

1

2 ***Transcript-specific characteristics determine the***
3 ***contribution of endo- and exonucleolytic decay***
4 ***pathways during the degradation of nonsense-mediated***
5 ***decay substrates***

6

7 ***Franziska Ottens^{1#}, Volker Boehm^{1#}, Christopher R. Sibley^{2#}, Jernej Ule³, Niels H. Gehring¹***

8 ¹*Institute for Genetics, University of Cologne, 50674 Cologne, Germany*

9 ²*Division of Brain Sciences, Imperial College London,*
10 *Du Cane Road, London, W12 0NN, UK*

11 ³*Department of Molecular Neuroscience, UCL Institute of Neurology,*
12 *Queen Square, London, WC1N 3BG, UK*

13

14

15

16 **Contact**

17 Niels H. Gehring, University of Cologne, Institute for Genetics, Zulpicher Str. 47a, 50674 Cologne,
18 Germany; email: ngehring@uni-koeln.de

19

20 ***#These authors contributed equally to this work***

21

22

23

24 Running title: Differential contribution of decay pathways during NMD

25

26 Keywords: NMD, endonucleolytic cleavage, decay intermediates, degradome sequencing, SMG6,
27 SMG7

28

1 **Abstract**

2 Nonsense-mediated mRNA decay (NMD) controls gene expression by eliminating mRNAs with
3 premature or aberrant translation termination. Degradation of NMD substrates is initiated by the
4 central NMD factor UPF1, which recruits the endonuclease SMG6 and the deadenylation-promoting
5 SMG5/7 complex. The extent to which SMG5/7 and SMG6 contribute to the degradation of individual
6 substrates and their regulation by UPF1 remain elusive. Here we map transcriptome-wide sites of
7 SMG6-mediated endocleavage via 3' fragment capture and degradome sequencing. This reveals that
8 endogenous transcripts can have NMD-eliciting features at various positions, including upstream
9 open reading frames (uORFs), premature termination codons (PTCs) and long 3' UTRs. We find that
10 NMD substrates with PTCs undergo constitutive SMG6-dependent endocleavage, rather than SMG7-
11 dependent exonucleolytic decay. In contrast, the turnover of NMD substrates containing uORFs and
12 long 3' UTRs involves both, SMG6- and SMG7-dependent endo- and exonucleolytic decay,
13 respectively. This suggests that the extent to which SMG6 and SMG7 degrade NMD substrates is
14 determined by the mRNA architecture.

15

1 **Introduction**

2 Nonsense-mediated mRNA decay (NMD) is a eukaryotic mRNA surveillance mechanism that
3 maintains the fidelity of gene expression by targeting aberrant transcripts. NMD degrades transcripts
4 containing premature termination codons (PTCs) and thereby prevents the synthesis of C-terminally
5 truncated proteins with potentially unphysiological or deleterious functions (Holbrook et al. 2004;
6 Chang et al. 2007; Nicholson et al. 2010). NMD also regulates many cellular mRNAs, which contain
7 one or more NMD-activating features, for example long 3' UTRs or upstream open reading frames
8 (uORFs). Thereby, NMD alters the expression levels of ~5% of the transcriptome in eukaryotes
9 (Kervestin and Jacobson 2012).

10 In mammalian cells, NMD is activated at termination codons that are located >50 nucleotides (nt)
11 upstream of at least one downstream exon-exon junction that is marked by an exon-junction
12 complex (EJC) (Nagy and Maquat 1998; Thermann et al. 1998; Neu-Yilik et al. 2001). In addition, NMD
13 targets mRNAs with an aberrant architecture downstream of the termination codon, such as
14 unusually long 3' UTRs (Muhlrad and Parker 1999; Amrani et al. 2004; Buhler et al. 2006; Eberle et al.
15 2008; Singh et al. 2008). This EJC-independent NMD is activated in response to improper termination
16 by the ribosome, for example, when the interaction of the cytoplasmic poly(A)-binding protein
17 (PABPC1) with the eukaryotic release factors (eRFs) is impaired (Amrani et al. 2004; Eberle et al.
18 2008; Singh et al. 2008; Fatscher et al. 2014).

19 Activation of NMD by aberrant translation termination triggers the assembly of a surveillance
20 complex onto the mRNA, consisting of the proteins UPF1, UPF2, and UPF3 (UPF3a and UPF3b in
21 humans). During the initial phases of NMD, the phosphoinositide 3-kinase related kinase SMG1
22 phosphorylates the helicase UPF1 within its extended N- and C-terminal regions (Kashima et al.
23 2006), which function as binding sites for SMG6 and the SMG5/7 heterodimer (Okada-Katsuhata et
24 al. 2012). It has been reported that SMG5, SMG6, and SMG7 interact preferentially with
25 phosphorylated UPF1 via their 14-3-3-like domains (Fukuhara et al. 2005; Okada-Katsuhata et al.

1 2012; Kurosaki et al. 2014). However, a phosphorylation-independent interaction between an
2 unstructured region of SMG6 and UPF1 has been recently identified (Chakrabarti et al. 2014;
3 Nicholson et al. 2014).

4 The degradation of NMD targets occurs via independent pathways that either involve
5 endonucleolytic cleavage or deadenylation and decapping (Muhlemann and Lykke-Andersen 2010).
6 Deadenylation is executed by the SMG5/7 heterodimer, which directly interacts with POP2, a
7 catalytic subunit of the CCR4–NOT deadenylase complex (Loh et al. 2013). Furthermore,
8 deadenylation-independent decapping of NMD substrates is stimulated by interactions of NMD
9 factors with decapping enzymes (Cho et al. 2009; Loh et al. 2013).

10 The endonucleolytic cleavage of NMD targets is mediated by the C-terminal PIN (PiIT N terminus)
11 domain of SMG6 (Glavan et al. 2006; Huntzinger et al. 2008; Eberle et al. 2009). Endocleavage occurs
12 in the vicinity of the termination codon and the 5' and 3' mRNA fragments are exonucleolytically
13 degraded by the exosome and XRN1, respectively (Gatfield and Izaurralde 2004; Huntzinger et al.
14 2008; Eberle et al. 2009). Hence, 3' decay intermediates are enriched in cells depleted of XRN1,
15 which enables their amplification and mapping by high-throughput sequencing approaches (Boehm
16 et al. 2014; Lykke-Andersen et al. 2014; Schmidt et al. 2015).

17 The existence of several degradation pathways used by the NMD machinery raises the question
18 how individual NMD targets are selected for a particular degradation pathway. It has been suggested
19 that NMD substrates are preferentially degraded via SMG6-mediated endonucleolytic cleavage
20 (Gatfield and Izaurralde 2004). SMG5/7-dependent deadenylation and decapping activity was found
21 to be elevated when SMG6 is inactive. However, for complete inhibition of the NMD machinery both
22 degradation routes have to be inactivated (Jonas et al. 2013; Lykke-Andersen et al. 2014; Schmidt et
23 al. 2015). Hence, alternative degradation pathways are currently considered to ensure efficient
24 degradation of NMD target mRNAs.

1 In this work, we map the sites of endocleavage and the NMD-eliciting features of many
2 endogenous NMD substrates via degradome sequencing. We find that the endocleavage efficiency of
3 NMD substrates with uORFs and long 3' UTRs is inversely regulated by the activity of the SMG7-
4 dependent exonucleolytic decay pathway. Interestingly, our results suggest a differential
5 contribution of SMG6- and SMG7-dependent decay routes towards the degradation of different
6 NMD target classes.

1 **Results**

2 ***Global identification of endogenous 3' fragments***

3 To identify mRNAs cleaved by SMG6 in a transcriptome-wide manner, we adapted a method
4 to capture and amplify polyadenylated RNAs with a 5' monophosphate (Fig. 1A) (Addo-Quaye et al.
5 2008; German et al. 2008; Gregory et al. 2008). We generated 3' fragment libraries with RNA isolated
6 from HeLa cells that were transfected with siRNA against XRN1 or XRN1 and SMG6, respectively
7 (Supplemental Fig. S1A). Comparison of these two conditions allows the identification of XRN1-
8 sensitive, SMG6-dependent cleavage products (Supplemental Fig. S1B). In addition, to determine the
9 background of 5' monophosphate RNAs, which are not degraded by XRN1, the same protocol was
10 carried out on cells transfected with a negative control siRNA directed against Luciferase (Luc).

11 High-throughput sequencing and bioinformatics analysis was carried out for two replicate
12 experiments, providing a total of 14,458,242 and 11,573,210 mapped reads for the siXRN1 and
13 siXRN1/siSMG6 conditions, respectively (Supplemental Fig. S1C). The lowest percentage of mapped
14 reads was detected in the control condition (Luc knockdown), as in this case only a small number of
15 3' fragments is stable and can be captured by the transcriptome-wide mapping approach. The overall
16 number of mapped reads corresponded to 149,346 low FDR degradation sites (see methods), with
17 replicate data sets revealing the most consistent degradation locations and normalized frequencies
18 for each site within conditions (Supplemental Fig. S1D). This demonstrates the high reproducibility of
19 the quantitative information contained within the degradome sequencing data. An average read
20 length of 27 nt (Supplemental Fig. S1E) confirms that we specifically amplified EcoP15I-digested
21 fragments. Annotation of genomic features revealed a strong enrichment of degradome sequencing
22 fragments within processed mRNAs (Supplemental Fig. S1F), as SMG6-dependent reads preferentially
23 map within the coding sequence (CDS) and 5' and 3' UTR, but not within introns or intergenic regions.
24 Interestingly, in the library from XRN1-depleted cells alone, we observed a strong peak at the ORF - 3'
25 UTR, but not 3' UTR - intergenic border, demonstrating a preference for SMG6 endocleavage at the

1 termination codon (Fig. 1B). In contrast, cleavage is enriched across the whole 5' UTR, but without
2 any positional preference. This could reflect cleavage in uORFs that can occur at various positions
3 within 5' UTRs of different genes, or generally indicate high cleavage across 5' UTRs from other
4 unknown phenomena.

5 To separate genes that had XRN1-sensitive, SMG6-dependent cleavage products associated with
6 different features, we used a sliding window approach to identify 100 nt segments across the
7 transcriptome that are most enriched in "XRN1 siRNA" reads relative to "XRN1/SMG6 siRNA" reads
8 (Supplemental Table S1). To assign genes to a feature (5' UTR, ORF, 3' UTR) we used a strict threshold
9 of >50 reads in the window for the XRN1 siRNA condition and a fold-difference of >10 between the
10 "XRN1 siRNA" and "XRN1/SMG6 siRNA" conditions. The overlapping feature (5' UTR, ORF, 3' UTR) of
11 the window with the highest z-score in each mRNA was then assigned to the corresponding gene. In
12 all identified SMG6-sensitive genes, the highest-scoring window was found within the following
13 features; 1079 in ORFs, 271 in 5' UTRs, 118 in 3' UTRs. We additionally assessed how many of these
14 identified target windows were cleaved in proximity to an annotated termination codon via the
15 positional evaluation of each window's highest peak. This analysis determined that 19% of 3' UTR
16 sites were within 50 nt of an annotated termination codon, in contrast to 11% for ORF windows and
17 0% for 5' UTR windows (Supplemental Fig. S1G). In agreement with genome-wide maps at the ORF-3'
18 UTR boundary (Fig. 1B), there was preference for max peaks to be located at or after the termination
19 codons (Supplemental Fig. S1H).

20 To further investigate the previously observed overall enrichment of SMG6-dependent cleavage sites
21 in the 5' UTR (Fig. 1B), we probed uORF-containing mRNAs for SMG6-specific cleavage events.
22 Analysis of all uORF-containing mRNAs showed no enrichment of reads (Fig. 1C) (Fritsch et al. 2012).
23 This could be due to the described NMD-escaping behavior of uORFs (Stockklauser et al. 2006) or
24 alternatively due to technical limitations associated with picking up 5' UTR degradation products with
25 a poly(A) enrichment method.

1 ***Implication of SMG6-specific cleavage events for different classes of*** 2 ***endogenous NMD targets***

3 Endogenous NMD targets can be classified by one or more NMD activating features, such as
4 long 3' UTRs, upstream open reading frames (uORFs), PTCs resulting from alternative or aberrant
5 splicing events or the presence of a selenocysteine codon (UGA). For each of these classes, we
6 identified representative targets during our transcriptome-wide analysis of decay intermediates. The
7 results from high-throughput target mapping also allowed us to obtain endocleavage patterns at
8 single nucleotide resolution. Previous studies have defined preferential sites of endocleavage at and
9 downstream of termination codons (Eberle et al. 2009; Boehm et al. 2014). For selected transcripts,
10 we find SMG6-dependent 3' fragments mapping to positions at and downstream of termination
11 codons located upstream of long 3' UTRs (Fig. 2A), selenocysteine codons (Fig. 2C) or out of frame
12 stop codons (Fig. 2D). Although we were not able to detect endocleavage at uORFs on a
13 transcriptome-wide scale (Fig. 1C), individual examples of SMG6 activity around uORF stop codons
14 could readily be detected (Fig. 2B). Hence, we reproduced the previously described pattern of SMG6-
15 mediated cleavage of mRNAs in the vicinity of the NMD-activating termination codon for
16 endogenous NMD substrates (Lykke-Andersen et al. 2014; Schmidt et al. 2015). Interestingly, some
17 targets (C11orf31 and TMEM222) showed an additional enrichment of mapped reads upstream of
18 the stop codon, but the relevance of this site remains to be determined (Fig. 2C and D). In order to
19 validate that these mRNAs are *bona fide* NMD targets, we used a semi-quantitative PCR-based
20 approach to visualize 3' decay intermediates. This PCR approach confirmed SMG6-dependent
21 endocleavage of PDRG1, MED10 and SURF6 (long 3' UTR, Fig. 2A), BAG1 and IFRD1 (uORF, Fig. 2B)
22 and C11orf31 (selenocysteine codon, Fig. 2C) in the close proximity of the termination codon. The
23 analysis of RPL10A and TMEM222 revealed SMG6-dependent 3' fragments in an annotated, intronic
24 region (RPL10A) or in an alternative exon with an in-frame stop codon (TMEM222), respectively (Fig.
25 2D). Endonucleolytic decay of all analyzed mRNAs was strictly UPF1-dependent, confirming that
26 endocleavage occurred during NMD.

1 ***Endocleavage efficiency of several NMD targets is regulated by SMG7 levels***

2 To study the regulation of NMD substrates, we constructed triosephosphate isomerase (TPI)
3 reporters with 3' UTRs of two endogenous targets, PDRG1 and SURF6, downstream of the TPI stop
4 codon (Supplemental Fig. S2A). Other reporter constructs containing the shortened SMG5 3' UTR, the
5 UPF3b 3' UTR and the GFP coding sequence were characterized before and shown to elicit SMG6-
6 dependent endocleavage (Boehm et al, 2014). We confirmed in addition for the PDRG1 and SURF6 3'
7 UTR-containing reporters that they undergo SMG6- and UPF1-mediated degradation and
8 endocleavage (Supplemental Fig. S2B).

9 Two pathways are involved in the degradation of NMD targets, SMG5/7-dependent
10 deadenylation and decapping (Loh et al. 2013) as well as endonucleolytic degradation via SMG6
11 (Glavan et al. 2006; Huntzinger et al. 2008; Eberle et al. 2009). Recent high-throughput analyses
12 showed an increase in decapped mRNAs upon XRN1 and SMG6 co-depletion, which was attributed to
13 enhanced deadenylation-mediated by SMG7 (Lykke-Andersen et al. 2014; Schmidt et al. 2015).
14 Therefore, we aimed to investigate if SMG7 expression levels influence the efficiency of SMG6-
15 dependent endocleavage. In XRN1 and SMG7 co-depleted cells expressing a TPI-SURF6 reporter
16 construct we readily detected 3' fragments, which were reduced when siRNA-resistant SMG7 was
17 expressed (Fig. 3A). It is conceivable that 3' decay intermediates are stabilized due to decreased
18 deadenylation activity upon SMG7 knockdown. To this end, we co-depleted XRN1 and CNOT1, a
19 crucial scaffolding component of the CCR4-NOT deadenylase complex. We detected comparable 3'
20 fragment levels in XRN1 and XRN1/CNOT1 depletion conditions, indicating that the 3' fragments are
21 not directly degraded via deadenylation. Furthermore, monitoring 3' fragment stability after
22 transcriptional shut-off of stable cell lines expressing the TPI-SURF6 reporter showed that the block
23 of deadenylase activity (XRN1/CNOT1 knockdown) did not alter the degradation rate of 3' decay
24 intermediates (Supplemental Fig. S2C and D). The influence of SMG7 on 3' fragment levels was also
25 observed using TPI constructs with a SMG5, UPF3b or GFP 3' UTR, which behaved similarly to the TPI-
26 SURF6 construct (compare Fig. 3A and Supplemental Fig. S2E). The inverse correlation of 3' fragments

1 with SMG7 expression levels was also detected by semi-quantitative PCR for many of the
2 endogenous NMD substrates. However, the 3' fragments of both targets with PTCs, incorporated due
3 to alternative splicing (TMEM222 and RPL10A), were not affected by SMG7 levels (Fig. 3B).
4 Additionally, we performed the PCR assay with cDNA that was generated by a poly(A)-independent
5 method (Supplemental Fig. S2F). This showed that the observed effect on 3' fragment abundance
6 indeed results from differential SMG7 expression levels and not from enhanced or impaired
7 deadenylation of the 3' endocleavage products. Taken together, these observations indicate a
8 differential mode of degradation of target mRNAs, depending on their NMD-inducing features. While
9 PTC-containing RNAs mainly undergo efficient endocleavage, the decay of transcripts with uORFs,
10 Sec codons and long 3' UTRs is regulated by both, SMG6-mediated endocleavage and SMG7-induced
11 deadenylation/decapping.

12 ***Endonucleolytic cleavage is not saturated by increasing levels of NMD*** 13 ***substrates***

14 Our results suggest substantial differences in the requirement of NMD factors and the
15 mechanism of degradation for different endogenous and transfected NMD substrates. Two
16 redundant degradation pathways may act as a failsafe mechanism to ensure efficient degradation of
17 target mRNAs even if the integrity of one of the pathways is compromised. We speculate that such a
18 failsafe mechanism may be relevant for cells that produce large amounts of NMD substrates and
19 increasing levels of NMD substrates may activate the alternative SMG7-dependent NMD pathway. To
20 this end, we transfected variable amounts of plasmids expressing long 3' UTR- or PTC-containing
21 mRNAs (Fig. 4A and B). We observed no change of endocleavage efficiency with increasing amounts
22 of NMD substrates, suggesting that the NMD machinery and specifically SMG6 is able to cope with
23 large quantities of target mRNA.

24 We also investigated whether increasing amounts of PTC-containing reporter mRNA, which is
25 cleaved efficiently, can influence the degradation of constant amounts of long 3' UTR mRNA, which is

1 on average a less efficient SMG6 substrate (Boehm et al. 2014). However, we observed neither an
2 increase in reporter levels nor a decrease in 3' fragments generated from the TPI-SMG5 reporter
3 when the levels of TPI-PTC160 reporter were increased (Fig. 4C). Interestingly, overexpression of
4 SMG6 WT increased endocleavage efficiency when TPI-SMG5, but not TPI-PTC160, was used as
5 reporter mRNA (Fig. 4). This observation further indicates a differential contribution of SMG6- and
6 SMG7-dependent decay routes towards the degradation of different NMD target classes.

7 ***SMG7-mediated regulation of endocleavage depends on functional UPF1*** 8 ***binding***

9 To better understand the interplay between SMG6 and SMG7, we performed overexpression
10 studies using different SMG7 constructs (Fig. 5A). Specifically, we asked whether SMG7 regulates
11 endocleavage via its interaction with UPF1 or by a UPF1-independent mechanism. To this end, we
12 compared the effects of full length wildtype SMG7, full length SMG7 with a mutation in the 14-3-3-
13 like domain (14-3-3^{Mut}; unable to interact with UPF1) and a C-terminally truncated SMG7 variant (1-
14 497) (Fukuhara et al. 2005). SMG7 (1-497) consists only of the 14-3-3-like and helical domain and
15 therefore lacks the region required for interaction with the deadenylase POP2 (Loh et al. 2013). The
16 overexpression of wildtype SMG7 (SMG7 WT) or 1-497 resulted in reduced 3' fragment generation of
17 TPI-SMG5, whereas overexpression of SMG7 14-3-3^{Mut} did not alter endocleavage activity (Fig. 5B).
18 Notably, elevated reporter mRNA levels were observed for SMG7 1-497, which is not able to activate
19 substrate deadenylation (compare lane 1 and 4). These results suggest that some NMD substrates
20 may be degraded either by the SMG6- or SMG7-dependent pathways and the cellular availability of
21 SMG6 and SMG7 therefore determines, which decay mechanism prevails. Interestingly, we did not
22 detect reduced endocleavage of a PTC reporter mRNA (TPI-PTC160) (Fig. 5C) when SMG7 was
23 overexpressed, indicating that SMG7 does not influence SMG6-mediated endocleavage of this
24 substrate (Fig. 5D).

1 ***The UPF1 C-terminal extension is required for endocleavage of long 3' UTR***
2 ***targets***

3 We next focused on UPF1, which represents a central key molecule of the NMD machinery and
4 serves as a binding platform for different decay factors. The kinase SMG1 phosphorylates UPF1 at SQ
5 and TQ motifs in the unstructured N- and C-terminal regions in order to recruit SMG5, SMG6 and
6 SMG7 (Yamashita et al. 2001; Chakrabarti et al. 2014; Kurosaki et al. 2014). Previously, T28 in the
7 UPF1 N-terminus was identified as the binding site for SMG6, whereas the phosphorylated UPF1 C-
8 terminus mediates the interaction with the SMG5/7 heterodimer (Okada-Katsuhata et al. 2012). As
9 revealed by SMG7 overexpression, the interaction of overexpressed SMG7 with UPF1 reduced
10 endocleavage of long 3' UTR targets (Fig. 5B). We therefore studied the role of the SMG7 binding
11 sites (C-terminal extension of UPF1) during endocleavage by analyzing truncated UPF1 proteins in
12 complementation assays (Fig. 6A). Both, deletion of the complete C-terminus (N-A2) as well as the
13 very C-terminal cluster of SQ/TQ motifs (N-1030), which retains the helicase regulating SQ region of
14 UPF1 (Fiorini et al. 2013), increased TPI-PTC160 reporter mRNA levels in complementation assays
15 (Fig. 6B). Although the endocleavage efficiency was reduced due to the accumulation of reporter
16 mRNA, the total amount of 3' fragments changed only moderately, indicating that endocleavage is
17 maintained in spite of decreased NMD activity. In contrast, deletions of the C-terminus strongly
18 reduced endocleavage of TPI-SMG5 mRNA, suggesting that the truncated UPF1 proteins are inactive
19 on NMD substrates with long 3' UTRs (Fig. 6C). This finding is in accordance with the results of the
20 SMG7 overexpression assay (Fig. 5) and thereby further strengthens the view that the combined
21 activity of SMG6 and SMG7 is implicated in the degradation of non-PTC NMD targets and that the
22 mode of degradation depends on the mRNA architecture and the availability of SMG6 and SMG7.

23

1 **Discussion**

2 NMD of many substrate mRNAs is initiated by endonucleolytic cleavage and SMG6 represents an
3 important NMD factor in human cells. In this work, we obtained insights into the molecular
4 mechanism of endonucleolytic cleavage by SMG6 and its regulation in the context of NMD.
5 Furthermore, we identified endogenous NMD targets on a transcriptome-wide scale and mapped
6 sites of SMG6 endocleavage at nucleotide resolution. This analysis confirmed that endocleavage of
7 NMD substrates occurs mainly at and downstream of termination codons. In fact, our unbiased
8 approach enabled us to identify termination codons that are recognized as aberrant and to
9 discriminate between NMD targets with uORFs, long 3' UTRs, selenocysteine codons and those
10 produced by alternative splicing. Interestingly, the “degradative” NMD of some transcripts is
11 dominated by rapid endocleavage, while “regulatory” NMD of other mRNAs uses endo- and
12 exonucleolytic decay pathways. Hence, the identification of two partially redundant degradation
13 pathways is an important aspect of this work. We note that our method would in principle allow the
14 detection of transcripts that are targeted for degradation due to transcription or splicing errors.
15 However, this biologically important class of NMD substrates, which escapes quality control during
16 production or processing, is currently not recorded because of the low abundance of individual
17 transcripts. Hence, it will be important to further improve the methodology to identify the whole
18 spectrum of SMG6 substrates in human cells.

19 We observe that some of the endocleavage sites we identify here overlap with previously
20 reported sites (Schmidt et al. 2015). In contrast, there seems to be little overlap with another study,
21 in which HEK 293 cells were used (Lykke-Andersen et al. 2014). Although this variation could be due
22 to a different sequencing technique that was used, it may also indicate that certain NMD substrates
23 are cell-type specific and others ubiquitous. A comparison of NMD substrates of different cells
24 and/or organisms will be required to identify a core set of NMD targets.

1 Recently, an increased decapping of NMD substrates was reported when SMG6, but not UPF1,
2 was depleted (Lykke-Andersen et al. 2014; Schmidt et al. 2015). This suggested that mRNAs, which
3 are normally degraded by SMG6, can be targeted by an alternative decay pathway involving SMG5/7-
4 mediated deadenylation and decapping. This observation is consistent with reports that NMD is
5 strongly inhibited by the combined depletion of SMG6 with either SMG5 or SMG7 (Jonas et al. 2013).
6 This raises the question how individual NMD substrates are assigned to different degradation
7 pathways, i.e. what determines whether an mRNA is degraded preferentially by endocleavage or
8 deadenylation and decapping. To address this question, we altered the expression levels of SMG7 by
9 overexpression or depletion in cultured cells. We chose SMG7, because it interacts directly with
10 phosphorylated UPF1 and several SNPs have been found to be associated with differences in the
11 expression of SMG7 (Nguyen et al. 2014). Hence, SMG7 may show highly diverse expression levels in
12 different individuals. Interestingly, the efficiency of endocleavage of a substrate mRNA with a long 3'
13 UTR inversely correlated with the expression levels of SMG7. This suggests that the degradation of
14 these NMD substrates uses both, endo- and exonucleolytic decay and their contribution depends on
15 the levels of SMG6 and SMG7, respectively. Therefore, the complete lack of a SMG7 gene in the
16 *Drosophila melanogaster* genome may explain why endocleavage of NMD targets has been originally
17 identified in cultured cells from *D. melanogaster* (Gatfield & Izaurralde, 2004).

18 Moreover, we find that different classes of endogenous NMD substrates show different levels of
19 partitioning between SMG6- and SMG7-dependent decay. Transcripts containing uORFs, long 3' UTRs
20 or selenocysteine codons undergo more efficient endocleavage when SMG7 levels are low, while
21 endocleavage efficiency of PTC-containing transcripts is not affected by SMG7 abundance. This
22 suggests a previously unrecognized complexity of NMD and indicates that endogenous targets are
23 either rapidly degraded by the SMG6-mediated endonucleolytic degradation, or slowly degraded
24 ("downregulated") by the weaker activity of SMG6, and the slower SMG7-dependent decay (Fig. 7).
25 The difference between these degradative and regulatory effects of NMD is consistent with the
26 different likelihood of NMD targets to encode functional proteins. PTCs are most often caused by

1 mutations or aberrant splicing in the ORF that can produce aberrant proteins. SnoRNA host
2 transcripts are also rapidly degraded by the SMG6-mediated endocleavage, and their mature mRNAs
3 that result from splicing-coupled snoRNA production do not encode functional proteins (Lykke-
4 Andersen et al. 2014). In contrast, mRNAs with uORFs or long 3' UTRs are most often canonical
5 transcript isoforms that encode functional proteins. It is thus appropriate that functional mRNAs are
6 not efficiently degraded by the SMG6-mediated endocleavage, and are instead more amenable to
7 the slower degradation by SMG7. Although we clearly observe differences between the mRNA
8 targets of “degradative” and “regulatory” NMD pathways, the molecular characteristics of the two
9 classes have not been explicitly determined and need to be investigated in the future. However, we
10 speculate that NMD caused by defective ribosome recycling may be a common feature of regulated
11 substrates (Fatscher et al. 2014; Joncourt et al. 2014).

12 The existence of two partially redundant degradation pathways might be required when cells
13 need to cope with increased amounts of NMD targets. This could occur when other quality control
14 systems are impaired or when cells are exposed to highly mutagenic conditions. However, our
15 experimental data (Fig. 4) demonstrate that SMG6 catalyzes endocleavage of overexpressed NMD
16 substrates as efficiently as small amounts thereof. Hence, we suggest that alternative degradation
17 mechanisms have not been implemented to process an excess of NMD substrates, but to regulate
18 NMD in a transcript-specific manner (Fig. 7). Inherent features of the transcript, such as mRNA
19 architecture, determine whether the NMD machinery executes a degradative or downregulating
20 function. While NMD has been originally considered as a mechanism to degrade faulty mRNAs, our
21 results underline its growing importance as a master regulator of gene expression (Nicholson et al.
22 2010; Karam et al. 2013; Lykke-Andersen and Jensen 2015).

23

1 ***Material and Methods***

2 ***Plasmids and cell culture***

3 Plasmid constructs β -globin, LacZ-4H, pCI-FLAG, pCI-mVenus, WT and PTC-containing TPI and
4 expression vectors for UPF1 and SMG6 were described previously (Gehring et al. 2009; Boehm et al.
5 2014). SMG7 was cloned from HeLa cDNA and mutated to confer siRNA insensitivity (targeting
6 sequence 5'- CGATTTGGAATACGCTTTA -3' replaced by 5'- TGACCTTGAGTATGCCCTG -3'). Point or
7 deletion mutants of UPF1 and SMG7 were generated by site-directed mutagenesis, cloned in the
8 designated vector and verified by sequencing. Standard protocols were used to generate the stable
9 HeLa Flp-In T-REx TPI-SURF6 cell line (HeLa Flp-In T-REx initially established by Elena Dobrikova and
10 Matthias Gromeier, Duke University Medical Center). Expression of stable cell lines was induced with
11 1 μ g/ml doxycycline for 24 h. All cell lines were cultured in DMEM (Gibco) supplemented with 9%
12 fetal bovine serum (FBS; Gibco) and 1x Pen Strep (Gibco) and the cells were incubated at 37 °C, 5%
13 CO₂ and 90% humidity.

14 ***siRNA transfections***

15 5x10⁵ HeLa Tet-Off cells (Clontech) were grown over night in 6 cm plates and transiently
16 transfected with 200 pmol siRNA for single or 400 pmol total siRNA for double knockdowns using
17 Lipofectamine RNAiMAX (Life Technologies). 24 h post transfection the cells were split 1:2 in 10 cm
18 plates and the day after transfected again with 600 pmol (single knockdown) or 1200 pmol (double
19 knockdown) siRNA. For triple knockdown, 400 pmol of the single siRNAs were used (1200 pmol in
20 total). For library preparation and verification, 1x10⁶ HeLa cells were cultured overnight in 10 cm
21 dishes and transiently transfected with 600 pmol total siRNA following the protocol for RNAiMAX
22 (Life Technologies). 24h after transfection, cells were split 1:8 in 10 cm dishes. 24 h later, cells again
23 were transfected with 600 pmol (300 pmol each for double knockdowns) siRNA and incubated for 48
24 h. Stable cell lines were reverse transfected using 2.5 μ l Lipofectamine RNAiMAX and 60 pmol siRNA
25 per 2x10⁵ HeLa cells. The following siRNA target sequences were used for luciferase 5'-

1 CGTACGCGGAATACTTCGA-3', for XRN1 5'-AGATGAACTTACCGTAGAA-3', for SMG6 5'-
2 GGGTCACAGTGCTGAAGTA-3', for UPF1 5'-GATGCAGTCCGCTCCATT-3', for CNOT1 5'-
3 GGAACUUGUUUGAAGAAUA-3' and for SMG7 5'-CGATTTGGAATACGCTTTA-3'.

4 ***Plasmid transfections***

5 HeLa cells were split to 6-well plates the day after siRNA transfection and transfected by calcium
6 phosphate precipitation with 0.5 µg of a mVenus expression plasmid, 2 µg control plasmid (LacZ-4H)
7 and 1.5 µg plasmid encoding for TPI-PTC160 reporter mRNA. For long 3' UTR reporters, 3 µg reporter
8 and 0.75 µg control plasmid (β-globin) were transfected. For rescue assays, 1 µg of FLAG-tagged
9 expression plasmid was included in the transfection mix.

10 ***RNA extraction and northern blotting***

11 Total RNA was extracted with peqGOLD TriFast (Peqlab), resolved on a 1% agarose and 0.4 M
12 formaldehyde gel using a tricine-triethanolamine buffer system and analyzed by northern blotting.
13 pSP65-globin plasmid was linearized with BamHI and used as template for in-vitro transcribed [α-
14 ³²P]-GTP body-labeled RNA probes, which were used for the detection of all reporter and control
15 RNA. 7SL endogenous RNA was detected using a 5'-³²P-labeled oligonucleotide (5'-
16 TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT-3'). Signals were scanned using a Typhoon FLA
17 7000 (GE Healthcare). For time-course assays, the stable HeLa Flp-In T-REx cells were treated with 5
18 µg/ml actinomycin D for the indicated time prior to harvesting.

1 ***Immunoblot analysis and antibodies***

2 SDS-polyacrylamide gel electrophoresis and immunoblot analysis was performed using protein
3 samples derived from TriFast extractions. The antibodies against tubulin (T6074) and FLAG (F7425)
4 were from Sigma, the antibody against SMG6 (ab87539) was from Abcam, the antibodies against
5 XRN1 (A300-443A) and SMG7 (A302-170A) were from Bethyl and the antibody against UPF1 was
6 kindly provided by Jens Lykke-Andersen.

7 ***3' fragment library preparation***

8 siRNA treated cells were used for poly(A)⁺ RNA isolation utilizing the magnetic mRNA isolation kit
9 (New England Biolabs). Two hundred pmol of 5' RNA linker, containing a recognition site for EcoP15I
10 (5' linker_EcoP15I) was ligated to 3 µg poly(A)⁺ RNA using T4 RNA Ligase I. After DNase I digest and
11 RNA purification (RNA Clean & Concentrator Kit, Zymo Research) RT-PCR was performed using a
12 linker oligo(dT)₁₈V (Harigaya and Parker 2012). The cDNA was PCR amplified for 9 cycles, digested
13 with EcoP15I overnight which cleaves 25 to 27 nt downstream of the recognition site, gel-purified
14 and ligated to a 3' dsDNA adaptor. After column purification of the ligation reaction, the ligated
15 material was PCR-amplified for subsequent Illumina sequencing, using P3 and P5 primers and
16 Accuprime Taq Polymerase (Life Technologies). For oligonucleotide sequences see Supplemental
17 Table S2.

18 ***Library processing***

19 Single-end sequencing reads had their barcodes and adapter trimmed with fastx_clipper from
20 the fastx_toolkit (v. 0.0.13) before being mapped to the human genome (hg19) using tophat (v.
21 2.0.11) and known splice junctions from ENSEMBL annotated transcripts using the following settings;
22 -g 1 -p 8 -library-type fr-secondstrand. Sorted bam files were converted to bed files with the
23 bedtools bamtobed command (v. 2.16.2), before single nucleotide resolution cleavage sites were
24 determined using awk (v. 3.1.5) and sed (v. 4.1.5). Specifically, based on the protocols design, the site

1 of cleavage was considered the nucleotide that was antisense to the nucleotide immediately
2 following the 3' end of the read. For evaluation of replicate consistency, we initially applied a peak-
3 finding algorithm that identified clusters of degradation sites with significant enrichment of
4 degradation events relative to the local environment (Konig et al. 2010; Zarnack et al. 2013). We
5 used all replicates from all conditions for this analysis, and a flank size of 15 nt on either side to
6 define significant clusters with FDR <5%. This yielded a total of 299,442 degradation sites. We then
7 evaluated the number of degradation events mapping to each degradation site for each replicate in
8 order to compare how individual libraries contributed to each defined degradation cluster.

9 ***RNA maps***

10 Maps were created through intersection of single nucleotide resolution cleavage files with
11 indicated features using the bedtools intersect command. Features were determined based on their
12 positioning in annotation files used for library mapping with the exception of uORFs (see below).
13 Counts at individual loci were normalized to the corresponding libraries total read count. Similarly,
14 for each junction type, counts were divided by the maximum value across all libraries to allow for the
15 comparison across different features. Maps were then created from normalized values around
16 indicated features using R (v. 3.1.1). Enrichment analysis of the different features of the
17 transcriptome was performed using the Homer annotatePeaks.pl command (Heinz et al. 2010),
18 utilizing the same annotation file as that used for mapping.

19 Separation into 5' UTR, 3' UTR and ORF categories was achieved by first applying a 100 nt sliding
20 window across the transcriptome in order to identify windows most enriched between the siRNA
21 XRN1 and siXRN1/SMG6 conditions. To then assign genes to a given feature we applied a strict
22 threshold of >50 reads in the window for the XRN1 siRNA condition, a fold-difference of >10 between
23 the XRN1 siRNA and XRN1/SMG6 siRNA conditions, and, using a ranked list of z-scores, we took the
24 first occurrence of each gene and assigned it to the feature in which that window overlapped.
25 Thereby windows were assigned based on their dominant overlap with genomic features. In cases

1 where windows overlapped features equally, the hierarchical order for assignment is "ncRNA; ORF; 3'
2 UTR; 5' UTR; intron; telomere; intergenic". To then determine proximity of cleavage sites to
3 annotated stop codons, we intersected the coordinates of all annotated stop codons with 50 nt
4 windows surrounding the highest peak in each SMG6-sensitive window. The highest peak was used
5 under the assumption that it represented the dominant cleavage site within the indicated window.

6 Coordinates of experimentally verified uORFs were taken from ribosome footprinting
7 experiments (Fritsch et al. 2012). This included 2107 unique uORF start and end coordinate pairs,
8 although starts and ends could be shared between different uORFs for the same gene. Using these
9 sites, we determined 50 nt windows surrounding all unique uORF end coordinates and intersected
10 these with single nucleotide resolution cleavage files. We removed all windows which had 0 counts in
11 the XRN1-sensitive condition prior to drawing maps.

12 ***PCR-based analysis of potential NMD targets***

13 Two hundred pmol of a 5' RNA linker (5' RNA linker_VB2) were ligated to 3 µg of poly(A)⁺ RNA
14 or 10 µg of total RNA, respectively. Reverse transcription was performed using an oligo(dT)₂₀VNN
15 primer or random hexamers. To functionally analyze NMD targets identified from the high-
16 throughput sequencing data, PCR was carried out using a primer directed against the 5' linker
17 (VB2_linker_se) and a target-specific antisense primer. Total amounts of target RNA were
18 determined using a gene-specific sense primer located downstream of the estimated endocleavage
19 site. For all targets, PCR fragments were verified via Sanger sequencing. All used oligonucleotides are
20 listed in Supplemental Table S2.

1 ***Supplemental Material***

2 Supplemental material is available for this article.

3 ***Acknowledgements***

4 We thank Heidi Thelen and Juliane Hancke for technical assistance, the Leptin and Schnetz lab for
5 sharing equipment and members of the Gehring lab for useful discussions. We are grateful to Jens
6 Lykke-Andersen for antibodies against UPF1. We thank Tomaž Curk and Nejc Haberman for
7 bioinformatics advice and assistance. This research was funded by a grant from the Deutsche
8 Forschungsgemeinschaft (GE2014/4-1) to N.H.G.

9 ***Author Contributions***

10 N.H.G, F.O. and V.B. designed the study. V.B. and F.O performed the experiments. C.S. and J.U.
11 carried out high-throughput sequencing. C.S. and J.U. analyzed and interpreted the resulting data,
12 V.B., F.O. and N.H.G. analyzed and interpreted all other data. All authors wrote the manuscript,
13 discussed the results and implications and commented on the manuscript at all stages.

14

1 **References**

- 2 Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. 2008. Endogenous siRNA and miRNA targets
3 identified by sequencing of the Arabidopsis degradome. *Curr Biol* **18**: 758-762.
- 4 Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. 2004. A faux 3'-UTR
5 promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**: 112-118.
- 6 Boehm V, Haberman N, Ottens F, Ule J, Gehring NH. 2014. 3' UTR length and messenger
7 ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep*
8 **9**: 555-568.
- 9 Buhler M, Steiner S, Mohn F, Paillusson A, Muhlemann O. 2006. EJC-independent degradation of
10 nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat Struct Mol Biol* **13**: 462-464.
- 11 Chakrabarti S, Bonneau F, Schussler S, Eppinger E, Conti E. 2014. Phospho-dependent and
12 phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and
13 SMG6. *Nucleic Acids Res* **42**: 9447-9460.
- 14 Chang YF, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance
15 pathway. *Annu Rev Biochem* **76**: 51-74.
- 16 Cho H, Kim KM, Kim YK. 2009. Human proline-rich nuclear receptor coregulatory protein 2
17 mediates an interaction between mRNA surveillance machinery and decapping complex. *Mol Cell* **33**:
18 75-86.
- 19 Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH. 2009. SMG6 promotes endonucleolytic
20 cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* **16**: 49-55.
- 21 Eberle AB, Stalder L, Mathys H, Orozco RZ, Muhlemann O. 2008. Posttranscriptional gene
22 regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol* **6**: e92.
- 23 Fatscher T, Boehm V, Weiche B, Gehring NH. 2014. The interaction of cytoplasmic poly(A)-
24 binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay.
25 *RNA* **20**: 1579-1592.
- 26 Fiorini F, Boudvillain M, Le Hir H. 2013. Tight intramolecular regulation of the human Upf1
27 helicase by its N- and C-terminal domains. *Nucleic Acids Res* **41**: 2404-2415.
- 28 Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M,
29 Krawczak M, Hampe J et al. 2012. Genome-wide search for novel human uORFs and N-terminal
30 protein extensions using ribosomal footprinting. *Genome Res* **22**: 2208-2218.
- 31 Fukuhara N, Ebert J, Unterholzner L, Lindner D, Izaurralde E, Conti E. 2005. SMG7 is a 14-3-3-like
32 adaptor in the nonsense-mediated mRNA decay pathway. *Mol Cell* **17**: 537-547.
- 33 Gatfield D, Izaurralde E. 2004. Nonsense-mediated messenger RNA decay is initiated by
34 endonucleolytic cleavage in *Drosophila*. *Nature* **429**: 575-578.
- 35 Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE. 2009. The hierarchy of exon-junction complex
36 assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay.
37 *PLoS Biol* **7**: e1000120.

- 1 German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA,
2 Nobuta K, German R et al. 2008. Global identification of microRNA-target RNA pairs by parallel
3 analysis of RNA ends. *Nat Biotechnol* **26**: 941-946.
- 4 Glavan F, Behm-Ansmant I, Izaurralde E, Conti E. 2006. Structures of the PIN domains of SMG6
5 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J* **25**: 5117-5125.
- 6 Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, Ecker JR. 2008. A
7 link between RNA metabolism and silencing affecting Arabidopsis development. *Dev Cell* **14**: 854-
8 866.
- 9 Harigaya Y, Parker R. 2012. Global analysis of mRNA decay intermediates in *Saccharomyces*
10 *cerevisiae*. *Proc Natl Acad Sci U S A* **109**: 11764-11769.
- 11 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK.
12 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory
13 elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.
- 14 Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. 2004. Nonsense-mediated decay approaches
15 the clinic. *Nat Genet* **36**: 801-808.
- 16 Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E. 2008. SMG6 is the catalytic
17 endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14**: 2609-2617.
- 18 Jonas S, Weichenrieder O, Izaurralde E. 2013. An unusual arrangement of two 14-3-3-like
19 domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay.
20 *Genes Dev* **27**: 211-225.
- 21 Joncourt R, Eberle AB, Rufener SC, Muhlemann O. 2014. Eukaryotic initiation factor 4G
22 suppresses nonsense-mediated mRNA decay by two genetically separable mechanisms. *PLoS One* **9**:
23 e104391.
- 24 Karam R, Wengrod J, Gardner LB, Wilkinson MF. 2013. Regulation of nonsense-mediated mRNA
25 decay: implications for physiology and disease. *Biochim Biophys Acta* **1829**: 624-633.
- 26 Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S.
27 2006. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex
28 triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* **20**: 355-367.
- 29 Kervestin S, Jacobson A. 2012. NMD: a multifaceted response to premature translational
30 termination. *Nat Rev Mol Cell Biol* **13**: 700-712.
- 31 Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP
32 reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol*
33 *Biol* **17**: 909-915.
- 34 Kurosaki T, Li W, Hoque M, Popp MW, Ermolenko DN, Tian B, Maquat LE. 2014. A post-
35 translational regulatory switch on UPF1 controls targeted mRNA degradation. *Genes Dev* **28**: 1900-
36 1916.
- 37 Loh B, Jonas S, Izaurralde E. 2013. The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT
38 deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev*
39 **27**: 2125-2138.

- 1 Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, Jensen TH. 2014. Human
2 nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes.
3 *Genes Dev* **28**: 2498-2517.
- 4 Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: an intricate machinery
5 that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**: 665-677.
- 6 Muhlemann O, Lykke-Andersen J. 2010. How and where are nonsense mRNAs degraded in
7 mammalian cells? *RNA Biol* **7**: 28-32.
- 8 Muhlrud D, Parker R. 1999. Aberrant mRNAs with extended 3' UTRs are substrates for rapid
9 degradation by mRNA surveillance. *RNA* **5**: 1299-1307.
- 10 Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing genes:
11 when nonsense affects RNA abundance. *Trends Biochem Sci* **23**: 198-199.
- 12 Neu-Yilik G, Gehring NH, Thermann R, Frede U, Hentze MW, Kulozik AE. 2001. Splicing and 3' end
13 formation in the definition of nonsense-mediated decay-competent human beta-globin mRNPs.
14 *EMBO J* **20**: 532-540.
- 15 Nguyen LS, Wilkinson MF, Gecz J. 2014. Nonsense-mediated mRNA decay: inter-individual
16 variability and human disease. *Neurosci Biobehav Rev* **46 Pt 2**: 175-186.
- 17 Nicholson P, Josi C, Kurosawa H, Yamashita A, Muhlemann O. 2014. A novel phosphorylation-
18 independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res* **42**:
19 9217-9235.
- 20 Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Muhlemann O. 2010.
21 Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality
22 control and the double-life of NMD factors. *Cell Mol Life Sci* **67**: 677-700.
- 23 Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, Ohno S. 2012. N- and C-
24 terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD.
25 *Nucleic Acids Res* **40**: 1251-1266.
- 26 Schmidt SA, Foley PL, Jeong DH, Rymarquis LA, Doyle F, Tenenbaum SA, Belasco JG, Green PJ.
27 2015. Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of
28 endogenous NMD targets in human cells. *Nucleic Acids Res* **43**: 309-323.
- 29 Singh G, Rebbapragada I, Lykke-Andersen J. 2008. A competition between stimulators and
30 antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol*
31 **6**: e111.
- 32 Stockklausner C, Breit S, Neu-Yilik G, Echner N, Hentze MW, Kulozik AE, Gehring NH. 2006. The
33 uORF-containing thrombopoietin mRNA escapes nonsense-mediated decay (NMD). *Nucleic Acids Res*
34 **34**: 2355-2363.
- 35 Thermann R, Neu-Yilik G, Deters A, Frede U, Wehr K, Hagemeyer C, Hentze MW, Kulozik AE. 1998.
36 Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* **17**: 3484-
37 3494.
- 38 Yamashita A, Ohnishi T, Kashima I, Taya Y, Ohno S. 2001. Human SMG-1, a novel
39 phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA

1 surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev*
2 **15**: 2215-2228.

3 Zarnack K, Konig J, Tajnik M, Martincorena I, Eustermann S, Stevant I, Reyes A, Anders S,
4 Luscombe NM, Ule J. 2013. Direct competition between hnRNP C and U2AF65 protects the
5 transcriptome from the exonization of Alu elements. *Cell* **152**: 453-466.

6

7

1 **Figure Legends**

2 **Figure 1. Computational analysis of 3' fragment libraries.**

3 (A) Workflow of 3' fragment library preparation for Illumina sequencing. Poly(A)⁺ RNA was used
4 for library construction. The free 5' phosphate of 3' fragments allows the ligation of a 5' RNA linker.
5 Subsequently, RT-PCR was performed with a linker oligo(dT) followed by second strand synthesis,
6 EcoP15I digest to equally shorten 3' fragment length down to 25-27 nt. Next a 3' dsDNA adaptor was
7 ligated harboring a fixed barcode and an index sequence to distinguish between knockdown
8 conditions, followed by construct amplification for Illumina sequencing.

9 (B) Normalized densities of cDNA reads plotted around indicated features for each degradome-
10 sequencing library. The termination codon is indicated by a red cross.

11 (C) Normalized density of cDNA reads plotted around uORF starts and uORF ends. All uORFs with
12 at least 1 read in the siXRN1 condition were evaluated. A red cross indicates the position of the the
13 uORF termination codon. See methods for full details.

14 **Figure 2. Functional analysis of potential endogenous NMD targets identified during high-** 15 **throughput sequencing.**

16 (A-D) A PCR based approach was used to quantify 3' fragment levels upon knockdown of NMD
17 effectors. HeLa cells were transiently transfected with the indicated siRNAs and poly(A)⁺ RNA was
18 extracted. 3' decay intermediates were ligated to a RNA linker, followed by reverse transcription with
19 an oligo(dT) primer and gene-specific PCR. Overall transcript levels were determined with primer
20 pairs located downstream of the estimated endocleavage site (second panel each, indicated as "gene
21 int."). PCR with TATA-Box binding protein (TBP) primers was used for cDNA level determination. For
22 each class of NMD targets (A, B, C, D respectively), the same set of cDNA was used and therefore the
23 same TBP profiles are shown for each target type. Degradome-sequencing reads were plotted against
24 their position on the indicated mRNAs. Endocleavage events within the selected targets were

1 visualized in an enlarged view spanning 150 nt (± 75 nt). The position of the second nucleotide of the
2 respective stop codon is set as 0. Mapped reads per nucleotide were plotted against the mRNA
3 length for each knockdown condition (Luciferase (Luc, black), XRN1 (orange), XRN1/SMG6 (blue)).

4 (A) PDRG1, SURF6 and MED10, transcripts with a long 3' UTR.

5 (B) IFRD1 and BAG1, transcripts with an uORF.

6 (C) C11orf31, encoding for Selenoprotein H, containing a selenocysteine codon.

7 (D) TMEM222, incorporation of an alternative exon (indicated in purple), harboring a PTC as well
8 as RPL10A, a transcript gaining a PTC probably due to alternative splice site usage within the third
9 intron.

10 **Figure 3. Endocleavage of endogenous targets is regulated SMG7-dependently.**

11 (A) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated
12 siRNAs, plasmids and reporter constructs using β -globin as control mRNA. Endocleavage products (3'
13 fragments) are indicated. A representative western blot is shown at the bottom, using tubulin as
14 loading control. Mean values of reporter and 3' fragment signal \pm SD ($n = 3$) were quantified and
15 normalized to the Luc control knockdown.

16 (B) Detection of endogenous 3' fragments for different classes of NMD targets ((1) long 3' UTR,
17 (2) uORF, (3) selenocysteine codon, (4) alternative splicing) dependent on SMG7 expression levels.
18 HeLa cells were transiently transfected with the indicated siRNAs and plasmids. The same PCR-based
19 approach as shown in Fig. 2 was applied. The cDNA was transcribed from total RNA with an oligo(dT)
20 primer and TBP served to determine overall cDNA levels. For each class ((1) - (4)), the same set of
21 cDNA was used.

22

1 **Figure 4. Robustness of the NMD system activated via different decay-inducing features.**

2 (A - C) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated
3 siRNAs, plasmids and reporter constructs. Co-transfected β -globin (A, C) or LacZ-H4 (B) served as
4 control mRNA. Endocleavage products (3' fragments) are indicated. To eliminate effects caused by
5 altered transfection efficiency or availability of the general gene expression machinery, decreasing
6 amounts of TPI-WT-OH (A and B: 3, 2 and 0 μ g; C: 4, 2 and 0 μ g) were co-transfected with increasing
7 amounts of reporter (A and B: 1, 2 and 4 μ g; C: constant 3 μ g) or competition vector (C: 0, 2 and 4
8 μ g). The OH constructs lack the northern blot probe binding sites and are not detected by the 3'
9 probe. Mean values of reporter and 3' fragment signal \pm SD (n = 3) were quantified and normalized to
10 the control. The ratio of 3' fragment to reporter mRNA levels is indicated below the bars. A
11 representative western blot is shown at the bottom, using tubulin as loading control (C).

12 **Figure 5. SMG7 abundance influences endocleavage efficiency for long 3' UTR targets.**

13 (A) Domain structure of SMG7, showing the N-terminal 14-3-3-like domain (interacts with UPF1)
14 and α -helical extensions as well as the C-terminal PC region (interacts with POP2).

15 (B) Northern blot of RNA samples extracted from HeLa cells transfected with the indicated
16 siRNAs and reporter constructs. Co-transfected β -globin served as control mRNA. Endocleavage
17 products (3' fragments) are indicated. Mean values of reporter and 3' fragment signal \pm SD (n = 3)
18 were quantified and normalized to the XRN1 control knockdown. A representative western blot is
19 shown at the bottom, using tubulin as loading control.

20 (C) Schematic representation of the transfected triosephosphate isomerase (TPI) reporter with a
21 PTC at amino acid position 160. Exons are depicted as white (untranslated) or black (translated)
22 boxes, introns as two connecting black lines and the northern probe binding sites as light gray boxes.
23 Vector derived 5' UTR intron and SV40 poly(A) signal (pA) are shown.

1 (D) Northern blot of RNA samples extracted from HeLa cells transfected with the indicated
2 siRNAs, plasmids and reporter constructs using LacZ-4H as control mRNA. Endocleavage products (3'
3 fragments) are indicated. Mean values of reporter and 3' fragment signal \pm SD (n = 3) were quantified
4 and normalized to the XRN1 control knockdown. A representative western blot is shown at the
5 bottom, using tubulin as loading control.

6 **Figure 6. Deletions of C-terminal UPF1 phosphorylation sites impair NMD differently.**

7 (A) UPF1 protein architecture is depicted schematically. All structural and functional domains
8 are indicated, the presence of potential phosphorylation sites (SQ/TQ) are shown in red and blue,
9 respectively.

10 (B - C) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated
11 siRNAs and reporter constructs. Co-transfected LacZ-H4 (B) or β -globin (C) served as control mRNA.
12 Endocleavage products (3' fragments) are indicated. Mean values of reporter and 3' fragment signal \pm
13 SD (n = 3) were quantified and normalized to the XRN1 control knockdown. Representative western
14 blots are shown at the bottom, using tubulin as loading control.

15 **Figure 7. Model of degradative and regulatory branches of NMD.**

16 Depending on the mRNA architecture, NMD substrates are either regulated (long 3' UTR and
17 uORF targets; encoding for mostly functional protein) or degraded (PTC targets; encoding for
18 aberrant proteins). During degradative NMD, the robust elimination of PTC-containing mRNAs is
19 mainly achieved by endonucleolytic decay via SMG6, which is supported by exonucleolytic
20 degradation induced by SMG5/7. Alternatively, the combined activity of SMG5/7 and SMG6 (exo-
21 and endonucleolytic degradation) during the decay of long 3' UTR- and uORF-containing targets
22 allows for regulated degradation ("downregulation"). For further details, see Discussion.

1 **Supplemental Figure Legends**

2 **Supplemental Figure S1. 3' fragment library features.**

3 (A) Knockdown efficiency of samples used for library preparation and library verification were
4 analyzed by western blot. Tubulin serves as loading control.

5 (B) Schematic outline of 3' fragment library preparation (for details see text and Fig. 1A).
6 Comparison of XRN1- and XRN1/SMG6-depleted samples allows the identification of XRN1-sensitive,
7 SMG6-dependent decay intermediates.

8 (C) Mapping efficiency of the 3' fragment libraries.

9 (D) Scatter plots comparing the degradome cleavage events within lowFDR clusters (black) or
10 ENSEMBL transcripts (red) for all replicate experiments of each condition. Sample types and replicate
11 numbers are given along the diagonal. The Spearman's rank correlation (r) for each pair is indicated.

12 (E) Distribution of degradome-sequencing read lengths.

13 (F) Density enrichments of degradome sequencing library reads within indicated genomic
14 features.

15 (H) Number of SMG6-sensitive windows identified within different gene features (5' UTR, ORF, 3'
16 UTR). The count of windows that are located within 50 nt of an annotated termination codon
17 (Windows Ter), was determined. For details see Material and Methods.

18 (G) Global distribution of SMG6-sensitive sites around annotated termination codons for ORF
19 and long 3' UTR genes. The position of the second nucleotide of the respective stop codon is set as 0.

20

21

1 **Supplemental Figure S2. Endocleavage of reporter constructs with 3' UTRs derived from**
2 **endogenous NMD targets.**

3 (A) Scheme of the transfected triosephosphate isomerase (TPI) reporter mRNAs, containing
4 different 3' UTRs. Exons are depicted as white (untranslated) or black (translated) boxes, the inserted
5 3' UTRs in color, introns as two connecting black lines and the northern probe binding sites as light
6 gray boxes. Vector derived 5' UTR intron and SV40 poly(A) signal (pA) are depicted.

7 (B) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated
8 siRNAs, plasmids and reporter constructs using β -globin as control mRNA. Endocleavage products (3'
9 fragments) are indicated.

10 (C) Northern blot analysis of RNA samples derived from HeLa stable cell lines transfected with
11 the indicated siRNAs. 24 h after induction of transcription of the TPI-SURF6 reporter construct by 1
12 μ g/ml doxycycline (Dox), actinomycin D (5 μ g/ml) was added and the cells were harvested at the
13 indicated time points.

14 (D) Mean values of reporter (left) and 3' fragment (right) signals \pm SD (n = 3) were quantified and
15 normalized to the 7SL endogenous control.

16 (E) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated
17 siRNAs, plasmids and reporter constructs using β -globin as control mRNA. Endocleavage products (3'
18 fragments) are indicated. A representative western blot is shown at the bottom, using tubulin as
19 loading control.

20 (F) PCR-based detection of endogenous 3' fragments that derive from different classes of NMD
21 targets ((1) long 3' UTR, (2) uORF, (3) selenocysteine codon, (4) alternative splicing) dependent on
22 SMG7 expression levels. HeLa cells were transiently transfected with the indicated siRNAs and
23 plasmids. The cDNA was transcribed from total RNA with random hexamers and TBP served to
24 determine general cDNA levels. For each target class ((1) - (4)), the same set of cDNA was used.

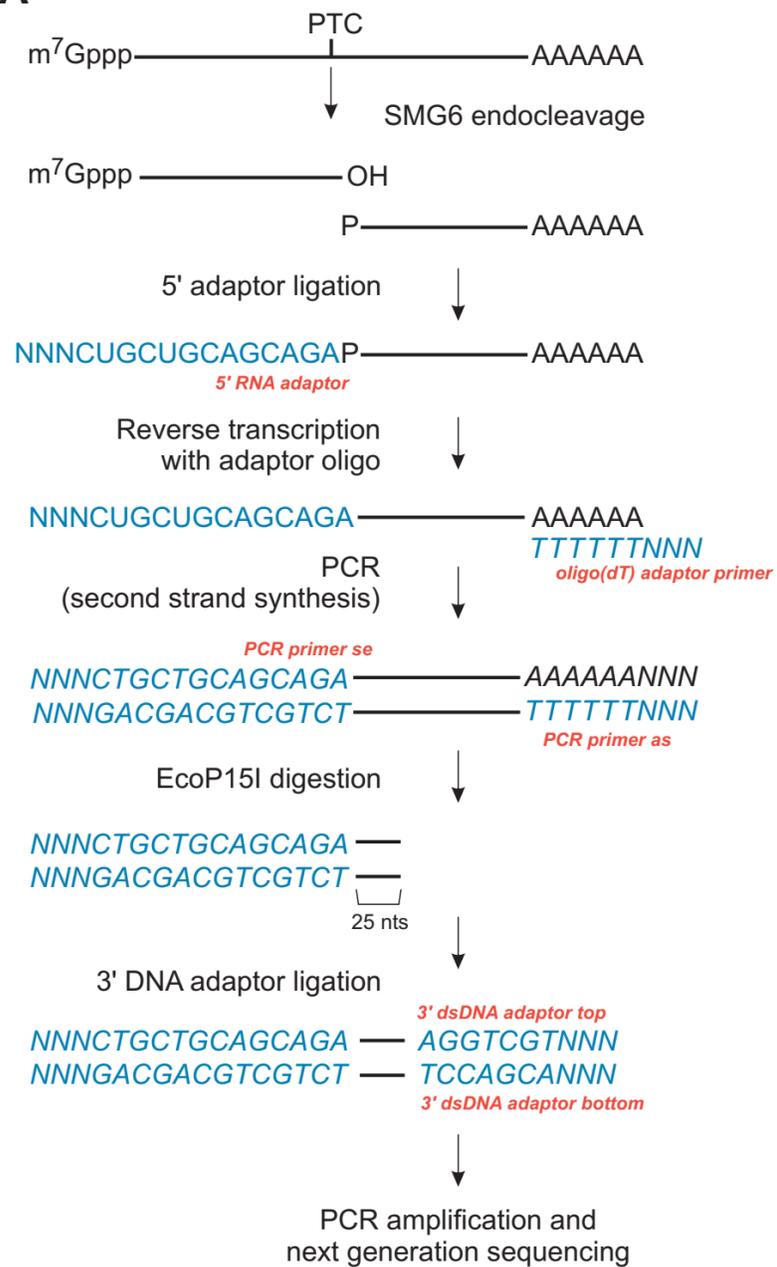
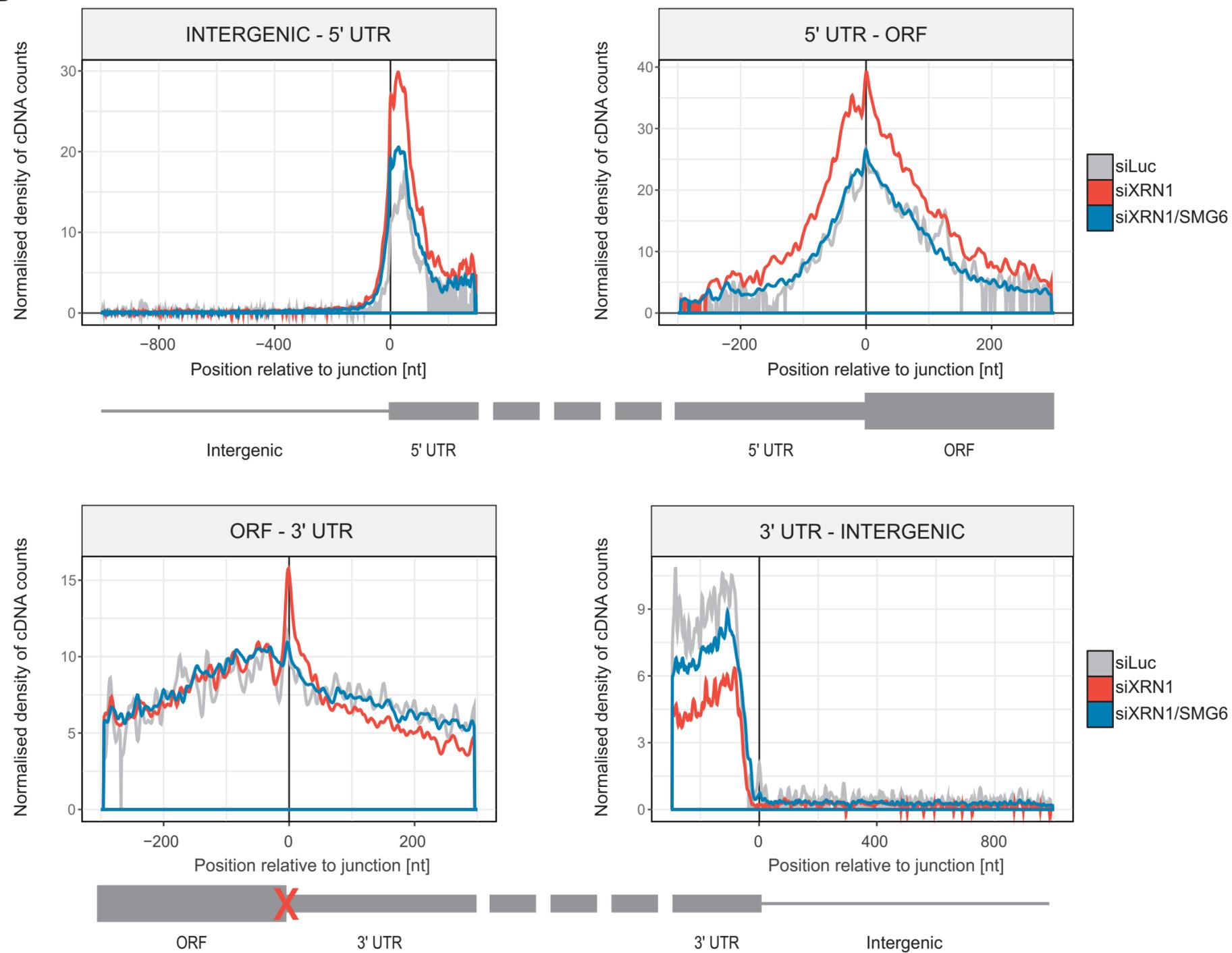
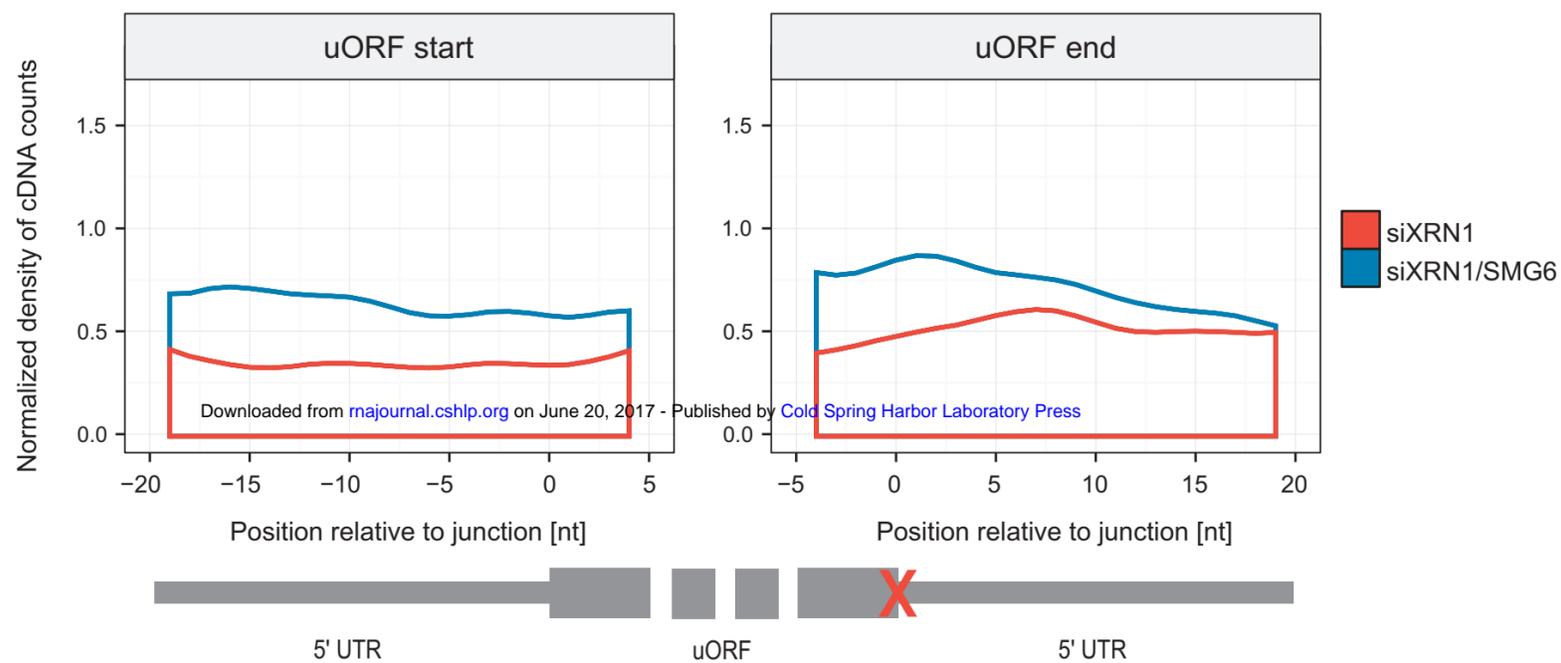
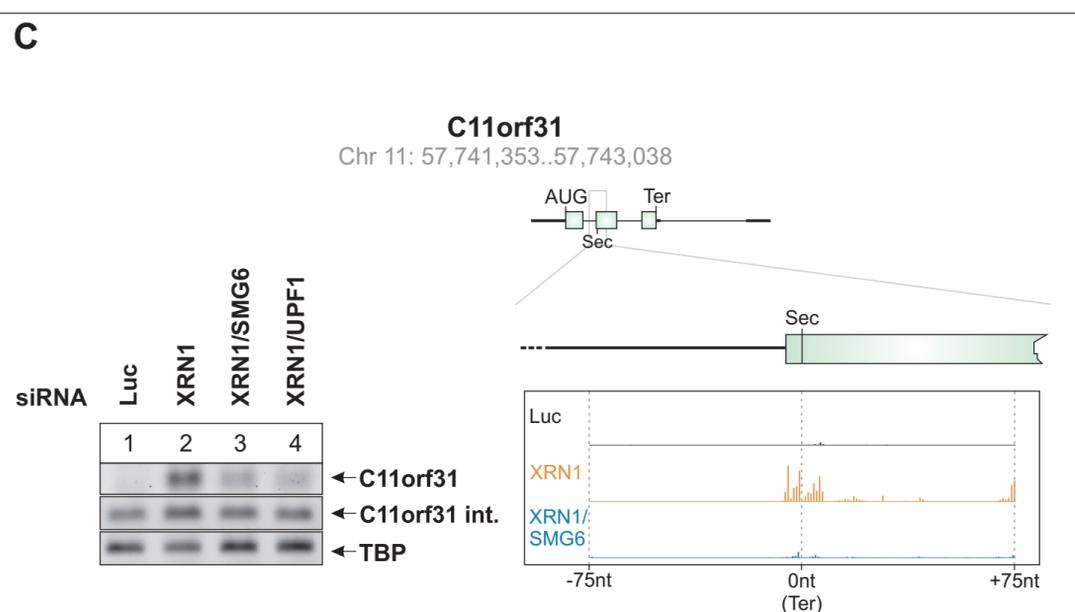
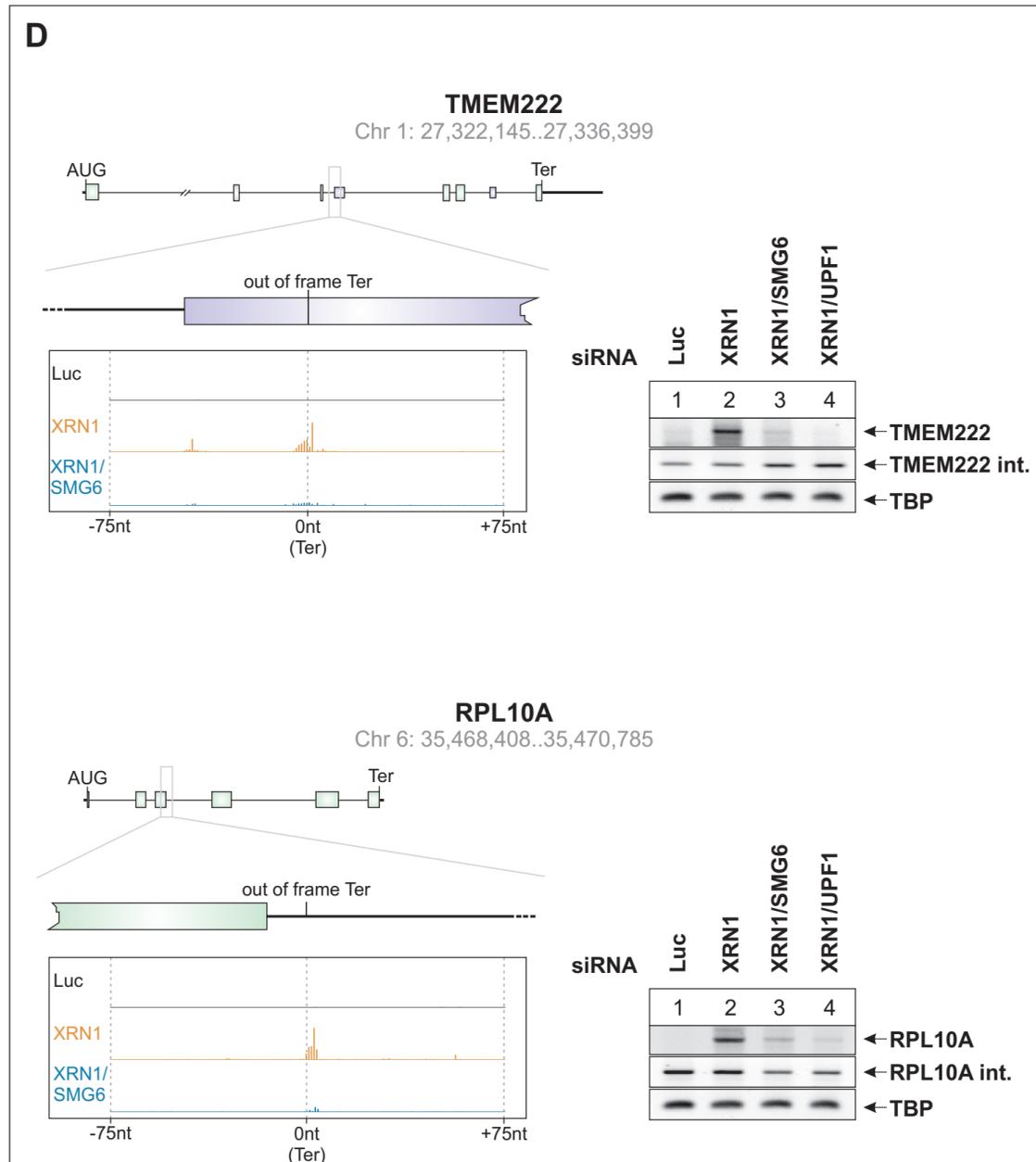
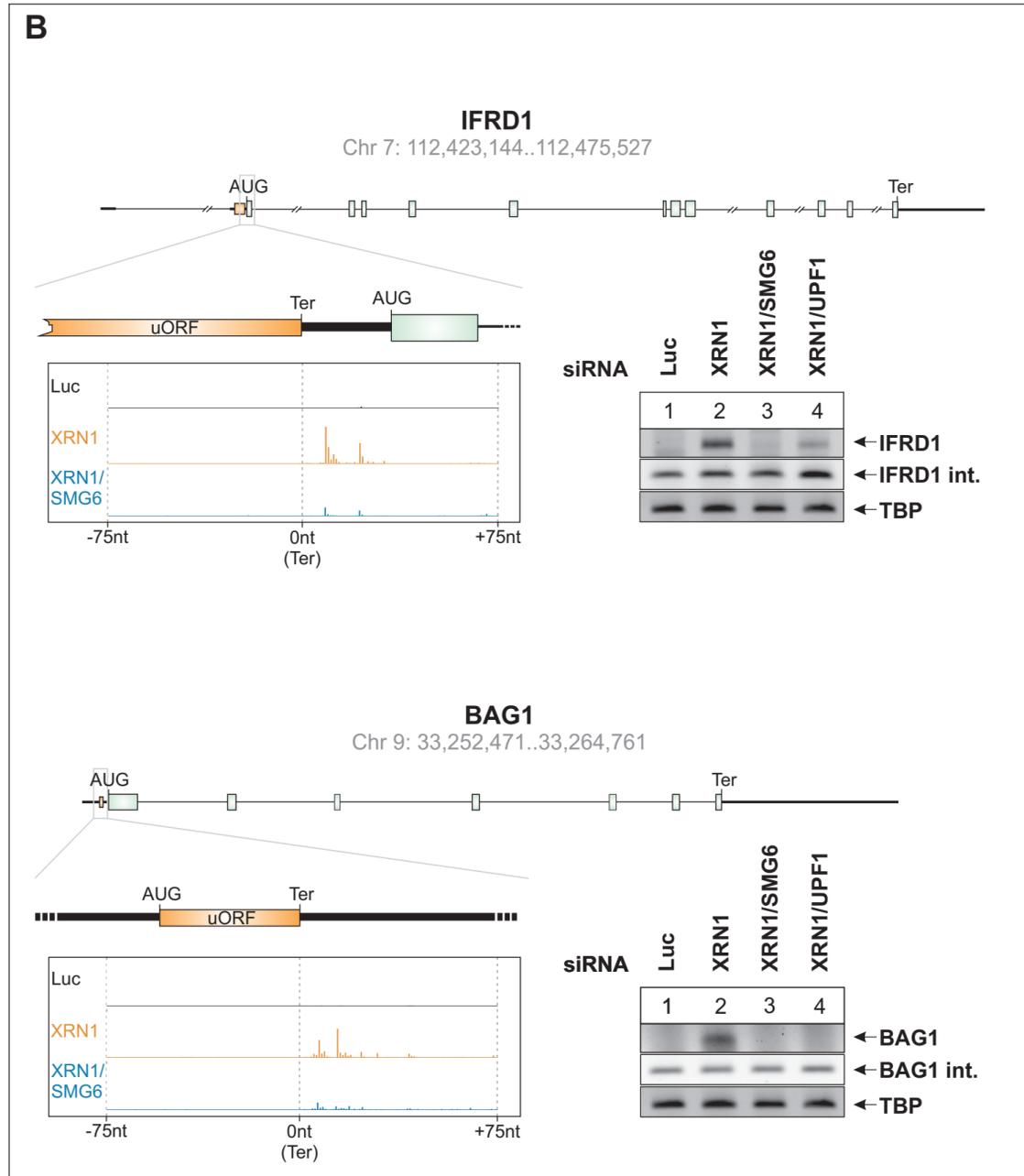
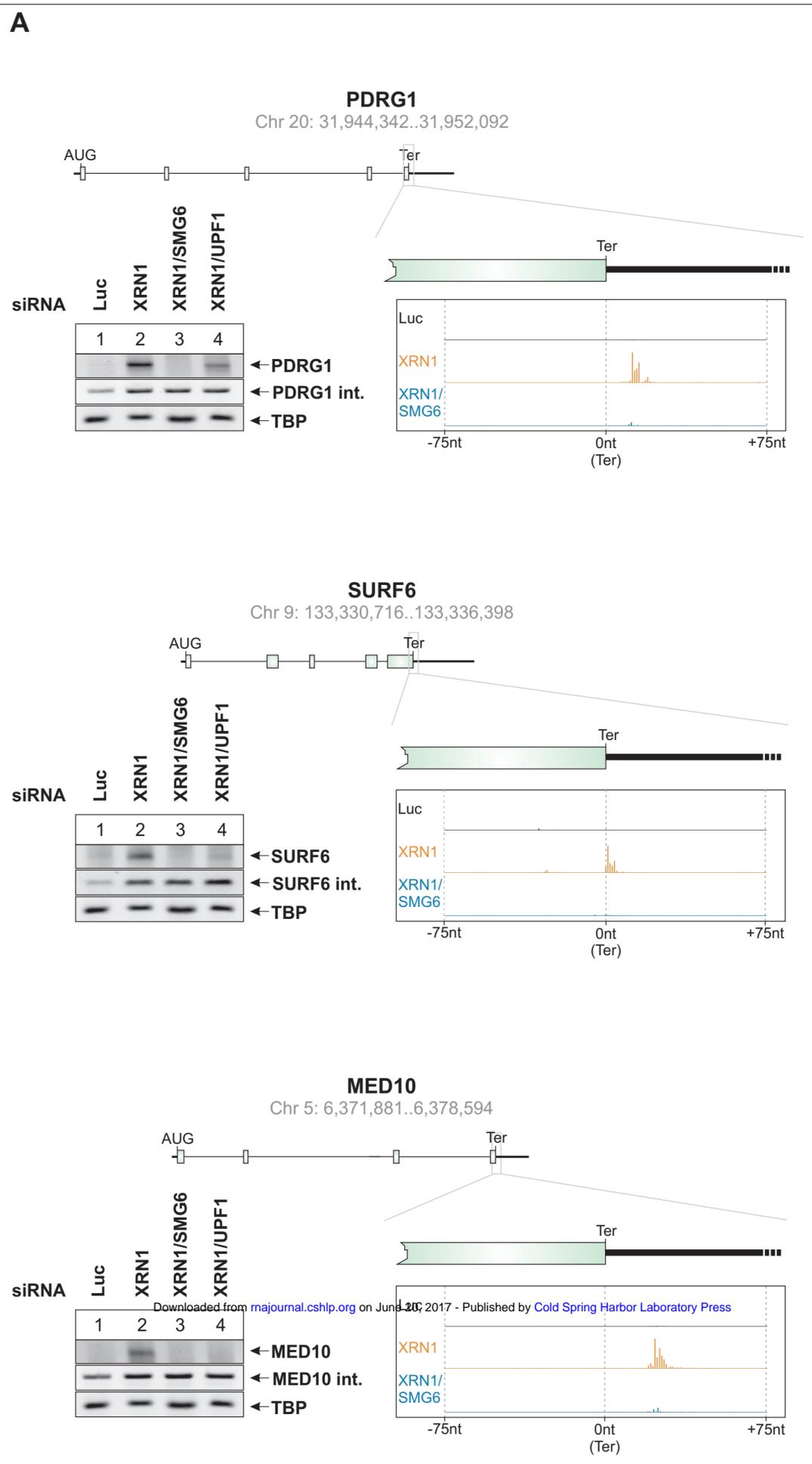
Figure 1**A****B****C**

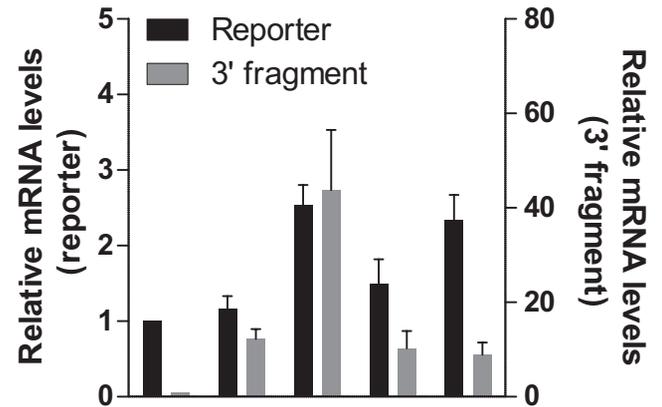
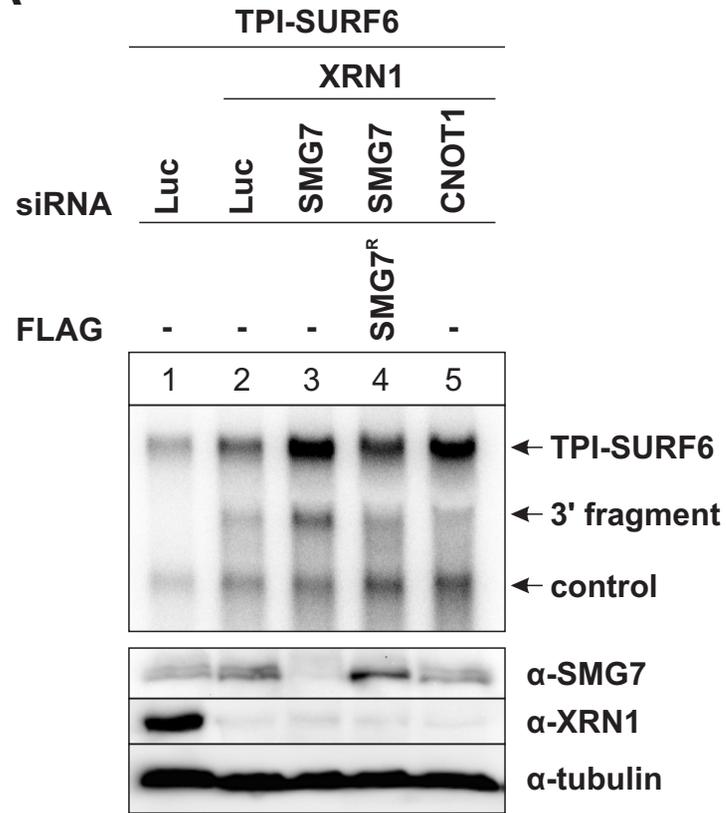
Figure 2



Constitutive exon
 Alternative exon
 Upstream open reading frame (uORF)
 Intron
 Untranslated region (UTR)

Figure 3

A



B

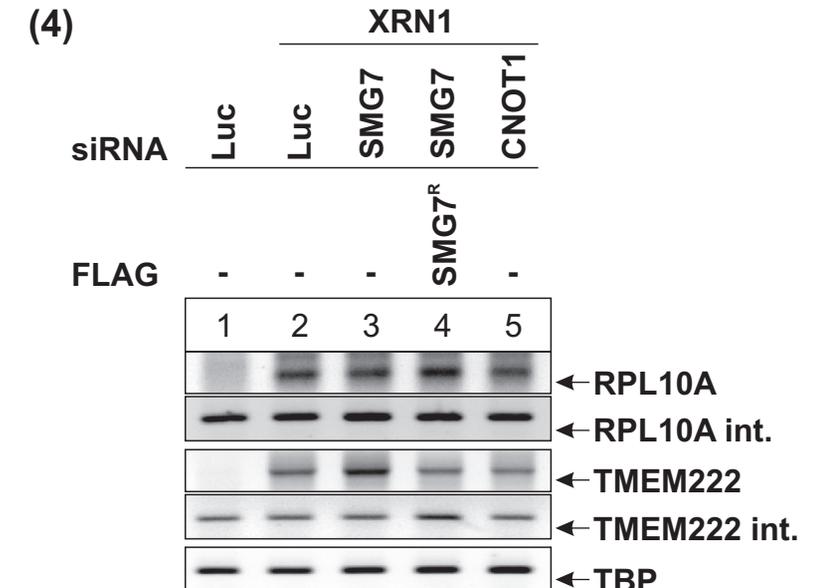
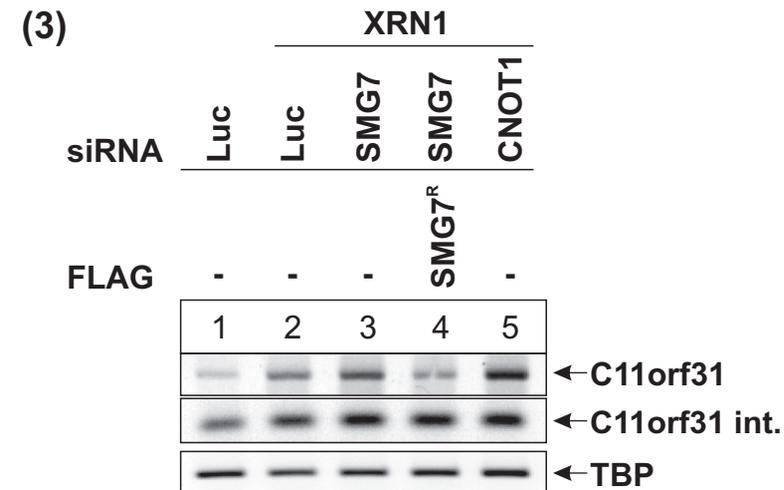
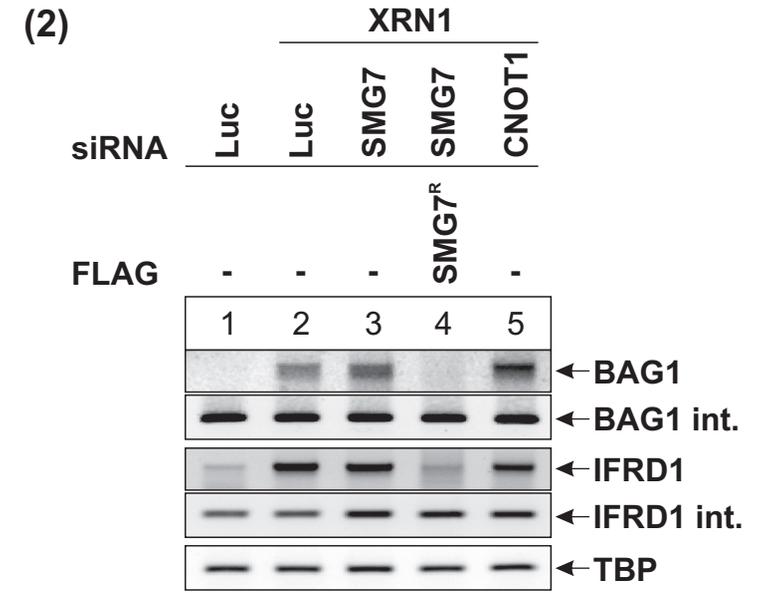
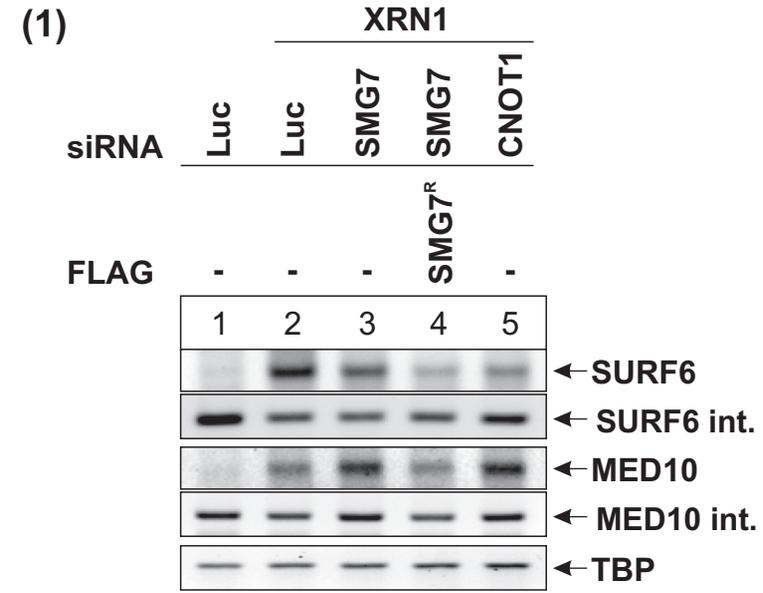


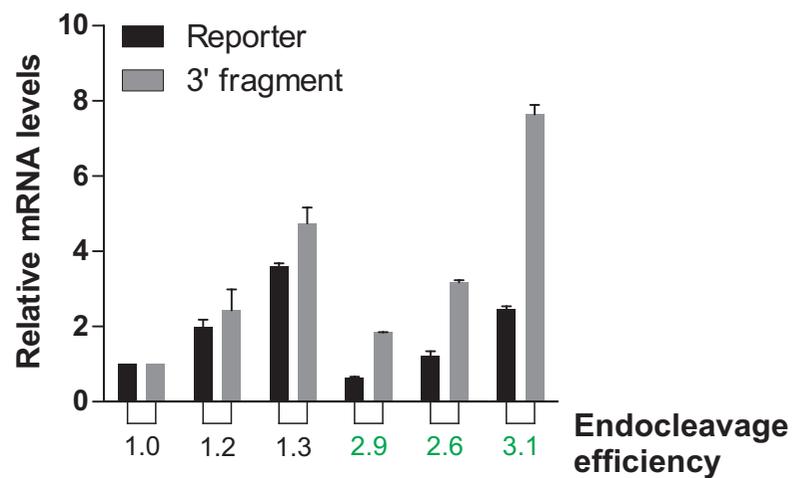
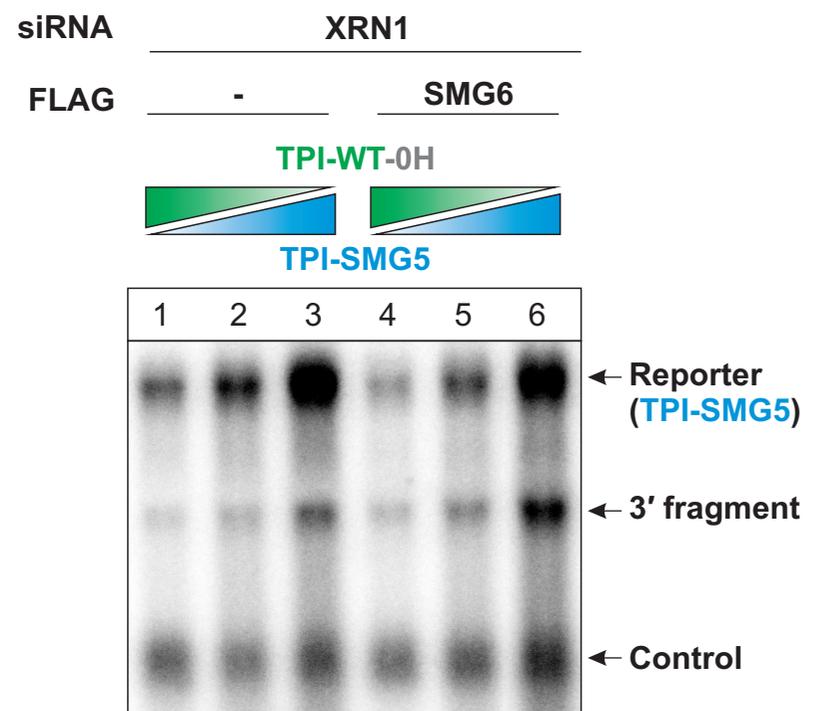
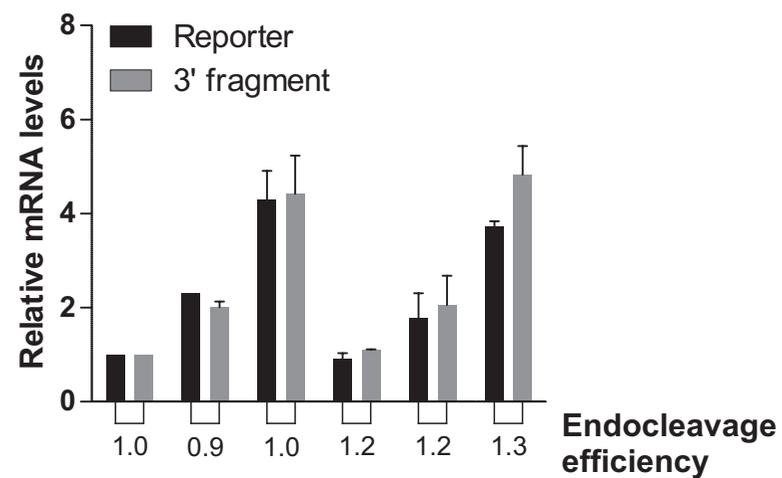
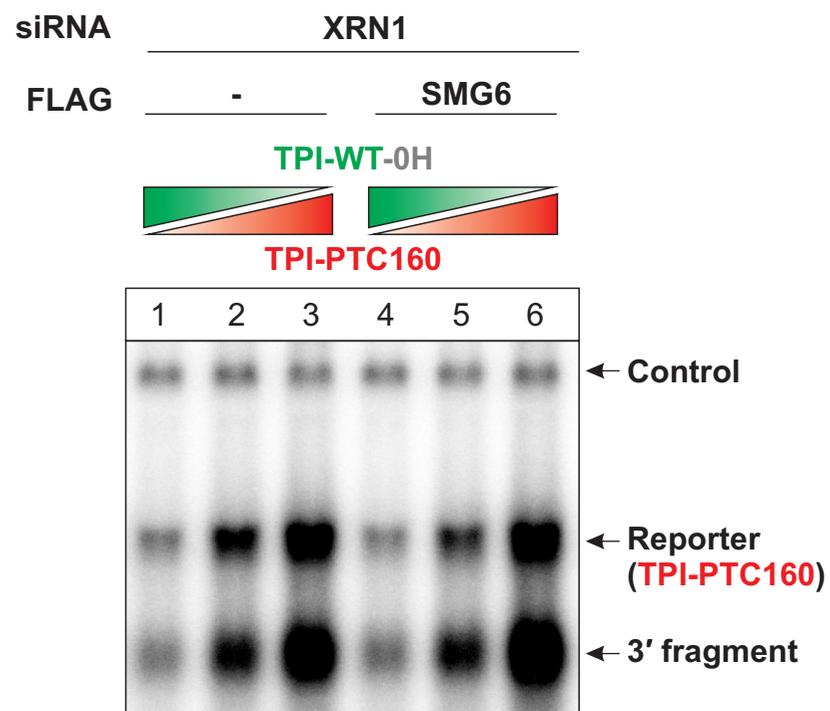
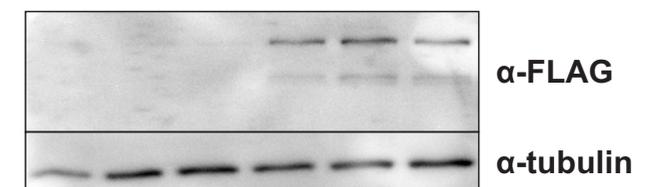
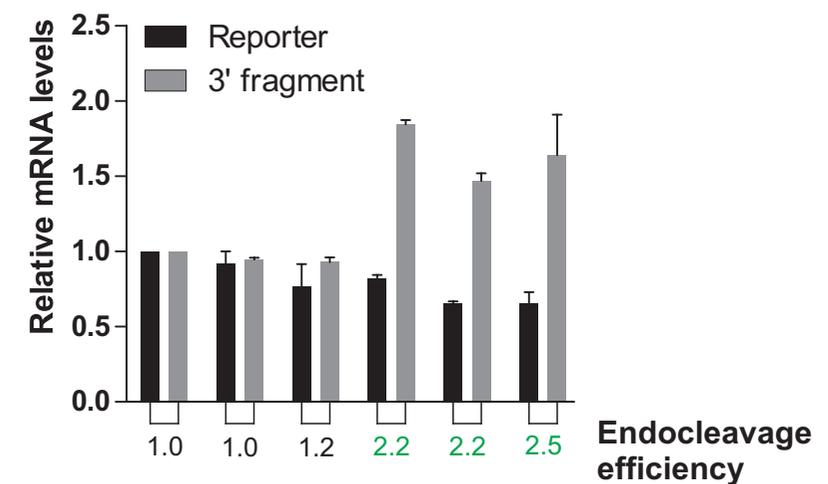
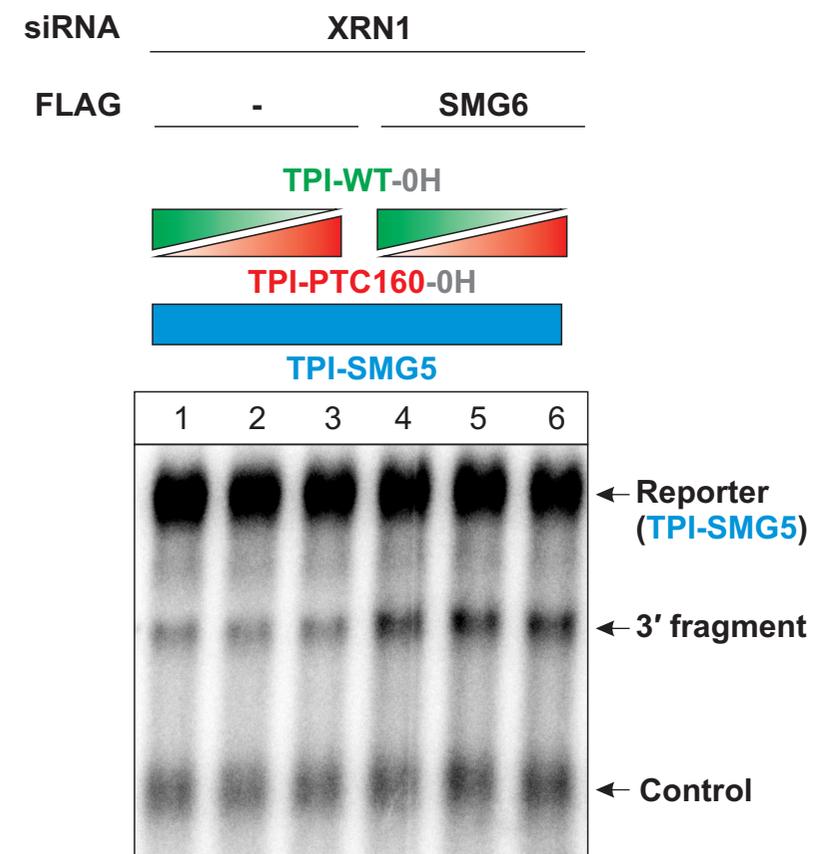
Figure 4**A****B****C**

Figure 5

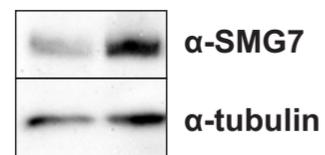
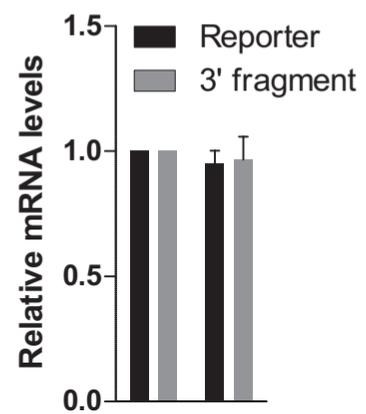
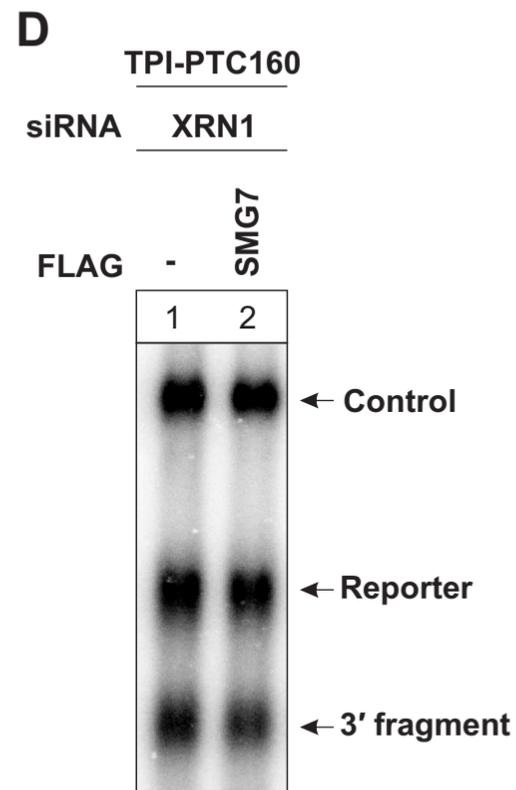
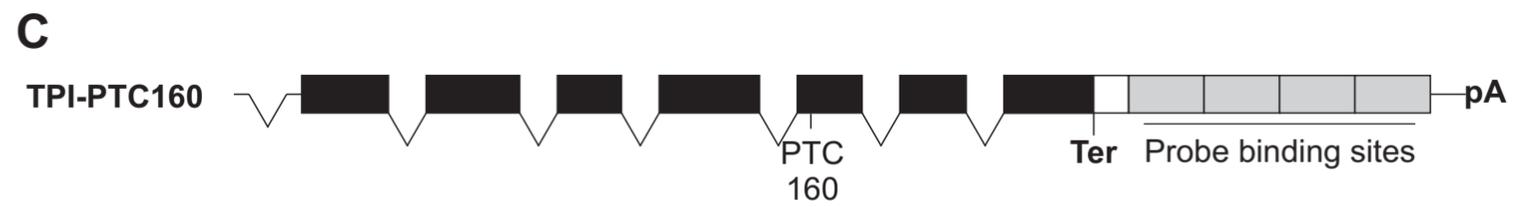
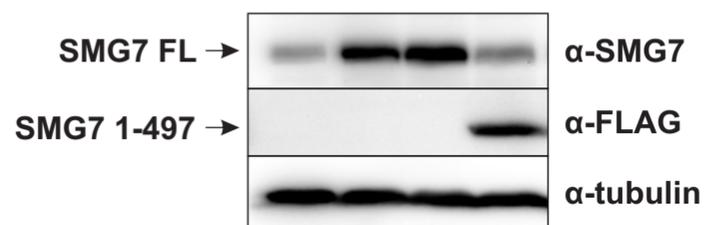
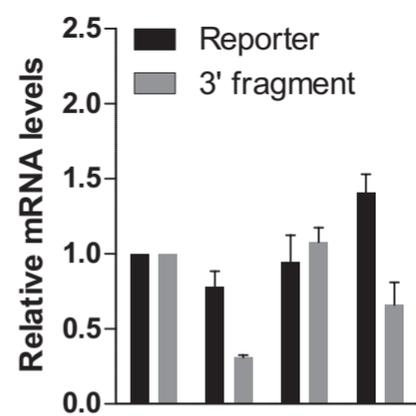
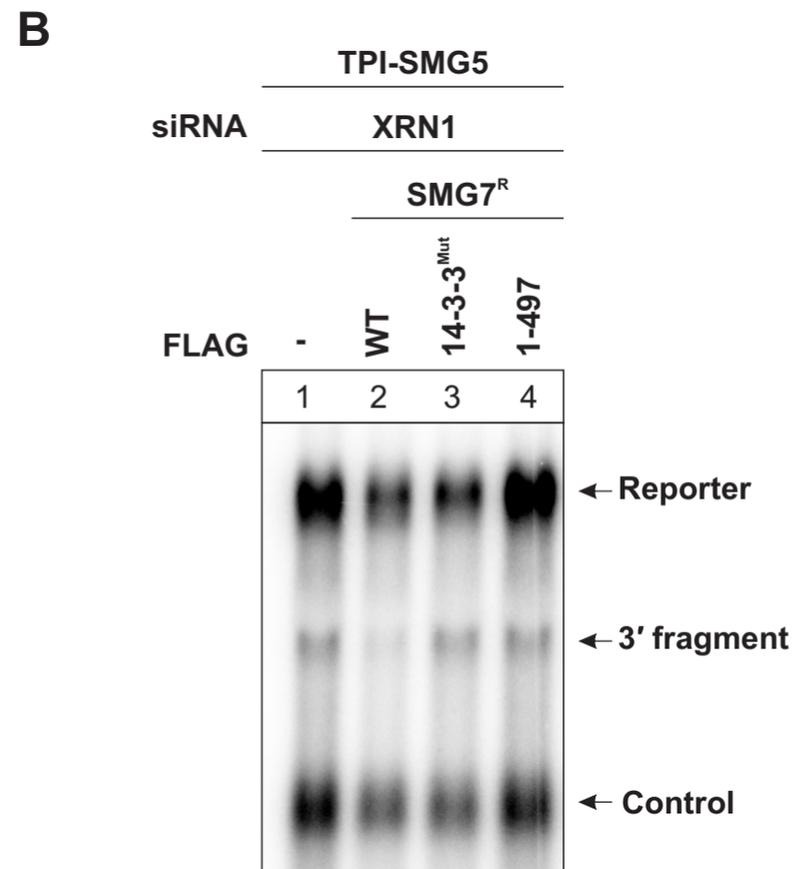
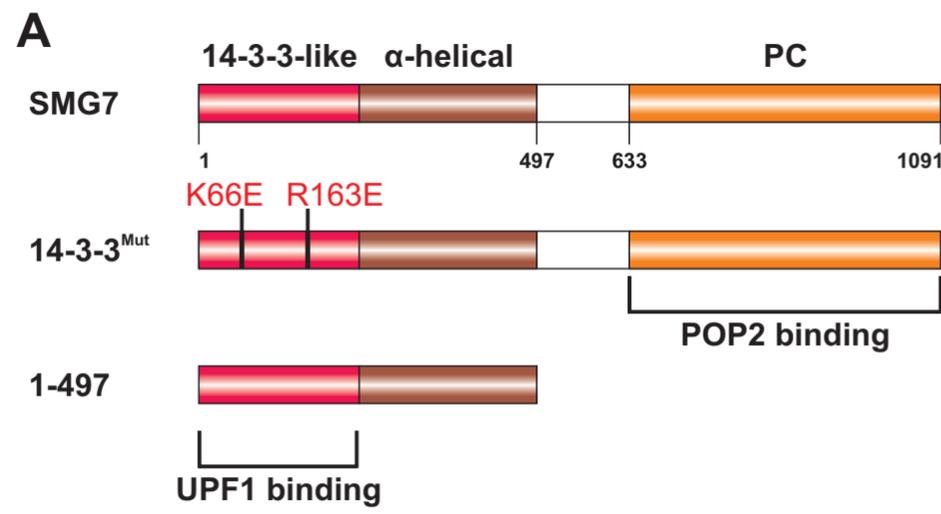
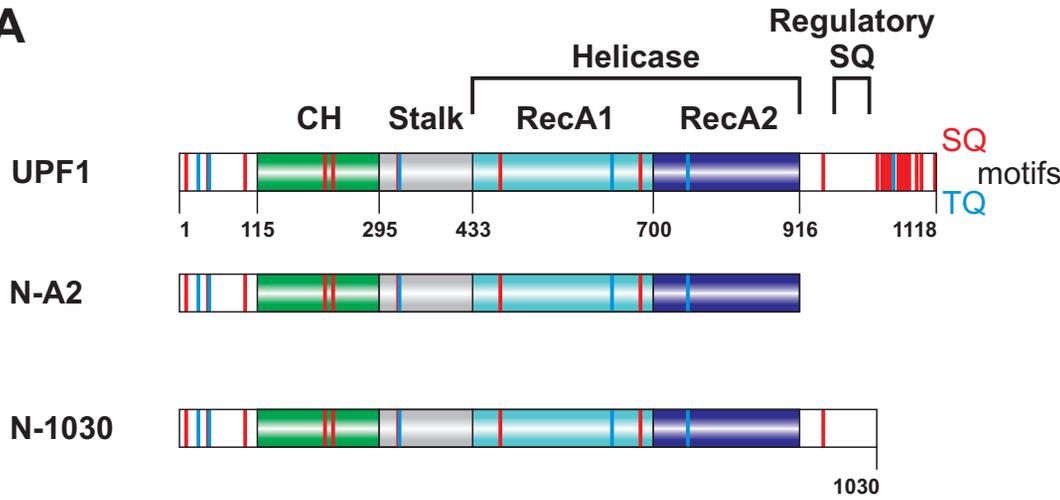
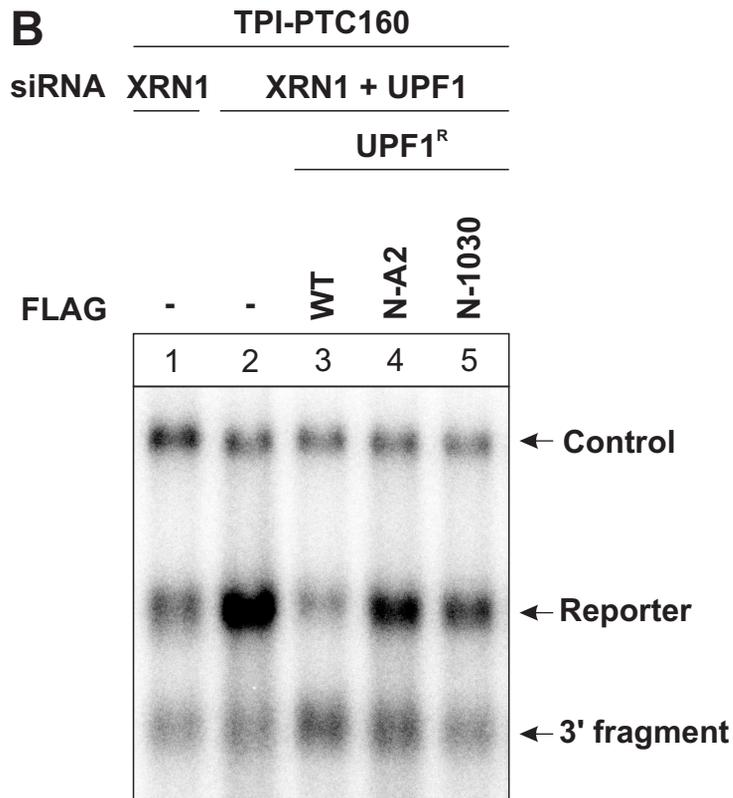


Figure 6

A



B



C

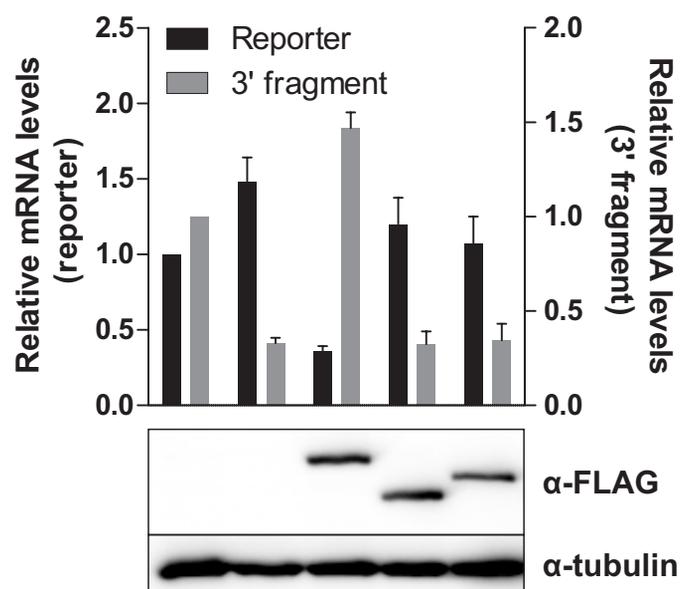
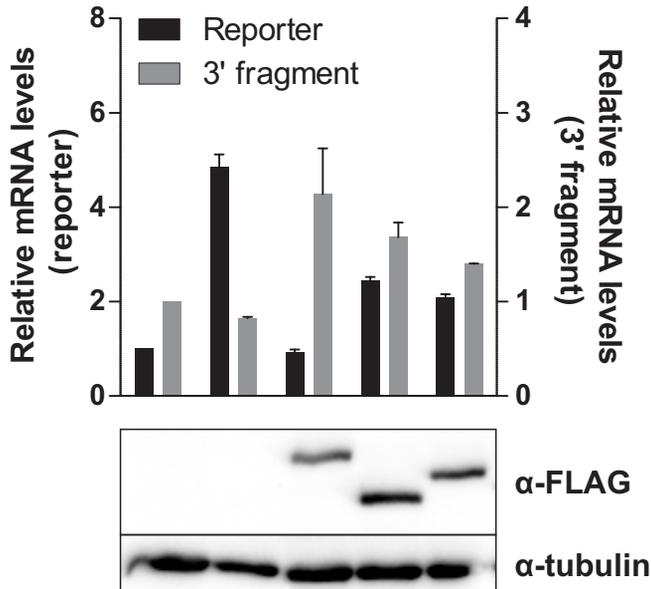
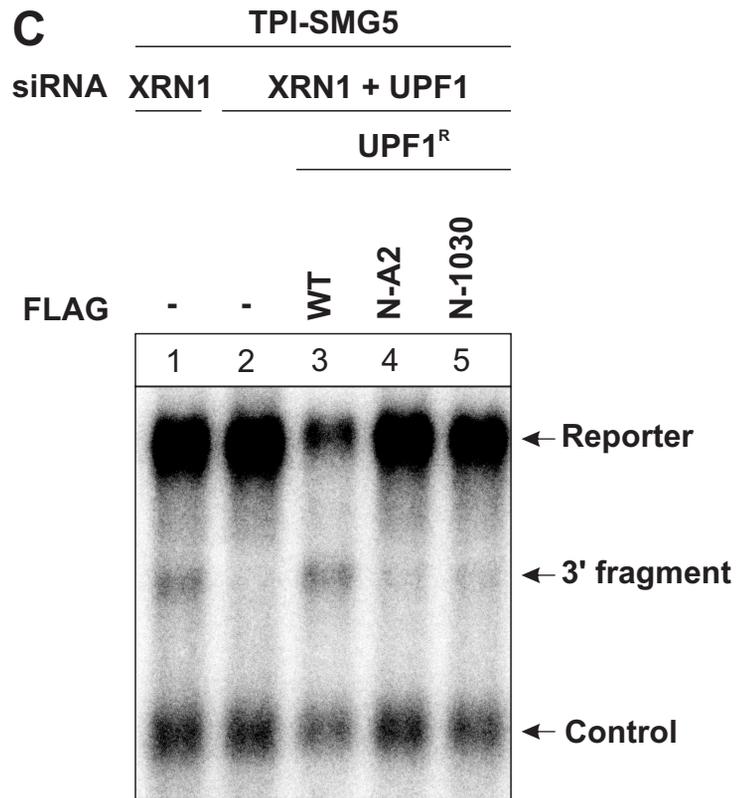
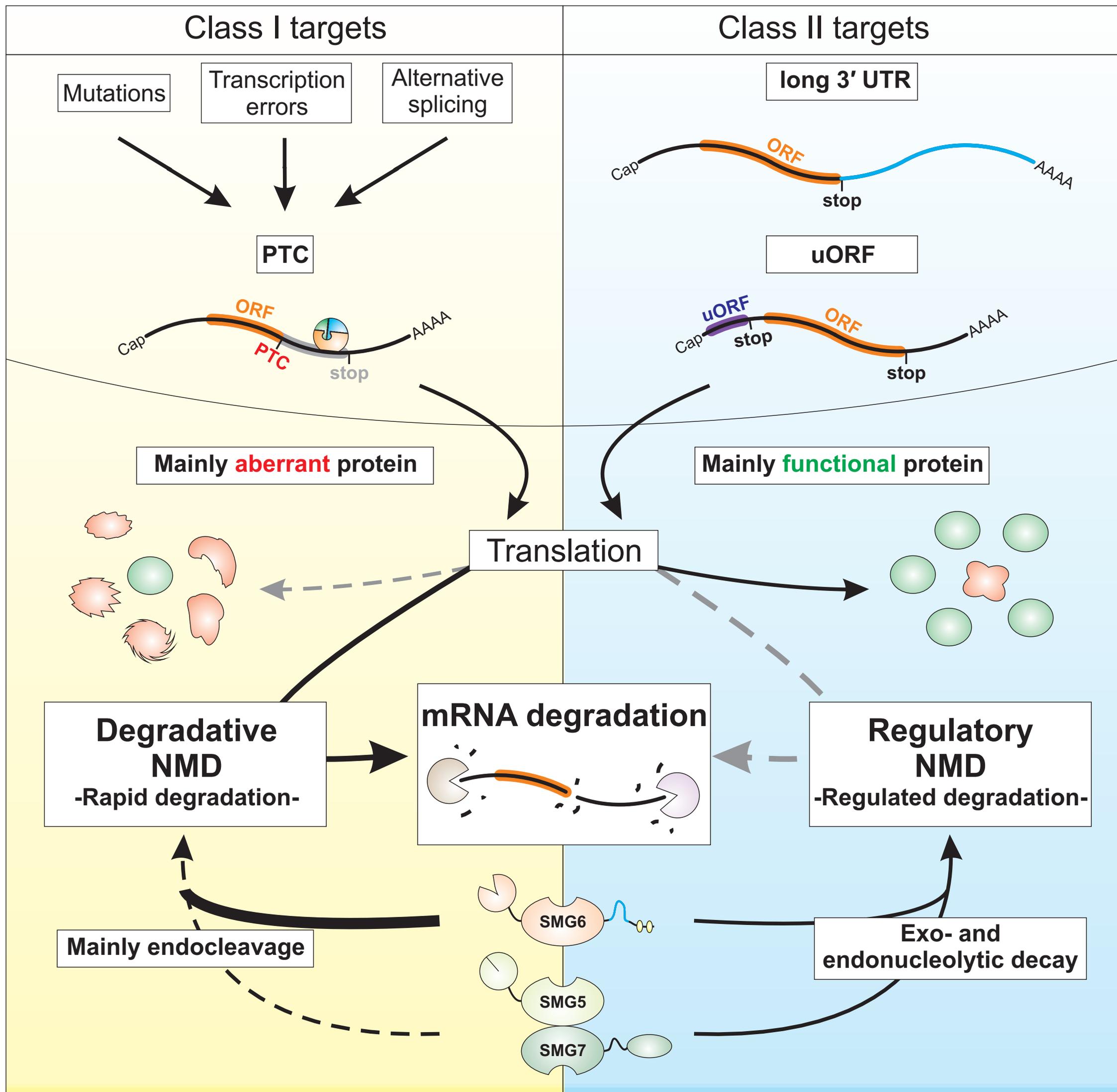


Figure 7





RNA

A PUBLICATION OF THE RNA SOCIETY

Transcript-specific characteristics determine the contribution of endo- and exonucleolytic decay pathways during the degradation of nonsense-mediated decay substrates

Franziska Ottens, Volker Boehm, Christopher R Sibley, et al.

RNA published online May 1, 2017

Supplemental Material <http://rnajournal.cshlp.org/content/suppl/2017/05/01/rna.059659.116.DC1>

P<P Published online May 1, 2017 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner for Exiqon. On the left is a blue, textured spherical object. The text "Work with our RNA experts to find biomarkers in exosomes." is written in white on an orange background. The Exiqon logo is in the top right corner.

Work with our RNA experts to find biomarkers in exosomes. **EXIQON**

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
