRSF6 PTC_{50nt}^{-} transcript isoform, 77% of the transcripts were bound by eight or more ribosomes and for the minor SRSF6 PTC_{50nt}^{-} isoform 87% were bound by eight or more ribosomes (Fig 1A). This is in contrast to the PTC_{50nt}^{+} SRSF6 transcript isoform; only 6% of this transcript isoform was bound by eight or more ribosomes (Fig 1A). The majority of the SRSF6 PTC_{50nt}^{+} transcripts were bound by either a single ribosome (39%) or by two ribosomes (23%; Fig 1A). A similar pattern emerges for hnRNP DL, where 65% of one PTC_{50nt}^{-} transcript isoforms was bound by eight or more ribosomes as was 36% of another, while 65% of the PTC_{50nt}^{+} transcript isoform was bound by a

7

single ribosome (Fig 1B). The relatively high abundance of PTC^{-}_{50nt} transcripts bound by eight or more ribosomes is consistent with these transcripts being productive and the substrate for most translation of these genes. in contrast, the relatively high abundance of PTC^{+}_{50nt} transcripts being bound by a single ribosome is consistent with these transcripts being NMD targets.

For each transcript isoform in the transcriptome, we measured the distance from the last exon-exon junction to the stop codon. We then examined how this related to peak abundance in the monosome fraction (Fig 1C). In agreement with the EJC model, we find a sharp increase in the proportion of transcripts mostly bound to a single ribosome when the last exon-exon junction is ≥50 nt downstream of the last exon-exon junction (Fig 1C). A relatively low proportion (15%) of PTC^{-}_{50nt} transcripts are most abundant in the monosome fraction (Fig 1C), while a relatively higher fraction (35%) of PTC^{+}_{50nt} transcripts are most abundant in the monosome fraction (Fig 1C). PTC^{+}_{50nt} transcripts are also relatively enriched (7%) in the polysome 2 fraction relative to the PTC^{-}_{50nt} transcripts (3%) (Fig 1C). An example of this is the SRSF10 PTC^{+}_{50nt} transcript (Supp Fig 1A). If the initiation rate is significantly high or the coding sequence significantly long, a second round of translation may start before the pioneering ribosome has terminated and initiated NMD; this would lead to two or more ribosomes binding to the NMD targeted transcripts and explain the enrichment of PTC^{+}_{50nt} transcripts in the polysome 2 fraction (Fig 1C). Although per-gene initiation rate data are not available, we do know the sequence lengths of PTC^{+}_{50nt} transcripts. By comparing the coding sequence length of PTC^{+}_{50nt} transcripts that are most abundant in the monosome fraction and the polysome 2 fraction (Supp Fig 1B), we find a small and non-significant difference in the lengths (average 213nt vs 258nt, Wilcoxon rank sum test p = 0.09). This suggests that longer coding sequences might partly account for why some NMD-targeted transcripts are most abundant in the polysome 2 fraction but other differences in these

transcripts are the main determinants. Taken together, these data show that exon-exon junctions, when located ≥50 nt downstream of a stop codon are highly associated with transcript abundance in the monosome and polysome 2 fractions. The low number of ribosomes on transcripts with a PTC_{50nt} is consistent with the role of the EJC in targeting transcripts to NMD, consistent with the EJC model.

Long 3' UTRs are not correlated with relatively high abundance in the monosome fraction

If long 3' UTRs generally act to target transcripts to NMD, we would expect to see a correlation between 3' UTR length and proportion of isoforms whose highest abundance is in the monosome fraction. To explore this, we restricted our analysis to PTC_{50nt}^{-} transcripts—since a PTC_{50nt} is already known to act as a potent trigger of NMD (Nagy & Maquat, 1998; Zhang et al., 1998a; Zhang et al., 1998b; Drechsel et al., 2013; Hurt et al., 2013; Kurosaki & Maquat, 2013; French, 2016; Lindeboom et al., 2016; Colombo et al., 2017; Lloyd et al., 2018; Hoek et al., 2019), and would confound the results. We found that increasing 3' UTR length does not correlate with an increase in transcripts that have peak abundance in the monosome fraction (Fig 2A). Indeed, we see a gentle decline in the proportion of transcripts that have peak abundant in the monosome fraction as 3' UTR length increases (Fig 2A). We also examined the profiles of transcripts whose 3' UTRs have been reported to trigger NMD in reporter construct assays (Yepiskoposyan et al., 2011; Toma et al., 2015). We find that SMG5 has a flat abundance profile across the different polysome fractions (Fig 2B) and that SMG7 and MFN2 are most abundant in the polysome 8+ fraction (Fig 2B), which is indicative of highly translated transcripts. Taken together, these data show that transcripts with long 3' UTRs are not enriched for peak abundance in the monosome fraction. For those targeted to NMD, this suggests differences in NMD of these transcripts relative to those with a PTC_{50nt}.

Physiologically relevant NMD target GADD45A is not most abundant in the monosome fraction

NMD is essential for normal development in many animals and removal of a core NMD factor leads to embryonic lethality in flies and mammals (Medghalchi et al., 2001; Metzstein & Krasnow, 2006; Weischenfeldt et al., 2008). The cause of the lethality in some animals has been linked to NMD targeting GADD45, which encodes a regulator of programmed cell death (Nelson et al., 2016). It is unclear what features of GADD45 target it to NMD, as it lacks a PTC_{50nt} and a long 3' UTR but appears to be a direct NMD target (Nelson et al., 2016). In agreement with previous reports (Viegas et al., 2007; Huang et al., 2011; Tani et al., 2012; Nelson et al., 2016; da Costa et al., 2019), we found that GADD45A did increase after NMD inhibition (Fig 3A; Supp Table 5). However, we found that both of the expressed GADD45A transcripts were bound by multiple ribosomes (Fig 3B). This suggests that while GADD45A is a direct target, many of the transcripts are efficiently translated, similar to what we observed for transcripts with long 3' UTRs that are NMD targets (Fig 2B).

Transcripts with upstream ORFs and a consensus Kozak sequence often have high abundance in the monosome fraction

Many upstream open reading frames (uORFs) have been identified as features that target a transcript to NMD or translational repression (Wang & Wessler, 1998; Mendell et al., 2004; Yepiskoposyan et al., 2011; Rayson et al., 2012; Son et al., 2017; Ishitsuka et al., 2019), likely due to the presence of exon-exon junctions downstream of the uORF's stop codon. Because not all uORFs are efficiently translated, it is difficult from sequence analysis alone to identify whether a uORF acts to repress a transcript by altering translation, stability, or both. By examining polysome fractionation data, we can gain insights into whether a uORF-containing transcript is affecting translation, including via the activity of NMD. We examined two well characterized targets of NMD (CHOP and ATF4) that are recognized for degradation through

10

the translation of their uORFs (Mendell et al., 2004; Gardner, 2008; Wang et al., 2011; Karam et al., 2015). In both cases, these transcripts are most abundant in the monosome fraction (Fig 4A), consistent with either these transcripts being translationally repressed, being targeted to NMD, or both. Transcriptome-wide analysis reveals that uORF-containing transcripts are more abundant in the monosome fraction, which is in contrast to transcripts without a uORF (Fig 4B). This is most striking for transcripts whose uORF start codon has a consensus Kozak sequence and is therefore in a strong context for the initiation of translation. Indeed, 12% of transcripts without a uORF are most abundant in the monosome fraction compared to 31% of transcripts with a uORF in a Kozak sequence (Chi-squared p = 6.29 x 10^{-39}; Fig 3B). Transcripts with a uORF start codon not in a Kozak sequence were much less enriched in the monosome fraction with only 15% for with peak abundance in the monosome fraction (Chi-squared p = 6.86 x10^{-20} when compared to no uORF transcripts; Fig 4B). This is not surprising, given that translation of the uORF is a prerequisite for it to target a transcript to NMD. Taken together, these data suggest that when uORFs act to target a transcript to NMD, they often lead to high abundance in the monosome fraction. Polysome fractionation data could have a broader role in identifying uORFs that exert a regulatory effect on the transcripts via NMD.

Most transcripts annotated as non-coding associated with ribosomes have peak abundance in the monosome fraction and many are likely NMD targets

A growing number of transcripts annotated as non-coding RNAs are being assigned functions; however, the fate of the majority of these transcripts is still unclear (Palazzo & Lee, 2015; Thiel et al., 2019). There are reported examples of non-coding RNAs being targeted to NMD (Kurihara et al., 2009; Smith et al., 2014; Colombo et al., 2017; Li et al., 2017; Zeng et al., 2018), despite the general presumption that RNAs not encoding proteins are not translated. Some annotated non-coding transcripts appear to generate short peptides (Ingolia et al., 2011; Magny et al., 2013; Bazzini et al., 2014; Ingolia et al., 2014; Fesenko et al., 2019). We

examined if any transcripts annotated as "non-coding" were found in the polysome fractions, and whether this could be informative regarding their NMD-sensitivity. We compared the proportion of the transcript types across the different polysome fractions. Approximately 15% and 35% of protein-coding PTC_{50nt}^{-} transcripts and PTC_{50nt}^{+} transcripts, respectively, are most abundant in the monosome fraction (Fig 5A). However, for predicted non-coding transcripts, we see a much larger proportion are most abundant in the monosome fraction; 60% for pseudogenes, 52% for long non-coding intergenic RNAs (lincRNAs), 61% for antisense non-coding transcripts, and 57% for other non-coding RNAs ("processed transcripts") (Fig 5A). The high monosome fraction abundance of pseudogenes is not surprising given that pseudogenes have been reported as NMD targets (Mitrovich & Anderson, 2005; Colombo et al., 2017; Muir et al., 2018). Short ORFs, including those from near-cognate start codons, likely account for the translation of these non-coding transcripts (Guttman et al., 2013). Using previously released analysis of transcript changes upon NMD inhibition in human cells (French, 2016) (Supp Table 5), we found that a majority of non-coding transcripts that are most abundant in the monosome fraction have increased expression upon NMD inhibition (Supp Fig 2).

We explored in detail SNHG15, which is annotated as a lincRNA and its three transcripts are candidates for being degraded by NMD. SNHG15 expression is associated with poor prognosis in gastric cancer and promoting cancer cell proliferation (Chen et al., 2016; Saeinasab et al., 2019). We find that SNHG15 is most abundant in the monosome fraction and increases in expression when NMD is inhibited (Fig 5B). Previously, the expression of SNHG15 was found to be responsive to the translation inhibitor cycloheximide (Tani & Torimura, 2013, 2015), consistent with SNHG15 being an NMD-targeted transcript. Taken together, these data show that when predicted non-coding transcripts associate with ribosomes, the majority are only

bound by a single ribosome and many of these transcripts are NMD-sensitive, including the oncogenic SNHG15.

Retained intron transcripts are highly abundant in the monosome fraction

We examined if different types of alternative splicing events could affect the abundance of a transcript across the polysome fractions. It has previously been demonstrated that transcripts with retained introns are often localized to the nucleus (Gohring et al., 2014; Boutz et al., 2015; Braun et al., 2017). We found that the splice type with the largest proportion of transcripts in the monosome fraction was intron retention (Fig 6A). As expected, retained intron PTC^{+}_{50nt} transcripts are highly abundant in the monosome fraction (Fig 6A). To our surprise, PTC^{-}_{50nt} transcripts were also highly abundant in the monosome fraction. Of the PTC^{-}_{50nt} transcripts with retained introns, 22% are most abundant in the monosome fraction (Fig 6A). In contrast, only 11% of PTC^{-}_{50nt} transcripts with cassette exons are most abundant in the monosome fraction (Chi-squared test $p = 1.42 \times 10^{-13}$). One possible explanation for the enrichment of retained intron transcripts in the monosome fraction is that many of these transcripts are targeted to NMD, despite the absence of a PTC_{50nt}. To test this, we examined the expression change of PTC^{-}_{50nt} transcripts whose expression is most abundant in the monosome fraction, upon NMD inhibition (French, 2016) (Supp Table 5). We found that PTC^{-}_{50nt} transcripts with cassette exons and those with retained introns have similar responses to the inhibition of NMD (Wilcoxon rank sum test $p = 0.81$; Fig 5B). Therefore, it is unlikely that NMD is leading to the higher than expected abundance of retained intron transcripts in the monosome fraction, meaning another mechanism is likely involved.

Discussion

NMD is a pervasive mechanism, influencing the expression of thousands of genes in humans (Mendell et al., 2004; Colombo et al., 2017); however, it is not a trivial matter to determine which transcripts are direct NMD targets and what features of a transcript trigger NMD. Studying the steady-state expression of transcripts after NMD inhibition has shown that exon-exon junctions located ≥50 nt downstream of a stop codon are potent triggers of NMD, while it is less clear as to what extent long 3' UTRs act to trigger NMD (Lindeboom et al., 2016; Colombo et al., 2017; Tian et al., 2017; Lloyd et al., 2018; Hoek et al., 2019). By examining mutations in tumors and the corresponding changes in expression of these genes, it was shown that the EJC model was able to account for the majority of expression changes in mutations leading to early stop codons, and found that the length of the 3' UTR was not a factor (Lindeboom et al., 2016). In single molecule studies of transcripts, exon-exon junctions were found to be a potent trigger of NMD while the long 3' UTRs tested were not (Hoek et al., 2019). In this study, we used the abundance of transcripts in the different polysome fractions to explore the features of transcripts linked to NMD. Previous studies have shown that NMD targets are abundant in the monosome fraction compared to polysome fractions (Barberan-Soler et al., 2009; Floor & Doudna, 2016; Heyer & Moore, 2016). This is to be expected given that the first ribosome to terminate at a PTC is likely to induce NMD, which inhibits further translation (Muhlrad & Parker, 1999; Sheth & Parker, 2006; Isken et al., 2008). By analyzing transcript abundance across the polysome fractions, we have been able to explore the features that trigger NMD in physiologically normal cells, without the secondary effects of inhibiting the NMD pathway. Through this approach, we found that the presence of a PTC_{50nt} led to a marked increase in peak abundance in the monosome fraction (Fig 1), in agreement with the EJC model (Nagy & Maquat, 1998; Zhang et al., 1998a; Zhang et al., 1998b; Le Hir et al., 2000a; Le Hir et al., 2000b). NMD-targeted

transcripts are often low abundance, meaning that in cells with an active NMD pathway NMD-targeted transcripts do not appear to be expressed. Eighty-four percent of identified NMD-targeted transcript isoforms (French, 2016) (Supp Table 5) were not observed in the polysome fractionization experiment. Despite being unable to detect the majority of the previously NMD-targeted transcripts in the polysome fractions, we were still able to find a strong signal for NMD-triggering features (Fig 1 and 4), and therefore we are likely underestimating the potency of NMD-triggering features such as the EJC model and uORFs.

We did not find that longer 3' UTRs of PTC^{-}_{50nt} transcripts were associated with higher relative abundance in the monosome fraction (Fig 2). These results suggest that degradation of long 3' UTR transcripts known to be subject to NMD is not occurring during the pioneer round of translation and therefore these transcripts are not most abundant in the monosome fraction. This would suggest that NMD due to long 3 'UTR transcripts has aspects distinct from that of PTC^{+}_{50nt} and uORF-containing transcripts.

Approximately half of all human transcripts are computationally identified to contain an uORF (Iacono et al., 2005). However, it is difficult to predict whether these uORFs will be translated and if so, whether they will have a repressive effect on the transcript through NMD, translational repression, or both. Consider the two uORF-containing NMD targets, CHOP and ATF4. In normal conditions, these transcripts are targeted to NMD, but when the integrated stress response is activated, both transcripts become highly translated to produce transcription factors integral to the stress response (Gardner, 2008; Wang et al., 2011; Karam et al., 2015; Pakos-Zebrucka et al., 2016). Both transcripts were highly abundant in the monosome fraction (Fig 4A) of this study, consistent with their being targeted to NMD under normal cellular conditions. Globally, we find that a larger fraction of uORF-containing transcripts are most abundant in the

monosome fraction than transcripts without a uORF (Fig 4B), consistent with many being translationally repressed or NMD targets or both. Transcripts whose uORF start codon was in a consensus Kozak sequence have larger enrichment in the monosome fraction than do those whose uORF lacks a strong Kozak sequence (Fig 4B).

Many transcripts have been identified in the human genome that appear to lack significant coding potential. These predicted non-coding transcripts come in many flavors, including pseudogenes, lincRNAs and antisense transcripts. Functions have only been assigned to a handful of non-coding transcripts (Palazzo & Lee, 2015; Thiel et al., 2019). Some such transcripts have been identified to produce small peptides (Ingolia et al., 2011; Magny et al., 2013; Bazzini et al., 2014; Ingolia et al., 2014; Fesenko et al., 2019), while others have RNA only functions (Rinn et al., 2007; Gutschner et al., 2011). Using our polysome fractionation analysis, we found that of the predicted non-coding transcripts that do associate with ribosomes, at least 50% in each category of predicted non-coding transcripts are most abundant in the monosome fraction (Fig 5A). Of the non-coding transcripts that are most abundant in the monosome fraction, a majority increase in observed expression when NMD is inhibited (Supp Fig 2), suggesting that these transcripts are direct NMD targets. We discovered that the oncogenic lincRNA SHNG15 appears to be an NMD target, indicating that NMD could be acting to control the expression of physiologically important predicted non-coding transcripts.

Many transcripts with retained introns have been shown to evade NMD because they are not exported from the nucleus (Gohring et al., 2014; Boutz et al., 2015; Braun et al., 2017). We also find ribosome-associated transcripts with retained introns, and these are enriched for greatest abundance in the monosome fraction (Fig 6A, retained intron bars). This would be expected for NMD targets; however, NMD does not appear to explain this large over-representation of retain intron transcripts in the monosome fraction. For instance, retained intron transcripts are also

over-represented in the monosome fraction when only intron-retaining PTC^-_{50nt} transcripts are considered (Fig 6A). Additionally, the expression of retained intron PTC^-_{50nt} transcripts after NMD inhibition does not indicate that they are NMD sensitive (Fig 6B). Given that uORFs also act as potent triggers for NMD and enrichment in the monosome fraction, we examined PTC^-_{50nt} transcripts without uORFs and found that the retained intron were still enriched in the monosome fraction relative to other splice types (Supp Fig 3). It is unclear what is responsible for the high abundance of these transcripts in the monosome fraction, but it is possible other RNA quality control pathways such as No-Go decay or Non-Stop decay (Frischmeyer et al., 2002; van Hoof et al., 2002; Doma & Parker, 2006) could be involved. Alternatively, intronic sequences could act to recruit factors to inhibit translation, activate degradation during translation, or both. Analysis of the translation of splice variants with ribosome profiling found that intron retention events within the coding regions of transcripts were poorly translated (Weatheritt et al., 2016), consistent with our findings that transcripts with retained introns are often only bound by a single ribosome (Fig 6A) and other polysome fractionation studies (Floor & Doudna, 2016; Blair et al., 2017). Further study of the fate of transcript with retained introns will lead to a better understand of the true diversity of roles intron retention plays in gene expression.

In this study, we have shown that using polysome fractionation coupled to RNA-Seq, in a technique called TrIP-Seq (Floor & Doudna, 2016), is a useful tool for the interrogation of the features which trigger NMD in physiologically normal cells. Normally, core NMD factors are depleted, or translation is inhibited, prior to changes in transcript steady state levels being measured to identify NMD-targeted transcripts (Mendell et al., 2004; Drechsel et al., 2013; Hurt et al., 2013; French, 2016; Colombo et al., 2017; Tian et al., 2017; Lloyd et al., 2018). Given that many NMD factors have functions unrelated to NMD (Azzalin & Lingner, 2006; Isken & Maquat,

2008; Lloyd, 2018), indirect changes to the transcriptome could mask the effects of NMD. By measuring transcript abundance in the polysome fractions of physiologically normal cells, without an inhibited NMD pathway, we were able to reveal that transcript features such as a PTC_{50nt} or uORFs, but not long 3' UTRs, act as potent triggers of NMD. The use of polysome fractionation has the potential to uncover more detailed understanding of the control of gene expression beyond what can be gained from measuring total RNA abundance through RNA-Seq and open up new ways to investigate the fates of transcripts after transcription.

Materials and methods

Datasets

We used the polysome fraction RNA-Seq data published by Floor and Doudna (2016) that was deposited in the NCBI SRA database under the GEO Accession ID GSE69352. For NMD inhibition data, we used the data analyzed by French (2016) (https://doi.org/10.6078/D1H019), which can be found in Supp Table 5. The raw reads are deposited in the NCBI SRA database under the BioProject Accession ID PRJNA294972 (SRA Accession ID SRP063462). The transcriptome was a GTF generated by French (2016) (https://doi.org/10.6078/D1H019) that is compatible with the hg19 release of the human genome and can be found in Supp Table 1.

RNA-Seq analysis

Polysome fraction data from GSE69352 was re-analyzed by mapping to a transcriptome (Supp Table 1) constructed using data from NMD inhibited cells from French (2016) (https://doi.org/10.6078/D1H019). Salmon (Patro et al., 2017) version 0.6.1 was used to quantify these transcripts from the transcriptome FASTA file. The --biasCorrect option was applied and library type was set to -l UI (inwards paired end reads and unstranded).

Transcript abundance in each sample (measured in Transcript per million; TPM) underwent a bespoke between-sample normalization; the relative amount of RNA in each fraction was estimated from the polysome fractionation experiment, taking into account the proportion of the fraction which was rRNA and non-rRNA molecules. The relative RNA abundance in each polysome fraction was estimated from a polysome profile (Figure 1—figure supplement 1 A of Floor and Doudna, 2016). By measuring the area under each curve of each fraction with ImageJ

(Schneider et al., 2012), we were able to estimate the relative amount of total RNA in each fraction (Supp Table 2).

Different polysome fractions will contain different proportions of rRNA and non-ribosomal RNA given the differing numbers of ribosomes bound to a transcript in each polysome fraction. Therefore, we normalized the relative mRNA abundance based on the expected proportions of mRNA in the fraction, as follows. The average length of non-ribosomal RNA molecules was estimated by taking the median length of transcripts (Supp Table 1). This was weighted by the abundance of each transcript, using the raw TPM value from Salmon (Patro et al., 2017). The total length of human rRNA in a ribosome is 7216 nt (sum of GenBank accessions of human rRNA: NR_023379, NR_003285, NR_003287, and NR_003286). We then estimated the proportion of rRNA to non-ribosomal RNA molecules for each fraction with these RNA lengths as follows: For the monosome fraction, it should be one ribosome on each non-ribosomal RNA transcript. However, approximately half of monosomes are empty (Liu & Qian, 2016), therefore we estimate two ribosomes to each transcript. The same ratio is applied to the polysome 2 fraction. The ratio of three to one for the polysome 3 fraction. Four to one for the polysome 4 fraction and so on until the polysome 7 fraction. For the polysome 8+ fraction, we used the estimate used by Floor and Doudna (2016) of, on average, 24 ribosomes to one transcript. This process yielded transcript abundances normalized based on the amount of RNA in each fraction from the polysome gradient.

Analysis of transcript abundances across the polysome

For each transcript, the fraction containing the highest (peak) abundance was calculated from the average of the two biological replicates, after between-sample normalization (see above). In Fig 1, the proportion of transcripts with peak abundance in each of the polysome fractions was plotted against the distance between the stop codon and the last exon-exon junction. To

mitigate effects of noise, the proportion of transcripts with peak abundance in each of the polysome fractions for each position was calculated as an average within a window of distances. If fewer than 200 transcripts had a particular distance between the stop codon and the last exon-exon junction, then the window was expanded by one, in each direction. This was repeated until at least 200 transcripts were captured in the window of distances to be averaged. No more than 350 transcripts were present in any range of window sizes. In Fig 2, the proportion of transcripts with peak abundance in each of the polysome fractions was plotted against the length of the 3' UTR. The same windowing approach was taken for the 3' UTR lengths as for the distance between the stop codon and the last exon-exon junction. For plotting, the R (R Development Core Team, 2011) package ggplots2 (Wickham, 2009) or the python module matplotlib (Hunter, 2007) was used.

When analyzing uORFs, the start codons were characterized as being in a Kozak sequence if the sequence matched the pattern RNN**ATG**GV (where R = A or G and V = A or C or G). For gene type analysis, the gene types were drawn from GENCODE version 19 (Harrow et al., 2012); we linked the gene ID from the NMD assembled transcriptome (Supp Table 1), to the Ensembl gene ID via the UCSC gene names (Rosenbloom et al., 2015).

Annotation of splicing events

We used the annotation script (generateEvents) of the splicing analysis tool SUPPA (Alamancos et al., 2015) to identify alternative splicing events within our transcriptome (Supp Table 1). Taking a GTF file as the input, generateEvents characterizes all of the alternative splicing events in all of a gene's transcripts and identifies which transcripts include or exclude a particular splicing event, such as a cassette exon. We then annotated each transcript with the splice types associated with it, and whether it was an inclusion or exclusion event for each type.

Data and code availability

Datasets and code used in this study are available for download at dash.berkeley.edu available

at: https://doi.org/10.6078/D1ZM39.

Acknowledgments

Author Contributions

Conceptualization: JPBL and SEB.

Data analysis: JPBL CEF and SEB.

Project administration: JPBL and SEB.

Writing – original draft: JPBL.

Writing – review & editing: JPBL CEF and SEB.

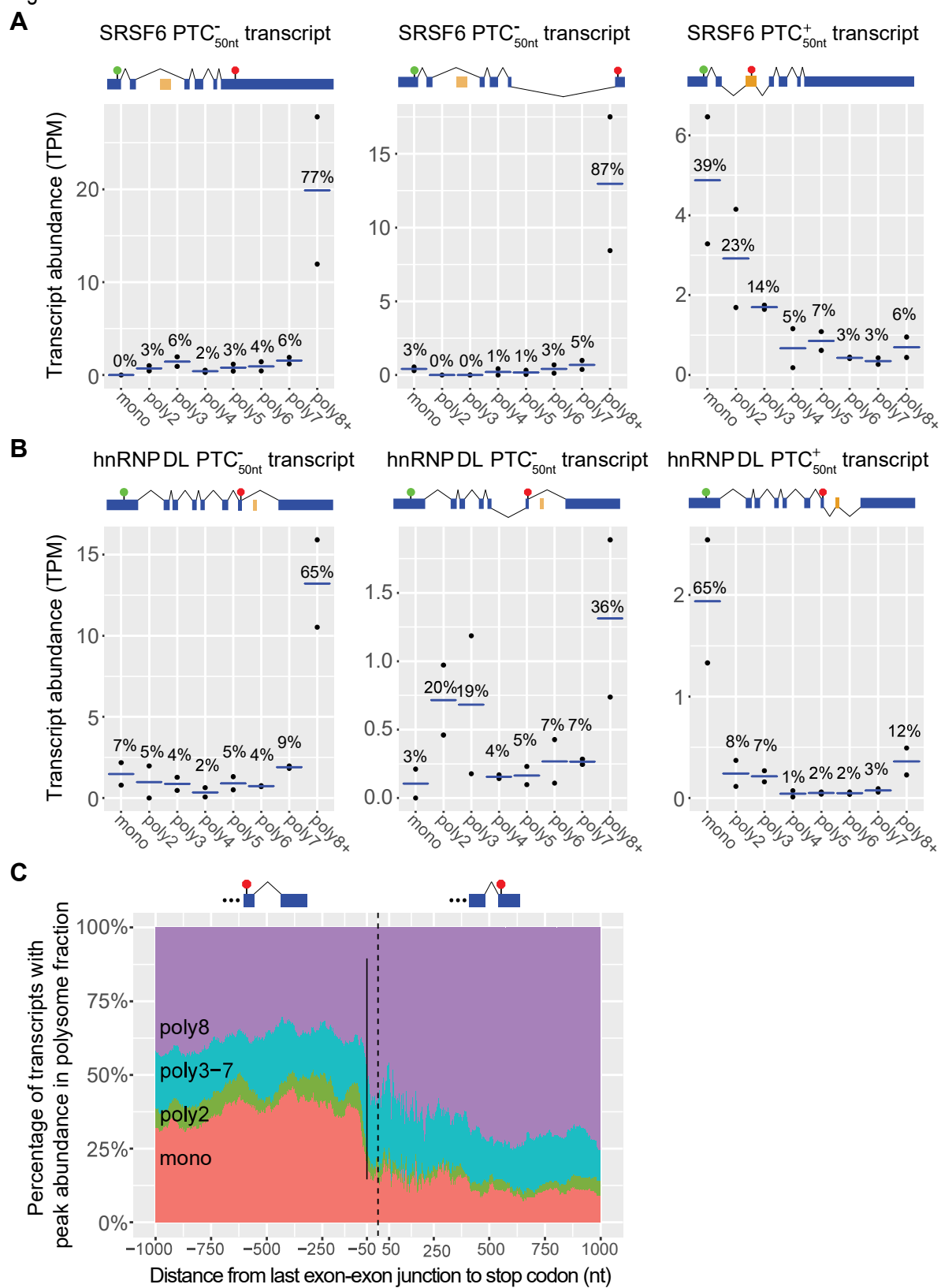S.E.B reviewed the penultimate version but not the final submitted version due to an accident injury.
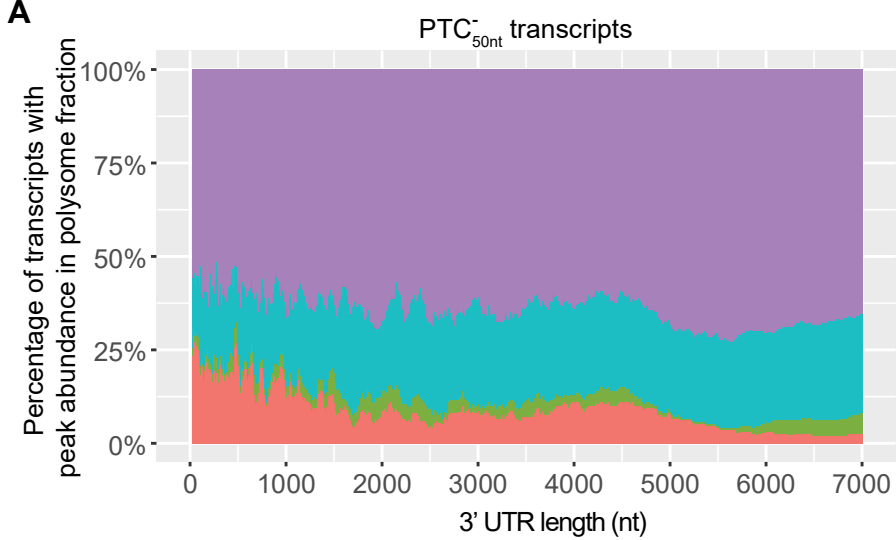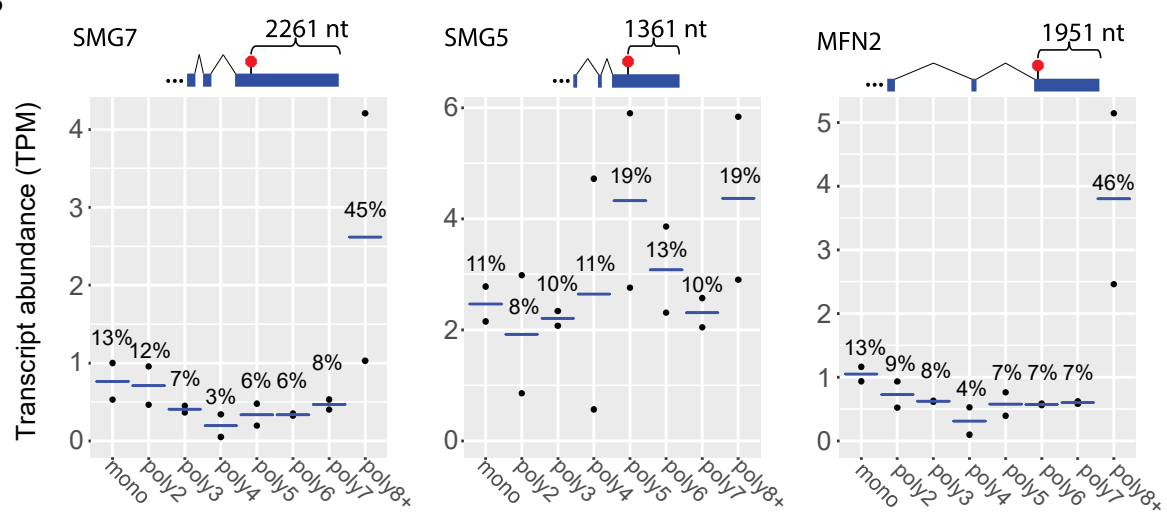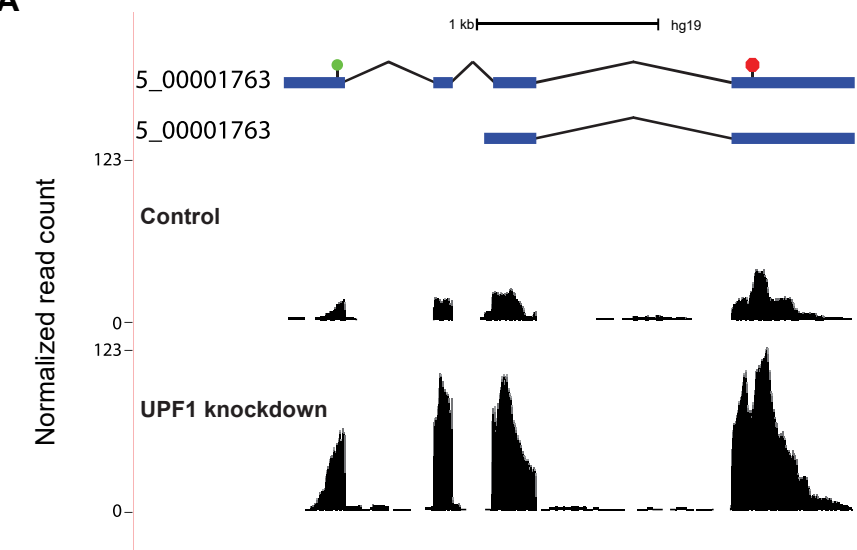
Fig 1

A

SRSF6 PTC$_{50nt}^{-}$ transcript   SRSF6 PTC$_{50nt}^{-}$ transcript   SRSF6 PTC$_{50nt}^{+}$ transcript

B

hnRNP DL PTC$_{50nt}^{-}$ transcript   hnRNP DL PTC$_{50nt}^{-}$ transcript   hnRNP DL PTC$_{50nt}^{+}$ transcript

C

Fig 2

A

PTC$^-_{50nt}$ transcripts

B

Fig 3

A

B

Fig 4

A

B

Fig 5

A

B

SNHG15

SNHG15
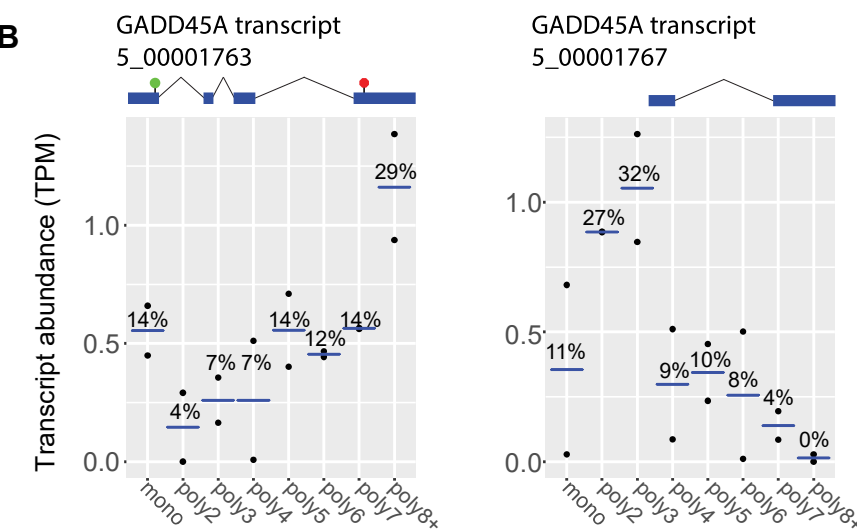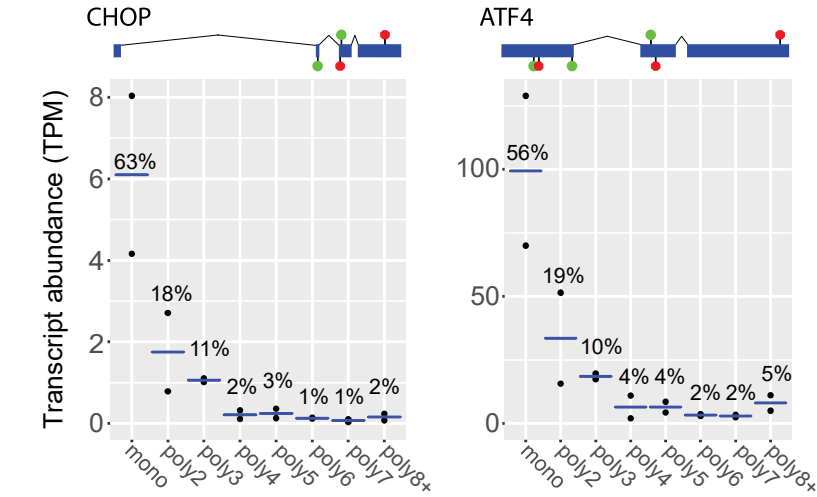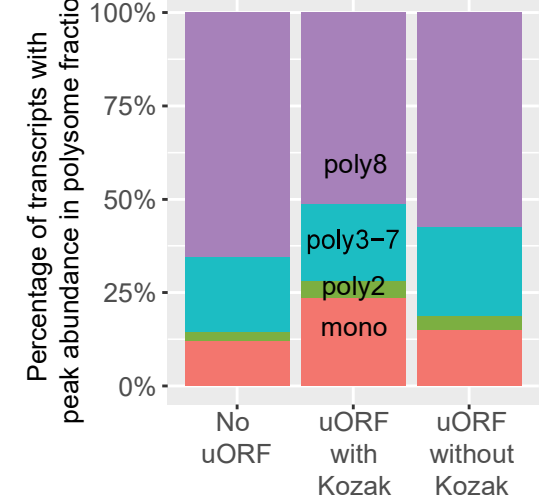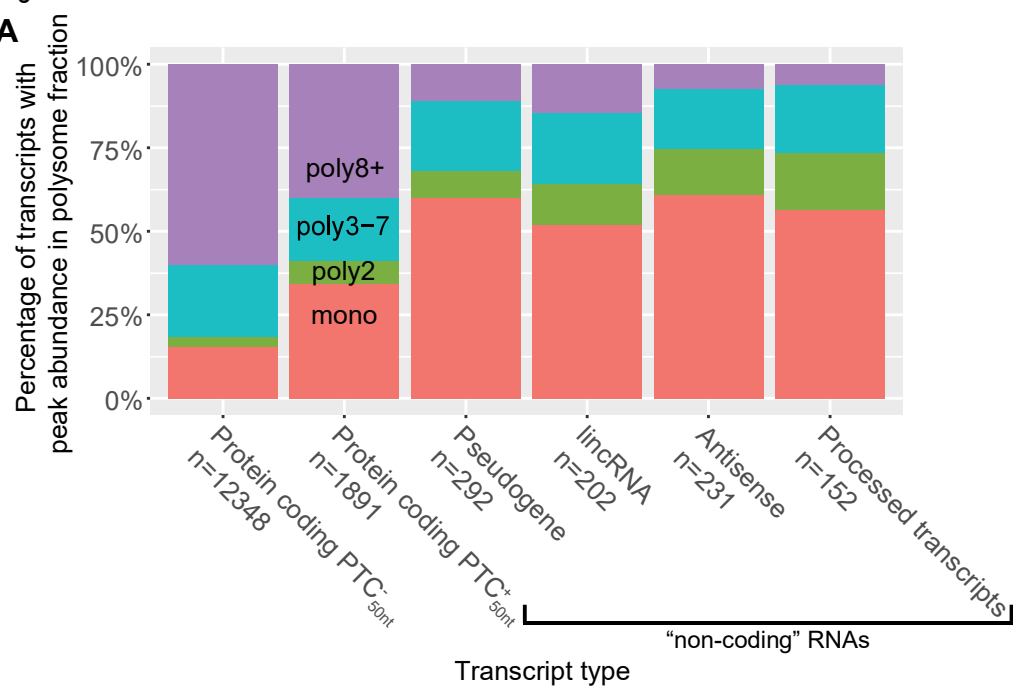
Fig 6

A

B

Figure legends

Fig 1. Transcripts with a PTC_{50nt} are enriched in the monosome fraction.
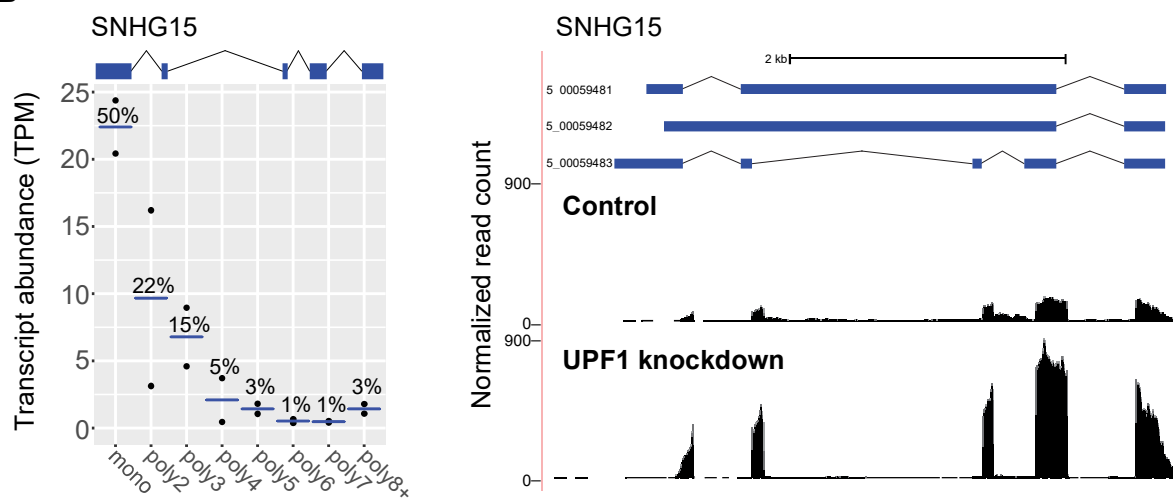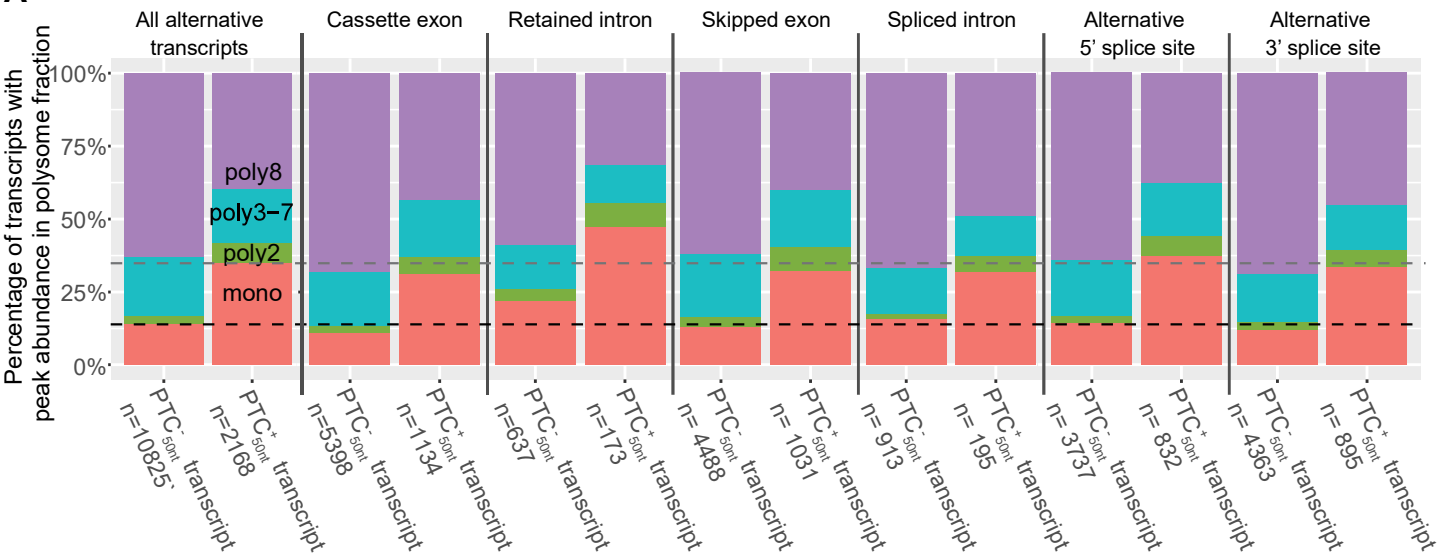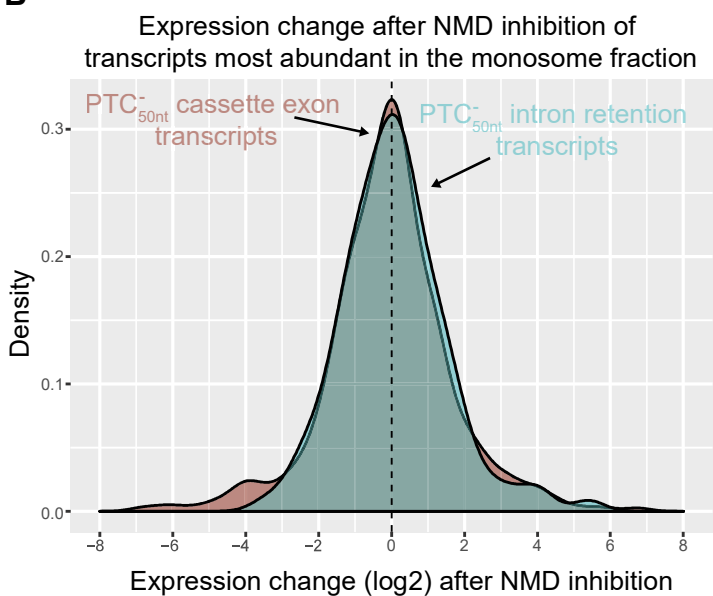
(A) The abundance of SRSF6 transcripts in the different polysome fractions according to whether they have a PTC_{50nt}. SRSF6 (ENSG00000124193) transcript IDs from left to right: 5_00146653 (PTC^-_{50nt} transcript), 5_00146654 (PTC^-_{50nt} transcript), and 5_00146652 (PTC^+_{50nt} transcript).

(B) The abundance of hnRNP DL transcripts in the different polysome fractions. hnRNP DL (ENSG00000152795) transcript IDs from left to right: 5_00037797 (PTC^-_{50nt} transcript), 5_00037799 (PTC^-_{50nt} transcript), and 5_00037796 (PTC^+_{50nt} transcript).

For (A) and (B), transcript abundance is measured in TPM normalized across the polysome fractions. Black dots represent the abundance in each of the two replicates and the blue bar represents the average abundance between the replicates. Blue exons are constitutively spliced exons. Orange exons are alternatively spliced exons. Circumflexes above the transcript model denote splicing of the major transcript isoform, while circumflexes below the transcript model represent alternative usage of a junction.

(C) The distance between stop codon final exon-exon junction versus the proportion of transcripts most abundant in each polysome fraction. Abundance of transcripts found most predominately in monosome fraction increases sharply when the stop codon is more than 50 nt before the last exon-exon junction. Positive values on the right of the plot represent transcripts whose last exon-exon junction are located before the stop codon. Negative values located on the left of the plot represent transcripts whose last exon-exon junctions are located after the stop codon. Dotted line marks zero, where the first base of the stop codon is immediately after an exon-exon junction, while the solid bar marks -50 nt. Intronless transcripts are excluded from this plot. The proportion of transcripts most abundant in each polysome fraction at each position

31

along the x axis is a average of a sliding window analysis, ensuring that each position represents at least 200 transcripts.

Fig 2. Transcripts with a long 3' UTR are not enriched in the monosome fraction.

(A) Length of transcript 3' UTRs versus the proportion of transcripts most abundant in the different polysome fractions. The proportion of transcripts most abundant in different polysome fractions at each position along the x axis is a moving average by sliding window analysis, ensuring that each position represents at least 200 transcripts.

(B) Example transcripts whose 3' UTRs have been shown to target transcripts to NMD in reporter assays (graphical conventions follow Fig 1). SMG7 (ENSG00000116698) transcript ID: 5_00003622, SMG5 (ENSG00000198952) transcript ID: 5_00007602, and MFN2 (ENSG00000116688) transcript ID: 5_00000356.

Fig 3. An NMD-targeted transcript responsible for animal lethality is not most abundant in the monosome fraction.

(A) The GADD45A transcript isoform 5_00001763 is responsive to NMD inhibition (Supp Table 5). NMD inhibition data was generated with the UCSC Genome Browser. Normalized read counts (# reads/10 million mapped reads) for control cells (top track) and NMD-inhibited cells (bottom track) were plotted.

(B) The abundance of the two expressed GADD45A transcripts in the polysome fractions (graphical conventions follow Fig 1). GADD45A (ENSG00000116717) transcript IDs from left to right: 5_00001763 and 5_00001767.

Fig 4. uORF-containing transcripts with a Kozak sequence are enriched in the monosome fraction.

(A) Example transcripts whose uORFs have previously been demonstrated to target these transcripts to NMD (graphical conventions follow Fig 1). From left to right, CHOP (ENSG00000175197) transcript ID: 5_00101020 and ATF4 (ENSG00000128272) transcript ID: 5_00152708.

(B) Comparison of PTC_{50nt}^{-} transcripts with and without a uORF according to their peak abundance polysome fraction. uORF-containing transcripts were sub-divided into those with the start codon of the uORF in a Kozak sequence and those without a Kozak sequence.

Fig 5. A majority of transcripts annotated as non-coding, that are associated with ribosomes, are most abundant in the monosome fraction.

(A) Comparison of transcript types, showing to their peak abundance polysome fraction. Transcript type is defined by the GENCODE gene type, and then further defined by the presence or absence of a PTC_{50nt} within the protein-coding transcripts.

(B) Cancer-associated lincRNA SNHG15 (ENSG00000232956; transcript ID: 5_00059483) is an NMD target (graphical conventions follow Fig 1). NMD inhibition data was generated with the UCSC Genome Brower. Normalized read counts (# reads/10 million mapped reads) for control cells (top track) and NMD-inhibited cells (bottom track) were plotted.

Fig 6. Retained intron transcripts are enriched in the monosome fraction relative to other splice types.

(A) Comparison of transcripts with different alternative splicing types according to their peak abundance polysome fraction. Dashed lines represent the average percentage of transcripts with a PTC_{50nt} (grey) and without a PTC_{50nt} (black) that are most abundant in the monosome fraction.

(B) The expression change after NMD inhibition of cassette exon or retained intron containing

PTC^-_{50nt} transcripts who are most abundant in the monosome.

References

Alamancos GP, Pages A, Trincado JL, Bellora N, Eyras E. 2015. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA 21*:1521-1531. PMCID:PMC4536314. doi:10.1261/rna.051557.115

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. 2004. A faux 3'-utr promotes aberrant termination and triggers nonsense-mediated mrna decay. *Nature 432*:112-118. doi:10.1038/nature03060

Azzalin CM, Lingner J. 2006. The double life of upf1 in rna and DNA stability pathways. *Cell Cycle 5*:1496-1498. doi:10.4161/cc.5.14.3093

Balagopal V, Beemon KL. 2017. Rous sarcoma virus rna stability element inhibits deadenylation of mrnas with long 3'utrs. *Viruses 9*. PMCID:PMC5580461. doi:10.3390/v9080204

Bao J, Vitting-Seerup K, Waage J, Tang C, Ge Y, Porse BT, 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. PMCID:PMC4858225. doi:10.1371/journal.pgen.1005863

Barberan-Soler S, Lambert NJ, Zahler AM. 2009. Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in c. Elegans. *RNA 15*:1652-1660. PMCID:PMC2743056. doi:10.1261/rna.1711109

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ. 2014. Identification of small orfs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J 33*:981-993. PMCID:PMC4193932. doi:10.1002/embj.201488411

Behm-Ansmant I, Gatfield D, Rehwinkel J, Hilgers V, Izaurralde E. 2007. A conserved role for cytoplasmic poly(a)-binding protein 1 (pabpc1) in nonsense-mediated mrna decay. *EMBO J 26*:1591-1601. PMCID:PMC1829367. doi:10.1038/sj.emboj.7601588

Blair JD, Hockemeyer D, Doudna JA, Bateup HS, Floor SN. 2017. Widespread translational remodeling during human neuronal differentiation. *Cell Rep 21*:2005-2016. PMCID:PMC5759054. doi:10.1016/j.celrep.2017.10.095

Boutz PL, Bhutkar A, Sharp PA. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes & development 29*:63-80. PMCID:4281565. doi:10.1101/gad.247361.114

Braun CJ, Stanciu M, Boutz PL, Patterson JC, Calligaris D, Higuchi F, Neupane R, Fenoglio S, Cahill DP, Wakimoto H, Agar NYR, Yaffe MB, Sharp PA, Hemann MT, Lees JA. 2017. Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. *Cancer Cell 32*:411-426 e411. PMCID:PMC5929990. doi:10.1016/j.ccell.2017.08.018

Bühler M, Steiner S, Mohn F, Paillusson A, Mühlemann O. 2006. Ejc-independent degradation of nonsense immunoglobulin-mu mrna depends on 3' utr length. *Nat Struct Mol Biol 13*:462-464. doi:10.1038/nsmb1081

Chen SX, Yin JF, Lin BC, Su HF, Zheng Z, Xie CY, Fei ZH. 2016. Upregulated expression of long noncoding rna snhg15 promotes cell proliferation and invasion through regulates mmp2/mmp9 in patients with gc. *Tumour Biol 37*:6801-6812. doi:10.1007/s13277-015-4404-0

Chiu SY, Lejeune F, Ranganathan AC, Maquat LE. 2004. The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. *Genes Dev 18*:745-754. PMCID:PMC387415. doi:10.1101/gad.1170204

Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O. 2017. Transcriptome-wide identification of nmd-targeted human mrnas reveals extensive redundancy between

smg6- and smg7-mediated degradation pathways. *RNA 23*:189-201. PMCID:PMC5238794. doi:10.1261/rna.059055.116

da Costa PJ, Menezes J, Saramago M, Garcia-Moreno JF, Santos HA, Gama-Carvalho M, Arraiano CM, Viegas SC, Romao L. 2019. A role for dis3l2 over natural nonsense-mediated mrna decay targets in human cells. *Biochem Biophys Res Commun*. doi:10.1016/j.bbrc.2019.08.105

Doma MK, Parker R. 2006. Endonucleolytic cleavage of eukaryotic mrnas with stalls in translation elongation. *Nature 440*:561-564. PMCID:PMC1839849. doi:10.1038/nature04530

Drechsel G, Kahles A, Kesarwani AK, Stauffer E, Behr J, Drewe P, Rätsch G, Wachter A. 2013. Nonsense-mediated decay of alternative precursor mrna splicing variants is a major determinant of the arabidopsis steady state transcriptome. *The Plant Cell 25*:3726-3742. doi:10.1105/tpc.113.115485

Fagg WS, Liu N, Fair JH, Shiue L, Katzman S, Donohue JP, Ares M, Jr. 2017. Autogenous cross-regulation of quaking mrna processing and translation balances quaking functions in splicing and translation. *Genes Dev 31*:1894-1909. PMCID:PMC5695090. doi:10.1101/gad.302059.117

Fanourgakis G, Lesche M, Akpinar M, Dahl A, Jessberger R. 2016. Chromatoid body protein tdrd6 supports long 3' utr triggered nonsense mediated mrna decay. *PLoS Genet 12*:e1005857. PMCID:PMC4858158. doi:10.1371/journal.pgen.1005857

Fesenko I, Kirov I, Knyazev A, Khazigaleeva R, Lazarev V, Kharlampieva D, Grafskaia E, Zgoda V, Butenko I, Arapidi G, Mamaeva A, Ivanov V, Govorun V. 2019. Distinct types of short open reading frames are translated in plant cells. *Genome Res*. doi:10.1101/gr.253302.119

Floor SN, Doudna JA. 2016. Tunable protein synthesis by transcript isoforms in human cells. *Elife 5*. PMCID:PMC4764583. doi:10.7554/eLife.10921

French CE. 2016. Prevalence and significance of nonsense-mediated mrna decay coupled with alternative splicing in diverse eukaryotic organisms. *Molecular and Cell Biology*: University of California, Berkeley.

Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002. An mrna surveillance mechanism that eliminates transcripts lacking termination codons. *Science 295*:2258-2261. doi:10.1126/science.1067338

Gardner LB. 2008. Hypoxic inhibition of nonsense-mediated rna decay regulates gene expression and the integrated stress response. *Molecular and Cellular Biology 28*:3729-3741. doi:10.1128/MCB.02284-07

Ge Z, Quek BL, Beemon KL, Hogg JR. 2016. Polypyrimidine tract binding protein 1 protects mrnas from recognition by the nonsense-mediated mrna decay pathway. *Elife 5*. PMCID:PMC4764554. doi:10.7554/eLife.11155

Gehring NH, Lamprinaki S, Kulozik AE, Hentze MW. 2009. Disassembly of exon junction complexes by pym. *Cell 137*:536-548. doi:10.1016/j.cell.2009.02.042

Gohring J, Jacak J, Barta A. 2014. Imaging of endogenous messenger rna splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in arabidopsis. *Plant Cell 26*:754-764. PMCID:3967038. doi:10.1105/tpc.113.118075

Gutschner T, Baas M, Diederichs S. 2011. Noncoding rna gene silencing through genomic integration of rna destabilizing elements using zinc finger nucleases. *Genome Res 21*:1944-1954. PMCID:PMC3205578. doi:10.1101/gr.122358.111

Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. 2013. Ribosome profiling provides evidence that large noncoding rnas do not encode proteins. *Cell 154*:240-251. PMCID:PMC3756563. doi:10.1016/j.cell.2013.06.009

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R, Hubbard TJ. 2012. Gencode: The reference human genome annotation for the encode project. *Genome Res 22*:1760-1774. PMCID:PMC3431492. doi:10.1101/gr.135350.111

Heyer EE, Moore MJ. 2016. Redefining the translational status of 80s monosomes. *Cell 164*:757-769. doi:10.1016/j.cell.2016.01.003

Hilleren P, Parker R. 1999. Mechanisms of mrna surveillance in eukaryotes. *Annu Rev Genet 33*:229-260. doi:10.1146/annurev.genet.33.1.229

Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, Vermeulen M, Tanenbaum ME. 2019. Single-molecule imaging uncovers rules governing nonsense-mediated mrna decay. *Molecular Cell 75*:324-339 e311. doi:10.1016/j.molcel.2019.05.008

Hogg JR, Goff SP. 2010. Upf1 senses 3'utr length to potentiate mrna decay. *Cell 143*:379-389. doi:10.1016/j.cell.2010.10.005

Huang L, Lou C-H, Chan W, Shum EY, Shao A, Stone E, Karam R, Song H-W, Wilkinson MF. 2011. Rna homeostasis governed by cell type-specific and branched feedback loops acting on nmd. *Molecular Cell 43*:950-961. doi:10.1016/j.molcel.2011.06.031

Hunter JD. 2007. Matplotlib: A 2d graphics environment. *Comput Sci Eng 9*:90-95. doi:Doi 10.1109/Mcse.2007.55

Hurt JA, Robertson AD, Burge CB. 2013. Global analyses of upf1 binding and function reveals expanded scope of nonsense-mediated mrna decay. *Genome Research*. doi:10.1101/gr.157354.113

Hwang J, Sato H, Tang Y, Matsuda D, Maquat LE. 2010. Upf1 association with the cap-binding protein, cbp80, promotes nonsense-mediated mrna decay at two distinct steps. *Molecular cell*. doi:10.1016/j.molcel.2010.07.004

Iacono M, Mignone F, Pesole G. 2005. Uaug and uorfs in human and rodent 5'untranslated mrnas. *Gene 349*:97-105. doi:10.1016/j.gene.2004.11.041

Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep 8*:1365-1379. PMCID:PMC4216110. doi:10.1016/j.celrep.2014.07.045

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell 147*:789-802. PMCID:PMC3225288. doi:10.1016/j.cell.2011.10.002

Ishigaki Y, Li X, Serin G, Maquat LE. 2001. Evidence for a pioneer round of mrna translation: Mrnas subject to nonsense-mediated decay in mammalian cells are bound by cbp80 and cbp20. *Cell 106*:607-617.

Ishitsuka S, Yamamoto M, Miyamoto M, Kuwashiro Y, Imai A, Motose H, Takahashi T. 2019. Complexity and conservation of thermospermine-responsive uorfs of sac51 family genes in angiosperms. *Front Plant Sci 10*:564. PMCID:PMC6504692. doi:10.3389/fpls.2019.00564

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, Maquat LE. 2008. Upf1 phosphorylation triggers translational repression during nonsense-mediated mrna decay. *Cell 133*:314-327. doi:10.1016/j.cell.2008.02.030

Isken O, Maquat L. 2008. The multiple lives of nmd factors: Balancing roles in gene and genome regulation. *Nat Rev Genet*. doi:10.1038/nrg2402

Jumaa H, Nielsen PJ. 1997. The splicing factor srp20 modifies splicing of its own mrna and asf/sf2 antagonizes this regulation. *EMBO J 16*:5077-5085. PMCID:PMC1170142. doi:10.1093/emboj/16.16.5077

Karam R, Lou CH, Kroeger H, Huang L, Lin JH, Wilkinson MF. 2015. The unfolded protein response is shaped by the nmd pathway. *EMBO Reports*. doi:10.15252/embr.201439696

Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S. 2006. Binding of a novel smg-1-upf1-erf1-erf3 complex (surf) to the exon junction complex triggers upf1 phosphorylation and nonsense-mediated mrna decay. *Genes & development 20*:355-367. doi:10.1101/gad.1389006

Kemmerer K, Fischer S, Weigand JE. 2018. Auto- and cross-regulation of the hnrnps d and dl. *RNA 24*:324-331. PMCID:PMC5824352. doi:10.1261/rna.063420.117

Kerényi Z, Mérai Z, Hiripi L, Benkovics A, Gyula P, Lacomme C, Barta E, Nagy F, Silhavy D. 2008. Inter-kingdom conservation of mechanism of nonsense-mediated mrna decay. *The EMBO Journal 27*:1585-1595. doi:10.1038/emboj.2008.88

Kim WK, Yun S, Kwon Y, You KT, Shin N, Kim J, Kim H. 2017. Mrnas containing nmd-competent premature termination codons are stabilized and translated under upf1 depletion. *Scientific Reports 7*:15833. doi:10.1038/s41598-017-16177-9

Kishor A, Ge Z, Hogg JR. 2019. Hnrnp l-dependent protection of normal mrnas from nmd subverts quality control in b cell lymphoma. *EMBO J 38*. PMCID:PMC6356069. doi:10.15252/embj.201899128

Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A, Toyoda T, Shinozaki K, Seki M. 2009. Genome-wide suppression of aberrant mrna-like noncoding rnas by nmd in arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America 106*:2453-2458. doi:10.1073/pnas.0808902106

Kurosaki T, Maquat LE. 2013. Rules that govern upf1 binding to mrna 3' utrs. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1219908110

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of sr genes associated with highly conserved and ultraconserved DNA elements. *Nature 446*:926-929. doi:10.1038/nature05676

Le Hir H, Gatfield D, Izaurralde E, Moore MJ. 2001. The exon-exon junction complex provides a binding platform for factors involved in mrna export and nonsense-mediated mrna decay. *The EMBO Journal 20*:4987-4997. PMCID:PMC125616. doi:10.1093/emboj/20.17.4987

Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000a. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mrna exon-exon junctions. *EMBO J 19*:6860-6869. doi:10.1093/emboj/19.24.6860

Le Hir H, Moore MJ, Maquat LE. 2000b. Pre-mrna splicing alters mrnp composition: Evidence for stable association of proteins at exon-exon junctions. *Genes Dev 14*:1098-1108. PMCID:PMC316578.

Lejeune F, Cavaloc Y, Stevenin J. 2001. Alternative splicing of intron 3 of the serine/arginine-rich protein 9g8 gene. Identification of flanking exonic splicing enhancers and involvement of 9g8 as a trans-acting factor. *J Biol Chem 276*:7850-7858. doi:10.1074/jbc.M009510200

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mrna decay in humans. *Proceedings of the National Academy of Sciences of the United States of America 100*:189-192. doi:10.1073/pnas.0136770100

Li L, Geng Y, Feng R, Zhu Q, Miao B, Cao J, Fei S. 2017. The human rna surveillance factor upf1 modulates gastric cancer progression by targeting long non-coding rna malat1. *Cell Physiol Biochem 42*:2194-2206. doi:10.1159/000479994

Lindeboom RG, Supek F, Lehner B. 2016. The rules and impact of nonsense-mediated mrna decay in human cancers. *Nature Genetics 48*:1112-1118. PMCID:PMC5045715. doi:10.1038/ng.3664

Liu B, Qian SB. 2016. Characterizing inactive ribosomes in translational profiling. *Translation (Austin) 4*:e1138018. PMCID:PMC4909419. doi:10.1080/21690731.2015.1138018

Lloyd JPB. 2018. The evolution and diversity of the nonsense-mediated mrna decay pathway [version 2; peer review: 4 approved]. *F1000Research 7*:1299. PMCID:PMC6173133.2. doi:10.12688/f1000research.15872.2

Lloyd JPB, Lang D, Zimmer AD, Causier B, Reski R, Davies B. 2018. The loss of smg1 causes defects in quality control pathways in physcomitrella patens. *Nucleic acids research 46*:5822-5836. PMCID:PMC6009662. doi:10.1093/nar/gky225

Lopez-Perrote A, Castano R, Melero R, Zamarro T, Kurosawa H, Ohnishi T, Uchiyama A, Aoyagi K, Buchwald G, Kataoka N, Yamashita A, Llorca O. 2016. Human nonsense-mediated mrna decay factor upf2 interacts directly with erf3 and the surf complex. *Nucleic Acids Res 44*:1909-1923. PMCID:PMC4770235. doi:10.1093/nar/gkv1527

Lykke-Andersen J, Shu MD, Steitz JA. 2001. Communication of the position of exon-exon junctions to the mrna surveillance machinery by the protein rnps1. *Science 293*:1836-1839. doi:10.1126/science.1062786

Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, Couso JP. 2013. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science 341*:1116-1120. doi:10.1126/science.1238802

Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM, Dietz HC. 2001. Rent1, a trans-effector of nonsense-mediated mrna decay, is essential for mammalian embryonic viability. *Hum Mol Genet 10*:99-105.

Mencke R, Harms G, Moser J, van Meurs M, Diepstra A, Leuvenink HG, Hillebrands JL. 2017. Human alternative klotho mrna is a nonsense-mediated mrna decay target inefficiently spliced in renal disease. *JCI Insight 2*. PMCID:PMC5846909. doi:10.1172/jci.insight.94375

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genetics 36*:1073-1078. doi:10.1038/ng1429

Metzstein MM, Krasnow MA. 2006. Functions of the nonsense-mediated mrna decay pathway in drosophila development. *PLoS Genetics 2*:e180. doi:10.1371/journal.pgen.0020180

Mitrovich QM, Anderson P. 2005. Mrna surveillance of expressed pseudogenes in c. Elegans. *Curr Biol 15*:963-967. doi:10.1016/j.cub.2005.04.055

Muhlrad D, Parker R. 1999. Recognition of yeast mrnas as "nonsense containing" leads to both inhibition of mrna translation and mrna degradation: Implications for the control of mrna decapping. *Mol Biol Cell 10*:3971-3978. PMCID:PMC25692.

Muir VS, Gasch AP, Anderson P. 2018. The substrates of nonsense-mediated mrna decay in caenorhabditis elegans. *G3 (Bethesda) 8*:195-205. PMCID:PMC5765348. doi:10.1534/g3.117.300254

Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing genes: When nonsense affects rna abundance. *Trends Biochem Sci 23*:198-199.

Nelson JO, Moore KA, Chapin A, Hollien J, Metzstein MM. 2016. Degradation of gadd45 mrna by nonsense-mediated decay is essential for viability. *Elife 5*. PMCID:PMC4848089. doi:10.7554/eLife.12876

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M, Jr. 2007. Ultraconserved elements are associated with homeostatic control of

splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev 21*:708-718. PMCID:PMC1820944. doi:10.1101/gad.1525507

Nyiko T, Auber A, Szabadkai L, Benkovics A, Auth M, Merai Z, Kerenyi Z, Dinnyes A, Nagy F, Silhavy D. 2017. Expression of the erf1 translation termination factor is controlled by an autoregulatory circuit involving readthrough and nonsense-mediated decay in plants. *Nucleic Acids Res 45*:4174-4188. PMCID:PMC5397192. doi:10.1093/nar/gkw1303

Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM. 2016. The integrated stress response. *EMBO Rep 17*:1374-1395. PMCID:PMC5048378. doi:10.15252/embr.201642195

Palazzo AF, Lee ES. 2015. Non-coding rna: What is functional and what is junk? *Front Genet 6*:2. PMCID:PMC4306305. doi:10.3389/fgene.2015.00002

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. doi:10.1038/nmeth.4197

Peltz SW, Brown AH, Jacobson A. 1993. Mrna destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes Dev 7*:1737-1754.

R Development Core Team. 2011. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rayson S, Arciga-Reyes L, Wootton L, De Torres Zabala M, Truman W, Graham N, Grant M, Davies B. 2012. A role for nonsense-mediated mrna decay in plants: Pathogen responses are induced in arabidopsis thaliana nmd mutants. *PLoS One 7*:e31917. doi:10.1371/journal.pone.0031917.g007

Reixachs-Solé M, Ruiz-Orera J, Alba MM, Eyras E. 2019. Ribosome profiling at isoform level reveals an evolutionary conserved impact of differential splicing on the proteome. *bioRxiv*:582031. doi:10.1101/582031

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. 2007. Functional demarcation of active and silent chromatin domains in human hox loci by noncoding rnas. *Cell 129*:1311-1323. PMCID:PMC2084369. doi:10.1016/j.cell.2007.05.022

Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ. 2015. The ucsc genome browser database: 2015 update. *Nucleic Acids Res 43*:D670-681. PMCID:PMC4383971. doi:10.1093/nar/gku1177

Saeinasab M, Bahrami AR, Gonzalez J, Marchese FP, Martinez D, Mowla SJ, Matin MM, Huarte M. 2019. Snhg15 is a bifunctional myc-regulated noncoding locus encoding a lncrna that promotes cell proliferation, invasion and drug resistance in colorectal cancer by interacting with aif. *J Exp Clin Cancer Res 38*:172. PMCID:PMC6480895. doi:10.1186/s13046-019-1169-0

Schneider CA, Rasband WS, Eliceiri KW. 2012. Nih image to imagej: 25 years of image analysis. *Nat Methods 9*:671-675.

Sheth U, Parker R. 2006. Targeting of aberrant mrnas to cytoplasmic processing bodies. *Cell 125*:1095-1109. PMCID:PMC1858659. doi:10.1016/j.cell.2006.04.037

Singh G, Rebbapragada I, Lykke-Andersen J. 2008. A competition between stimulators and antagonists of upf complex recruitment governs human nonsense-mediated mrna decay. *PLoS Biology 6*:e111. doi:10.1371/journal.pbio.0060111

Smith JE, Alvarez-Dominguez JR, Kline N, Huynh NJ, Geisler S, Hu W, Coller J, Baker KE. 2014. Translation of small open reading frames within unannotated rna transcripts in saccharomyces cerevisiae. *Cell reports*. doi:10.1016/j.celrep.2014.05.023

Son HG, Seo M, Ham S, Hwang W, Lee D, An SW, Artan M, Seo K, Kaletsky R, Arey RN, Ryu Y, Ha CM, Kim YK, Murphy CT, Roh TY, Nam HG, Lee SV. 2017. Rna surveillance via nonsense-mediated mrna decay is crucial for longevity in daf-2/insulin/igf-1 mutant c. Elegans. *Nat Commun 8*:14749. PMCID:PMC5347137. doi:10.1038/ncomms14749

Sterne-Weiler T, Martinez-Nunez RT, Howard JM, Cvitovik I, Katzman S, Tariq MA, Pourmand N, Sanford JR. 2013. Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res 23*:1615-1623. PMCID:PMC3787259. doi:10.1101/gr.148585.112

Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N. 2012. Identification of hundreds of novel upf1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol 9*:1370-1379. PMCID:3597577. doi:10.4161/rna.22360

Tani H, Torimura M. 2013. Identification of short-lived long non-coding rnas as surrogate indicators for chemical stress response. *Biochem Biophys Res Commun 439*:547-551. doi:10.1016/j.bbrc.2013.09.006

Tani H, Torimura M. 2015. Development of cytotoxicity-sensitive human cells using overexpression of long non-coding rnas. *J Biosci Bioeng 119*:604-608. doi:10.1016/j.jbiosc.2014.10.012

Thiel D, Conrad ND, Ntini E, Peschutter RX, Siebert H, Marsico A. 2019. Identifying lncrna-mediated regulatory modules via chia-pet network analysis. *BMC Bioinformatics 20*:292. PMCID:PMC6540383. doi:10.1186/s12859-019-2900-8

Tian M, Yang W, Zhang J, Dang H, Lu X, Fu C, Miao W. 2017. Nonsense-mediated mrna decay in tetrahymena is ejc independent and requires a protozoa-specific nuclease. *Nucleic acids research 45*:6848-6863. PMCID:PMC5499736. doi:10.1093/nar/gkx256

Toma KG, Rebbapragada I, Durand S, Lykke-Andersen J. 2015. Identification of elements in human long 3' utrs that inhibit nonsense-mediated decay. *RNA*. doi:10.1261/rna.048637.114

van Hoof A, Frischmeyer PA, Dietz HC, Parker R. 2002. Exosome-mediated recognition and degradation of mrnas lacking a termination codon. *Science 295*:2262-2264. doi:10.1126/science.1067272

Viegas MH, Gehring NH, Breit S, Hentze MW, Kulozik AE. 2007. The abundance of rnps1, a protein component of the exon junction complex, can determine the variability in efficiency of the nonsense mediated decay pathway. *Nucleic Acids Res 35*:4542-4551. PMCID:PMC1935013. doi:10.1093/nar/gkm461

Wang D, Zavadil J, Martin L, Parisi F, Friedman E, Levy D, Harding H, Ron D, Gardner LB. 2011. Inhibition of nonsense-mediated rna decay by the tumor microenvironment promotes tumorigenesis. *Molecular and Cellular Biology 31*:3670-3680. doi:10.1128/MCB.05704-11

Wang L, Wessler SR. 1998. Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize r gene. *Plant Cell 10*:1733-1746. PMCID:PMC143946.

Weatheritt RJ, Sterne-Weiler T, Blencowe BJ. 2016. The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol 23*:1117-1123. doi:10.1038/nsmb.3317

Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Mönch K, Thoren LA, Nielsen FC, Jacobsen SEW, Nerlov C, Porse BT. 2008. Nmd is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes & development 22*:1381-1396. doi:10.1101/gad.468808

Wickham H. 2009. Ggplot2: Elegant graphics for data analysis. *Use R*:1-212. doi:10.1007/978-0-387-98141-3

Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Muhlemann O. 2011. Autoregulation of the nonsense-mediated mrna decay pathway in human cells. *RNA*:1-12. doi:10.1261/rna.030247.111

Zeng C, Fukunaga T, Hamada M. 2018. Identification and analysis of ribosome-associated
lncrnas using ribosome profiling data. *BMC Genomics 19*:414. PMCID:PMC5975437.
doi:10.1186/s12864-018-4765-z

Zhang J, Sun X, Qian Y, LaDuca JP, Maquat LE. 1998a. At least one intron is required for the
nonsense-mediated decay of triosephosphate isomerase mrna: A possible link between
nuclear splicing and cytoplasmic translation. *Mol Cell Biol 18*:5272-5283.

Zhang J, Sun X, Qian Y, Maquat LE. 1998b. Intron function in the nonsense-mediated decay of
beta-globin mrna: Indications that pre-mrna splicing in the nucleus can influence mrna
translation in the cytoplasm. *RNA 4*:801-815. PMCID:PMC1369660.