Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



# Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways

**Martino Colombo[1,2,3], Evangelos D. Karousis[1], Joël Bourquin[1,4], Rémy Bruggmann[2] and Oliver Mühlemann[1*]**

[1]Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland; [2]Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Bern, CH-3012 Bern, Switzerland; [3]Graduate School for Cellular and Biomedical Sciences, University of Bern, CH-3012 Bern, Switzerland; [4]present address: Adolphe Merkle Institute, University of Fribourg, CH-1700 Fribourg, Switzerland

*corresponding author: oliver.muehlemann@dcb.unibe.ch

Keywords: Nonsense-mediated mRNA decay; RNA turnover, post-transcriptional gene regulation; mRNA-seq; UPF1, SMG6, SMG7, bioinformatics analysis

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press




## ABSTRACT

**Besides degrading aberrant mRNAs that harbor a premature translation termination codon (PTC), nonsense-mediated mRNA decay (NMD) also targets many seemingly "normal" mRNAs that encode for full-length proteins. To identify a *bona fide* set of such endogenous NMD targets in human cells, we applied a meta-analysis approach in which we combined transcriptome profiling of knockdowns and rescues of the three NMD factors UPF1, SMG6 and SMG7. We provide evidence that this combinatorial approach identifies NMD-targeted transcripts more reliably than previous attempts that focused on inactivation of single NMD factors. Our data revealed that SMG6 and SMG7 act on essentially the same transcripts, indicating extensive redundancy between the endo- and exonucleolytic decay routes. Besides mRNAs, we also identified as NMD targets many long non-coding RNAs as well as miRNA and snoRNA host genes. The NMD target feature with the most predictive value is an intron in the 3' UTR, followed by the presence of upstream open reading frames (uORFs) and long 3' UTRs. Furthermore, the 3' UTRs of NMD-targeted transcripts tend to have an increased GC content and to be phylogenetically less conserved when compared to 3' UTRs of NMD insensitive transcripts.**


## INTRODUCTION

Nonsense-mediated mRNA decay (NMD) was initially described as a quality control mechanism clearing transcripts harbouring a premature termination codon (PTC) from the cell (Losson and Lacroute 1979; Maquat et al. 1981). Since a PTC can be caused by a mutation in the gene sequence or by aberrant pre-mRNA splicing, NMD was always associated with abnormal or pathological conditions. More recently however, a large number of mRNAs with no PTC were found to be downregulated, often only moderately, by the NMD pathway (Mendell et al. 2004; Rehwinkel et al. 2005; Tani et al. 2012; Wittmann et al. 2006; Yepiskoposyan et al. 2011). An emerging view of NMD

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



is therefore that of a post-transcriptional mechanism contributing to the fine-tuning of gene expression.

The molecular mechanism of NMD is only partially understood and depends on the interplay of many factors. On-going translation is a pre-requisite for NMD to take place (Carter et al. 1995; Thermann et al. 1998) and there is an emerging consensus that NMD results as a consequence of aberrant translation termination (He and Jacobson 2015; Karousis and Mühlemann 2016; Lykke-Andersen and Jensen 2015), which can be detected as ribosome stalling at NMD-eliciting termination codons (TCs) (Amrani et al. 2004; Peixeiro et al. 2012). The ATP-dependent RNA helicase UPF1 is central for NMD activation. UPF1 binds RNA rather unspecifically and independent of translation (Hurt et al. 2013; Zünd et al. 2013; Hogg and Goff 2010). Activation of NMD in metazoans involves phosphorylation of UPF1 by SMG1 (Kurosaki et al. 2014; Yamashita et al. 2001). In mammalian cells, two mechanistically distinct pathways have been described to execute the degradation of the target mRNAs (Mühlemann and Lykke-Andersen 2010). The endonuclease SMG6 is recruited to NMD-targeted transcripts by activated UPF1 (Okada-Katsuhata et al. 2012; Nicholson et al. 2014; Chakrabarti et al. 2014) and cleaves them in the vicinity of the TC (Huntzinger et al. 2008; Eberle et al. 2009; Lykke-Andersen et al. 2014; Schmidt et al. 2015; Boehm et al. 2014). For the second pathway, the SMG5/SMG7 heterodimer binds to phosphorylated SQ epitopes in the C-terminal part of UPF1 and recruits through the C-terminus of SMG7 the deadenylase CCR4/NOT (Jonas et al. 2013; Loh et al. 2013; Chakrabarti et al. 2014). Whether the SMG6- and the SMG7-mediated decay pathways act independently of each other and maybe even target a distinct subpopulation of mRNAs has so far not been addressed on endogenous targets at a genome-wide level.

The characteristics that render a seemingly “normal” endogenous mRNA to be degraded by NMD by and large still need to be elucidated, owing to our fragmented understanding of the mechanism underlying NMD-eliciting aberrant translation termination. The best-characterized NMD-inducing feature is the presence of an exon junction complex (EJC) farther than 50 nucleotides downstream of the TC, a situation that is very common also in

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



PTC-containing transcripts. However, the presence of a downstream EJC seems to enhance the efficiency of NMD (i.e. it increases the extent of mRNA level reduction) rather than being an essential signal to trigger NMD (Bühler et al. 2006; Singh et al. 2008). Several studies have also shown that long 3' UTRs and short ORFs located upstream of the main ORF (uORFs) can activate NMD (Mendell et al. 2004; Yepiskoposyan et al. 2011; Bühler et al. 2006; Eberle et al. 2008; Hansen et al. 2009; Singh et al. 2008). But these transcript characteristics have only a limited predictive value and many mRNAs with long 3' UTRs or uORFs are in fact insensitive to NMD (Yepiskoposyan et al. 2011; Boehm et al. 2014; Hogg and Goff 2010; Burroughs et al. 2010). It has been shown that poly(A) binding protein C1 (PABPC1) promotes translation termination *in vitro* (Ivanov et al. 2016) and effectively antagonizes NMD when tethered close to PTCs in cells (Ivanov et al. 2008; Eberle et al. 2008; Behm-Ansmant et al. 2007). It is thought that UPF1 competes with PABPC1 for interacting with the eukaryotic release factor 3 (eRF3) (Singh et al. 2008) and that an extended physical distance between the terminating ribosome and the poly(A) tail-associated PABPC1, which is a typical configuration of transcripts with uORFs, PTCs or long 3' UTRs, increases the chance for UPF1 to win this competition and elicit NMD (Mühlemann and Jensen 2012). On the other hand, RNA-binding proteins that prevent UPF1 from accessing the mRNA just downstream of the TC can tilt the balance towards proper termination and inhibit NMD, as recently shown for Rous sarcoma virus (RSV) by polypyrimidine tract binding protein 1 (PTBP1) (Ge et al. 2016).

In order to better identify the NMD-eliciting features of mRNAs, a high confidence set of endogenous NMD-sensitive transcripts is needed. Previous genome-wide studies showed little agreement regarding the endogenous NMD targets, and depletion of different NMD factors affected different sets of transcripts, which prevents a reliable meta-analysis of these data. Partly, these discrepancies can be explained by technical biases inherent to the different methods used. To overcome these limitations, we performed RNA-seq experiments following the latest best practices of the field. We carried out knockdowns of three well-characterized NMD factors (UPF1, SMG6 and

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



SMG7) and also operated the respective rescues, which allowed us to increase the stringency and accuracy of the analysis.

Our results show that despite the existing individual differences, UPF1, SMG6 and SMG7 similarly affect the abundance of a large number of mRNAs. Among the NMD-targeted genes, we found a significant enrichment of miRNA host genes, in addition to the already reported snoRNA host genes (Lykke-Andersen et al. 2014). Many non-coding RNAs also appear to be targeted by NMD, depending on the presence of open reading frames (ORFs) in their sequence, consistent with recent reports showing ribosome association of many supposedly non-coding transcripts (Ingolia et al. 2014). Furthermore, we also obtained evidence of transcription upstream of canonical start sites, which appears to be partially cleared by the NMD pathway, but to a lesser extent than has been reported for yeast (Malabat et al. 2015). Already known NMD-inducing features were also enriched among our NMD targets, including 3' UTR introns, uORFs and long 3' UTRs. Additionally, 3' UTRs of NMD targets have on average a higher GC content and are phylogenetically less conserved than NMD-immune transcripts.

## RESULTS

### Experimental setup to identify a high-confidence set of NMD-targeted transcripts

Based on the current literature it is unclear if what is commonly termed NMD constitutes a single biochemical pathway or a blend of several ones. In mammalian cells, evidence for UPF2- as well as for UPF3-independent NMD has been reported (Gehring et al. 2005; Chan et al. 2007) and it has further been suggested that SMG6 and SMG7 might represent two independent branches of NMD to initiate target RNA degradation (Mühlemann and Lykke-Andersen 2010). Given these uncertainties, we decided to operationally define NMD as an RNA degradation pathway that depends on UPF1 and SMG6 or SMG7. Accordingly, we performed shRNA-mediated knockdowns (KD) in HeLa cells for these three NMD factors and also operated the respective rescues by expressing an RNAi-resistant version of the respective

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



protein (Fig. 1A). To generate a reference and control dataset (Ctrl), a knockdown with a scrambled shRNA sequence was performed. To address the extent of redundancy between SMG6 and SMG7, we also performed double knockdowns (dKD) of both factors and rescued this condition with either SMG6 or SMG7. Western blotting showed efficient depletion of the respective NMD factors in all knockdowns and rescue protein levels were comparable or higher than the endogenous levels of the respective NMD factor (Figure 1A). Checking the relative RNA levels of several previously identified NMD-targeted transcripts showed increased RNA levels under our knockdown conditions and a partial to complete rescue under our rescue conditions (Fig. 1B). Thus, our experimental conditions resulted in the attempted inhibition and at least partial rescue of NMD activity.

**mRNA-seq and principal component analysis**

Three independent biological replicates of the aforementioned 10 different conditions were enriched for poly(A)+ RNA and subjected to high throughput sequencing. The obtained reads were mapped to the human genome (GRCh38) using TopHat (Kim et al. 2013) and gene counting was performed with the program featureCounts (Liao et al. 2014). Since the library preparations and sequencing was carried out in two batches (batch A: Ctrl, UPF1 KD, SMG6 KD, and respective rescue samples; batch B: Ctrl, SMG7 KD, dKD and respective rescue samples), a total of 6 Ctrl reference samples was produced. Although the principle component analysis (PCA) showed the absence of a significant batch effect, indicated by the close clustering of the Ctrl samples (Supplemental Fig. S1), we nevertheless opted to separately compare every sample to the controls of the respective batch. The PCA also revealed that the KD samples of the different factors do not cluster as close to each other as one might have expected, which can have different explanations. One reason is that the PCA maximizes the components of variation between all the samples and are equally influenced by both upregulated (i.e. NMD targeted) and downregulated genes (i.e. non-NMD-related effects), of which the downregulated ones are not expected to be conserved for different NMD factors. It should also be noted that the first two

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



principal components only report a fraction of the total variation present in the dataset (in this case 49%). Additionally, UPF1 is known to be involved in several biological processes other than NMD (Isken and Maquat 2008), which may explain why UPF1 KDs cluster farther away from the other samples. The fact that the rescue samples are all closer to their respective KDs than to the control (Ctrl) confirms our observation from individual transcripts (Fig. 1B) that we only achieved a partial rescue in our experiments, despite of an overall higher-than-endogenous expression of the RNAi-resistant constructs (Fig. 1A and Supplemental Table S1). This can be attributed to the fact that a puromycin selection marker on the shRNA encoding plasmids enabled us to achieve the knockdown in essentially every cell surviving the selection, which was not the case for the cells expressing the rescue construct. Therefore a fraction of the cells in the rescue conditions was depleted for the endogenous NMD factor yet did not express the corresponding rescue construct, resulting in the observed partial rescue.

**Bioinformatics approach to identify *bona fide* NMD targeted genes**

A first differential expression analysis was conducted at the gene level using DESeq2 (Love et al. 2014) to compute log2 fold changes (log2FC) between two conditions. To represent the knockdown and the rescue effect for each gene as a single value, the KD/Ctrl log2FC and the inverse of the rescue/KD log2FC were averaged, thus resulting in a positive value for NMD targets (see methods). The respective *p*-values of these two comparisons were combined using a method called “sum of *p*-value” (Edgington 1972) (Supplemental Fig. S2A), which allows the detection of differentially expressed genes with enhanced sensitivity and confidence. The overall good negative correlation between these two log2FC values justifies this approach (Supplemental Fig. S2B). From here on, when describing targets of e.g. UPF1 or the dKD rescued with SMG6, we will refer exclusively to these combined log2FC (KD-rescue log2FC) and *p*-values (KD-rescue *p*-values). Using this approach and Fisher’s method (see Supplemental Methods), we computed a global list of significant differentially expressed genes (Supplemental Table S2). Since the meta-analysis procedure tends to inflate the number of significant results and

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



since it was already observed that RNA-seq methods cannot control their true false discovery rate (FDR) with few replicates (Soneson and Delorenzi 2013), we decided to focus on the top 1000 most significant genes of our analysis, defining this as our high-confidence set of NMD targets. A comparison with the transcripts annotated as NMD sensitive in the Ensembl database proved the benefit of this combinatorial approach over relying on data from individual knockdown or rescue conditions (Supplemental Fig. S3). This comparison also revealed that the rescue conditions (log2FC rescue/KD) generally identified more annotated NMD transcripts than the corresponding knockdown conditions (log2FC KD/Ctrl), despite being only partial (see above). Besides genes that show the pattern expected for NMD targets (RNA increase upon KD and decrease in rescue; see Supplemental Table S3, “positive results”), comparable numbers of genes were affected in the opposite way (RNA decrease upon KD and increase in rescue; see Table S3, “negative results”). However, many more of the former are characterized by a higher change in expression (log2FC > 0.5, Table S3). Negative results are also less shared among the different factors (Supplemental Fig. S4), suggesting that they are more likely to arise from transcriptional noise. These observations together indicate that indirect gene expression variation caused by NMD inactivation is secondary to the changes in RNA levels that can be attributed to active NMD decay. Finally, it's important to underline the fact that the single SMG7 KD inhibited NMD to a much lesser extent than the KD of all other NMD factors, an observation that we already reported in an earlier study (Yepiskoposyan et al. 2011).

**Identified NMD targets show expected properties**

We expected the rescue experiments to increase the accuracy of identifying NMD targeted RNAs because of the reduction of off-target hits. To test the actual benefits of the rescues, we compared our data to a previous published dataset that was also produced by RNA-seq from RNA of HeLa cells (Tani et al. 2012). The correlation of the simple log2FC obtained in UPF1 KD is very low (Supplemental Fig. S5A). This is a common situation when comparing RNA-seq datasets of different publications, which is mainly caused by

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



environmental and technological variables that strongly impact the results. However, the overlap between the two datasets improves considerably when the UPF1 rescue is included in the analysis (Supplemental Fig. S5B). The correlation score increases from 0.06 to 0.21 and the 100 top targets of our meta-analysis (red dots) are also more evidently among the most strongly reacting genes in this comparison.

To further assess the quality of our data, we complemented it with a UPF1 CLIP dataset that we generated by expressing C-terminally FLAG-tagged UPF1 in HeLa cells depleted for the endogenous UPF1. Consistent with previously observed preferential steady-state association of UPF1 with NMD targets (Johansson et al. 2007; Johns et al. 2007; Kurosaki and Maquat 2013; Silva et al. 2008; Hwang et al. 2010; Lee et al. 2015), the top 1000 significant NMD targets are overall enriched in UPF1 CLIP tags (Supplemental Fig. S6), demonstrating a strong correlation between the two datasets.

We have further compared our top NMD targets with the SMG6 cleavage sites determined by Schmidt and colleagues (Schmidt et al. 2015). In this study, the authors employed parallel analysis of RNA ends (PARE) to determine 5' termini of RNA decay intermediates produced by SMG6 and dependent on UPF1. The determined SMG6 targets in this study are strongly enriched in our list of significant NMD targets (Supplemental Fig. S7). The overlap is nevertheless far from complete, as already observed in the original publication (Schmidt et al. 2015).

**UPF1, SMG6 and SMG7 define a homogenous pathway**

To have a comprehensive view of the entire data, we performed a cluster analysis without adding any *a priori* information (Fig. 2A). The most significant cluster (cluster 1), of the two we determined, comprises ~ 40'000 genes and shows no particular trend of differential expression, consistent with the expectation that the majority of poly(A)+ RNAs are not targeted by NMD. The second cluster (cluster 2) comprises ~ 4'000 genes and shows the differential expression pattern expected for NMD targets, in which the log2FC is positive for KD/Ctrl and negative for rescue/KD conditions. This shows that NMD is

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



responsible for the most relevant pattern in our data, having a stronger impact than any other effect or bias. It is also interesting to note that among the less significant additional clusters, none shows any factor-specific trend (Supplemental Fig. S8), highlighting the overall uniformity of our data and indicating the absence of significant sub-pathways, i.e. branched or alternative NMD routes.

One of the key features of our dataset is the extent of overlap between the different NMD factors studied. A high overlap would first of all confirm that our approach indeed yields a high confidence list of NMD targeted transcripts and at the same time justifies the definition of NMD as a UPF1, SMG6 and SMG7-dependent pathway. To get an overview on the extent of overlap in our data, we selected for every condition the top significant 1000 targets and examined what percentage of those targets was also identified in the other conditions (Fig. 2B). Overall, we find an extensive overlap between most conditions. For example, at least 90% of the top 1000 UPF1 targets also showed a positive KD-rescue log2FC in the other conditions, albeit not necessarily significant, with the only exception of SMG7. It is indeed evident that the SMG7 condition correlates least with the other ones by a large margin. At least in part, this is due to a smaller number of genes being statistically significantly affected by SMG7, resulting in a higher proportion of false positive hits among the top 1000 targets, which do not correlate with the targets of the other conditions. It is interesting to observe that SMG7 KD is leading to a notable upregulation of both UPF1 and SMG6 levels (Fig. 2C). This autoregulatory feedback might explain why SMG7 KD only weakly impaired NMD. This effect was also observed, albeit less pronounced, in UPF1 KD but not in SMG6 KD. The comparison of SMG6 and SMG7 single KDs is therefore biased by the autoregulation phenomenon. For this reason, the double KD of these two factors is essential to disentangle their individual contributions to the NMD pathway.

**SMG6 and SMG7 act on the same target genes**

With regards to the previously proposed independent pathways to degrade mammalian NMD targets (Mühlemann and Lykke-Andersen 2010), the

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



extremely high overlap of the most significant targets between dKD SMG6 and dKD SMG7 is intriguing (Fig. 2B). Since these lists are computed taking into account the same dKD and Ctrl samples, they are not independent and their similarity could be overestimated. We therefore directly compared the effects caused by SMG6 and SMG7 rescues in the dKD cells (Fig. 3A). As can be seen from the spindle-shaped cloud of dots, most genes were affected in the same way by the SMG7 rescue and the SMG6 rescue, with an overall stronger effect caused by the SMG6 rescue. This correlation is highly significant (Pearson's correlation coefficient 0.842) and there is no substantial group of genes reacting in only one of the two conditions. To have a more rigorous statistical view on the variance present in this comparison, we also performed some simulations. In these simulations we generated a hypothetical SMG7 rescue dataset, in which we reproduced the changes in RNA levels observed in SMG6 rescue. We could thus measure how much these two rescue conditions would correlate, if there were absolutely no difference in the specificity of the two factors. The correlation of this new variable with the SMG6 log2FC was 0.883 ± 0.001, close to the one observed for the real data. This shows that most of the variation present in this comparison is caused by transcriptional noise and that only a small amount of it is actually caused by the different activity of SMG6 and SMG7. To estimate this additional small variation, we applied some variability to the simulated gene expression and compared the resulting correlation with the observed one (Fig. 3B). This allowed us to estimate the average extent of gene expression specificity between SMG6 and SMG7 to 8.19% ± 0.15. From these analyses we can conclude that SMG7 activity has generally a weaker effect than SMG6, but they act on the same genes.

## Long "non-coding" RNAs, small-RNA host genes and pervasive transcripts are targeted by NMD

To get a first overview on what kind of RNAs we have among our 1000 most significant NMD targets, we categorized them according to their biotype (Fig. 4A). As expected, the majority (78%) of the genes codes for proteins. However, there are also a considerable proportion of various non-coding genes, with the main sub-classes being pseudogenes (9%), long intergenic

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



non-coding RNAs (lincRNAs; 6%) and antisense transcripts (4%). Given that NMD is a translation-dependent process, it might be surprising at first sight that several genes annotated as "non-coding" are affected. However, many pseudogenes are known to give rise to PTC-containing mRNAs (Mitrovich and Anderson 2005) and recent ribosome profiling studies found many transcripts categorized as lincRNAs to be associated with ribosomes (Ingolia et al. 2011; Calviello et al. 2015; Carlevaro-Fita et al. 2016). In few cases, the short polypeptides encoded by these lincRNAs were even detected (Ingolia et al. 2014) thus revealing them as a misnomer. Given their documented evidence for associating with ribosomes, one would in fact predict that these mostly short ORFs, similar to uORFs, would terminate translation in an mRNP context that leads to NMD activation. Supporting this view, we find a strong correlation between the number of predicted ORFs (minimal length of 3 codons) on a non-coding RNA and its likelihood to be identified as an NMD target in our study (Fig. 4B).

An interesting group of genes that is significantly enriched in our data comprises host genes for snoRNAs and miRNAs (Fig 4C). Consistent with a previous study reporting an overrepresentation of snoRNA host genes among NMD targets (Lykke-Andersen et al. 2014), snoRNA host genes are 3-fold enriched among our top 1000 hits, when compared to genes that neither host snoRNAs nor miRNAs (non hosts) (*p*-value = 6e-6, Fisher exact test). Similarly, we also detected a 3-fold enrichment of miRNA host genes (*p*-value = 4e-14). The snoRNAs and miRNAs are encoded in the introns of these genes and processed from the excised introns of the pre-mRNAs. In many cases, the spliced RNAs do not encode a functional protein and the non-conserved short ORFs occurring in these spliced RNAs will presumably be translated and trigger NMD because of aberrant translation termination in the same way as proposed for lincRNAs and mRNAs containing uORFs. Thus, while high transcription rates of the snoRNA and miRNA host genes are required to produce sufficient amounts of these small RNAs, NMD ensures that the spliced host RNAs, which can represent waste products for the cell, are quickly degraded.

In *S. cerevisae*, another group of transcripts has recently been revealed to be abundant among the NMD targets, namely transcripts whose

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



transcription starts upstream of the canonical transcription start sites (TSS) (Malabat et al. 2015). Owing to their additional sequence upstream of the main ORF, they have an increased likelihood to contain uORFs that will activate NMD. To see if such pervasive transcripts are also present among the NMD targets in human cells, we searched in our data for sequence coverage 200 bp upstream of annotated TSS, ignoring intervals where other genetic annotations were present. The log2FCs of these “upstream TSS” sequences, normalized with the expression level of the corresponding gene, follow an NMD-sensitive distribution, even though the effect size is rather small (Fig. 4D). We conclude that although detectable, pervasive transcripts are not commonly present in our data, possibly because pervasive transcription occurs at a much lower frequency in human cells than in yeast.

**Different 3’ UTR features are associated with NMD sensitivity**

In the absence of a detailed understanding of the mechanism of NMD, empirical identification of transcript features that can differentiate NMD targets from non-targets is an important and active area of research. Even though such feature searches only yield correlations without implying a causal connection, they can help characterize the pathway and formulate hypotheses on the molecular mechanism. To have the best-possible correlation between mRNA properties and their level of degradation by NMD, we performed a transcript-level analysis. In this analysis, the expression of all the different splicing isoforms of each gene is estimated independently, providing information on the behaviour of specific mRNA molecules. The difficulty of uniquely assigning reads to single transcripts, however, determines a lower accuracy, compared to a gene-level study (Soneson et al. 2015). The combination of all conditions into a single measure was carried out in the same way as the gene-level analysis.

By far, the most prominent and significant NMD feature in our analysis is the presence of an intron in the 3' UTR located more than 50 nucleotides downstream of the stop codon ($p$-value < 2e-16, Fisher exact test). 40% of the significant targets are characterized by this property (Fig. 5A). This confirms many previous studies and supports the model that an EJC downstream of a

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



stop codon highly facilitates the decay process on the mRNA. To analyze additional features, we focused only on the transcripts that do not contain a 3' UTR intron. This is motivated by the fact that this property can mask the presence of other features. For example, it was reported that mRNAs with an intron in the 3' UTR are more strongly degraded when the 3' UTR is short (Hurt et al. 2013). Focusing on this filtered set we analyzed the presence of uORFs. For this analysis, uORFs were defined as ≥ 3 codon-long ORFs with both an AUG and a stop codon upstream of the main ORF, being aware that *in vivo* not all of these uORFs will actually be translated. Notwithstanding this oversimplification, we observed a highly significant enrichment of uORFs among the NMD targets (*p*-value: 2e-10). All mRNAs with a uORF were similarly discarded from further analyses.

Besides the presence of an intron, additional characteristics of the 3' UTR are also important in determining if an RNA is targeted by NMD (Fig. 5B). In our data, we found that NMD targets show a longer 3' UTR than a control group that we defined as a set of mRNAs with similar Ctrl expression levels as the NMD sensitive ones (*p*-value: 2e-5, permutation test). This feature, despite extensive experimental validation, shows only a limited statistical significance in our data. Nevertheless, with a median 3' UTR length of 836 nucleotides, NMD targets tend to have on average ~ 50% longer 3' UTRs than transcripts of the control group (median length of 561 nucleotides). In addition, the GC content of the 3' UTRs of NMD targets is also significantly higher than in NMD-insensitive transcripts (*p*-value: 2e-10). This finding is in line with the higher UPF1 propensity to GC rich regions we observe in our CLIP experiment (Supplemental Fig. S9). It was suggested that UPF1 ATPase and helicase activity was reduced when associated with GC motifs (Bhattacharya et al. 2000), which could lead to its enrichment on GC-rich 3' UTRs and thereby promote NMD. Furthermore, we also determined the phylogenetic conservation of the 3' UTRs in NMD targets and the control group (PhyloP score) (Pollard et al. 2010) and found that 3' UTRs of NMD targets are significantly less conserved (*p*-value: 7e-13). The biological meaning of this lower phylogenetic conservation of the 3' UTRs of NMD targets is unknown and possible explanations remain speculative at this point (see Discussion).

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



## DISCUSSION

This study presents an attempt to determine a high-confidence set of endogenous transcripts targeted by NMD. Several such attempts have been previously reported, using either microarrays (Mendell et al. 2004; Wittmann et al. 2006; Viegas et al. 2007; Yepiskoposyan et al. 2011) or deep sequencing (Tani et al. 2012; Hurt et al. 2013; Schmidt et al. 2015). However, the overlap among the hits in these different studies was minimal, questioning the robustness of the results. Our approach differs from these previous studies in the high sequencing quality and an experimental design that combines data from 10 different experimental conditions. The rescue conditions substantially improved the accuracy of the differential expression detection by controlling for indirect effects and possible biases introduced by the shRNA procedure (Supplemental Fig. S3). Furthermore, the meta-analysis of UPF1, SMG6 and SMG7 also resulted in increased statistical power and helped to filter out individual false positive hits. In particular, studies investigating only the effect of UPF1 depletion on the transcriptome are prone to yield many false positives because UPF1 is involved in additional pathways beside NMD (Isken and Maquat 2008). Finally, the SMG6/SMG7 double KDs were crucial to reveal the redundancy in the activity of SMG6 and SMG7. The single SMG7 KD caused a 77% upregulation of SMG6 mRNA (Fig. 2C), which prevents an unbiased evaluation of the extent of SMG7 activity in normal cells, since many of its targets are likely masked by an increased SMG6 activity. If instead the cells were depleted of both factors, it became evident that rescue plasmids of either gene had very similar effects on the transcriptome. Although the magnitude of these changes was clearly higher for SMG6, we determined that SMG7 acts on essentially the same targets, with an estimated variability of only 8.2% (Fig. 3B). We therefore conclude that SMG6- and SMG7-mediated degradation routes appear to be two highly redundant branches of the mammalian NMD pathway. This is consistent with previous observations made with reporter genes (Luke et al. 2007; Jonas et al. 2013; Metze et al. 2013) and in line with what was observed in a study focusing on SMG6 endonucleolytic activity (Schmidt et al. 2015). In this work, the authors observed an accumulation of decapped transcripts upon depletion

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



of SMG6, indicating the presence of a complementary decay mechanism that most likely involves SMG7 mediated recruitment of the CCR4/NOT deadenylase followed by decapping of the deadenylated RNAs.

Our approach allowed us to confirm and expand several previous observations. For example the GAS5 transcript, which is not associated to any known peptide sequence, has been previously discovered to be stabilized upon NMD inactivation (Weischenfeldt et al. 2008; Tani et al. 2013). In our study, we found that a substantial fraction of the genes affected by NMD are associated in the databases to a non-coding biotype (Fig. 4A). The dependency of NMD on translation indicates that this classification as “non-coding” may be inaccurate and that in reality these RNAs engage the translation machinery. This hypothesis is supported by a clear correlation between the number of possible ORFs on these transcripts and the likelihood of being subject to NMD (Fig. 4B) and by the fact that recent ribosome profiling studies and polysome analyses found many non-coding RNAs to be associated with ribosomes (Ingolia et al. 2011, 2014; Carlevaro-Fita et al. 2016).

In our list of NMD targets, we found a significant enrichment of snoRNA and miRNA host genes (Fig. 4C). SnoRNA host genes have already been described to be frequently targeted by NMD and it was proposed that this uncouples the expression of the snoRNAs and the corresponding host gene (Lykke-Andersen et al. 2014). The same regulation appears to apply for miRNA host genes, which we also show to frequently undergo NMD. We speculate that from many miRNA host genes a cell only requires high numbers of the specific mature miRNA but not of the cognate spliced transcript. Splicing the pre-mRNA of a miRNA host gene to an NMD-sensitive transcript ensures low levels of that transcript despite a high transcription rate of the gene.

In yeast, it has been shown that RNA polymerase II transcription often initiates upstream the usual transcription start site (a phenomenon called pervasive transcription) and it was shown that NMD plays an important role in clearing these spurious transcripts (Malabat et al. 2015). In our data we could also observe a similar activity, even though the phenomenon appeared to be much less common than in yeast (Fig. 4D). It should however be noted

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



that Malabat and colleagues used an experimental method aimed at specifically identifying transcriptional start sites (TSS sequencing), which is much more sensitive in detecting even very low abundant pervasive transcripts than a normal RNA-seq like ours. Nevertheless, our data indicates that unlike in yeast, such pervasive transcripts only constitute a small fraction of the NMD-targeted transcriptome in mammalian cells, presumably because the frequency of spurious transcription initiation is much lower than in yeast.

An accurate list of *bona fide* NMD targets may help to uncover common features among NMD-sensitive transcripts, which in turn can give further insights into NMD target identification and eventually allow the computational prediction of NMD targets. The most prominent feature we could determine in our data is, as expected, the presence of an intron in the 3' UTR farther than 50 nucleotides downstream of the TC (Fig. 5A). The NMD-stimulating effect of EJCs is well known and characterized (Karousis and Mühlemann 2016). However, there is a significant portion of transcripts among the NMD targets that lacks a 3’ UTR intron, proving the existence of other NMD-triggering signals. Among the NMD targets without 3’ UTR introns we could observe a significant enrichment for the presence uORFs as the second most relevant feature. If we then focus on targets without either of these two characteristics, we observe longer 3' UTRs to be significantly correlated with NMD susceptibility (Fig. 5B). Models proposing a common mechanism through which these different features lead to NMD have been put forward (Amrani et al. 2004; Stalder and Mühlemann 2008; Schweingruber et al. 2013) and there is ample supporting evidence for them. In quantitative terms however, we want to emphasize that the presence of a 3’ UTR intron is by far the most important criteria with the most predictive power, whereas the predictive power of uORFs and long 3’ UTRs is rather limited. Many transcripts with long 3' UTRs or predicted uORFs are in fact known to escape NMD and for some of them the NMD-protecting factors are known. For example, PTBP1 binding to the RNA stability element in Rous sarcoma virus protects the viral genomic RNA from NMD by preventing interaction of UPF1 with the 200 nucleotide region downstream of the TC (Ge et al. 2016). Furthermore, a high AU content within the first 200 nucleotides downstream of the TC of several mRNAs with long 3’ UTRs has also been reported to confer

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



resistance to NMD, but no trans-acting factor was identified (Toma et al. 2015). These findings point towards a complex interplay between NMD-promoting and NMD-inhibiting determinants that to a large extent still remain to be elucidated.

In line with AU-rich elements in the 3' UTR correlating with NMD resistance, the GC content in the 3' UTRs of our 660 NMD targets lacking a uORF or 3' UTR intron was significantly higher than in a control group of NMD-immune transcripts. This finding is also consistent with our CLIP data, showing a higher propensity of UPF1 for GC-rich regions, and with recent reports of a significant enrichment for guanosine residues in UPF1 binding regions (Hurt et al. 2013). G-rich and GC-rich sequences have a higher propensity to form secondary structures and it has been speculated that such secondary structure might slow down the helicase/translocase activity of UPF1, thereby resulting in its enrichment in these regions (Hurt et al. 2013).

Finally, analysis of the phylogenetic conservation of the 3' UTR sequences revealed a significantly lower conservation of the 3' UTRs of NMD targets compared to NMD-insensitive 3' UTRs (Fig. 5B). We can currently only speculate about the biological meaning of this intriguing finding. Novel transcripts, arising for example from gene duplications, transposons insertions or viral infections, could often be detrimental for the cell. We hypothesize that, at this stage, the transcripts often would present features that render them susceptible to NMD, which in fact may be beneficial for the cell. If however such a transcript acquires a new function that educes a selective pressure for increased gene expression, transcript variants escaping NMD will now confer a selective advantage. This could occur for example by evolving binding sites in the 3' UTR that would protect the transcripts from NMD (as discussed above). Such NMD-avoiding motifs would then be conserved in the future, since they provide an evolutionary advantage. According to this view, NMD might have an important role for the evolution of genomes in that it enables cells to entertain an evolutionary playground by reducing the detrimental effects that could be caused by young and not yet fully functional genes. This scenario could explain our observation that younger genes appear to be more susceptible to NMD than evolutionary more ancient ones. Notably, the recent observation that some RNA virus genomes are recognized and degraded by

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



NMD would be consistent with this scenario (Balistreri et al. 2014; Garcia et al. 2014).

In summary, we believe that the set of endogenous NMD-targeted transcripts that we have identified herein will provide a highly valuable resource and reference for the scientific community for further investigations into both the biological role and the mechanism of NMD.

# MATERIALS AND METHODS

## Experimental methods

For knockdowns, $2x10^5$ HeLa cells were seeded into six-well plates and 24 hours later the cells were transiently transfected using Dogtor (OZ Biosciences). For single factor knockdowns, 400 ng of pSUPERpuro plasmids expressing shRNAs against UPF1, SMG6, SMG7 or control plasmids were transfected. For the knockdown and rescue conditions 400 ng pcDNA3-NG-UPF1-WT-Flag, pcDNA3-SMG6-FLAG or pcDNA3-SMG7-FLAG were included in the transfection mixtures. For double knock-down experiments 400 ng of each pSUPERpuro plasmid was added and the rescue of each factor was achieved by including 400 ng of pcDNA3-SMG6-FLAG or pcDNA3-SMG7-FLAG accordingly. The cells were split into a T25-$cm^2$ cell culture flask and selected with puromycin at a concentration of 1.5 µg/µL. 24 hours prior to harvesting the cells were washed with PBS and the puromycin-containing medium was exchanged with normal DMEM-FCS medium. Cells were harvested 4 days after transfection.

The shRNA target sequence for UPF1 and SMG6 were described in (Paillusson et al. 2005) and SMG7 was described in (Metze et al. 2013). Total RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich).

Cell harvesting for protein samples (derived from the same sample as RNA preparation) and measurement of relative mRNA levels by reverse transcription quantitative polymerase chain reaction (RT-qPCR) were done as described in (Nicholson et al. 2012). Briefly, $2x10^5$ cell equivalents were analyzed on a 10% PAGE and detection was performed using Anti-RENT1

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



(UPF1) (Bethyl, A300-038A), anti- EST1 (SMG6) (Abcam, ab87539), Anti-SMG7 (Bethyl, A302-170A) and Anti-CPSF73 (custom made) antibodies.
qPCR assays have been described elsewhere (Yepiskoposyan et al. 2011), except for the assays to measure the following genes:
GAS5 (5´-GCACCTTATGGACAGTTG-3´, 5´-GGAGCAGAACCATTAAGC-3´)
CDKN1A (5´-GACCAGCATGACAGATTTCTAC3´, 5´-CAAACTGAGACTAAG GCAGAAG), TMEM183A (5´-TGCTCCGGCCGAGTGA-3´, 5´-ACCGCCGGAT CCGAGTT-3´), RP9P (5´- CAAGCGCCTGGAGTCCTTAA-3´, 5´-AGGAGGTTT TTCATAACTCGTGATCT-3´), GADD45B (5´-TCAACATCGTGCGGGTGTCG-3´, 5´-CCCGGCTTTCTTCGCAGTAG-3´), ATF4 (5´- TCAACATCGTGCGGGT GTCG-3´, 5´- CCCGGCTTTCTTCGCAGTAG-3´)
A total of 33 samples were sequenced: control knockdowns (Ctrl) in 6 replicates, all other conditions in triplicates. TruSeq Stranded mRNA kit (chemistry v3) was used in the preparation of the library and in the poly(A) enrichment step. The first batch was sequenced on an Illumina HiSeq2500 and the second on an Illumina HiSeq3000 machine. Reads are single-end and 100bp long. The sequencing depth of every sample is reported in Supplemental Table S4.

**UV cross-linking and immunoprecipitation (CLIP) of UPF1-Flag**

Knockdown of endogenous UPF1 was induced in HeLa tTR-KRAB-shUPF1 cells (Metze et al. 2013) by addition of 5 µg/mL doxycycline and $8x10^6$ cells were transiently transfected with 4 µg of a pcDNA3 expression plasmid encoding a C-terminally Flag-tagged, RNAi-resistant version of UPF1 using 30 µL of Lipofectamine 2000. 44 hours post transfection, cells were washed and cross-linked in ice-cold PBS applying 150 $mJ/cm^2$ UV-C light (Bio-Link BLX-E, 254 nm). After irradiation, cells were scraped of the culture dish, collected by centrifugation, flash-frozen in liquid nitrogen and stored at -80 °C. After cell lysis in 3 mL hypotonic lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2 mM EDTA, 0.5% (v/v) Triton X-100, Halt Protease Inhibitor Cocktail) and removal of cell debris by centrifugation, the supernatant was adjusted to

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



160 mM NaCl and incubated with 30 U RNase I (Ambion) and 15 U Turbo DNase (Ambion) at 37 °C for 7.5 minutes. 160 µL Dynabeads Protein G were incubated with 18 µg of mouse anti-FLAG M2 antibody (Sigma Aldrich), washed and resuspended in 1 mL hypotonic lysis buffer and incubated with the cell lysate at 4 °C for 1.5 hours. The beads were then washed three times with IP-buffer (50 mM HEPES-NaOH pH 7.5, 300 mM KCl, 0.05% (v/v) NP-40, Halt Protease Inhibitor Cocktail). To label the co-precipitated RNA fragments, dephosphorylation with Antarctic phosphatase was followed by incubation with 22.5 µL γ-32P-ATP (10 mCi/mL), 5 µL 100 mM ATP, and 100 U T4 Polynucleotide Kinase in a total volume of 150 µL at 37 °C for 45 minutes. The protein-RNA adducts were heat-eluted from the beads, resolved at 70 °C on a 2x NuPAGE Novex buffer 4-12% Bis-Tris Midi Gel (Life Technologies), transferred to nitrocellulose membrane using the iBlot system, and visualized by phosphorimager scanning. The section of the membrane harboring the RNA-UPF1 adducts was excised and the RNA fragments were retrieved by Proteinase K digestion followed by phenol/chloroform extractions and ethanol precipitation. cDNA library preparations and Illumina sequencing was performed at Fasteris (Geneva, Switzerland) according to their standard small RNA sequencing protocol. All bioinformatics analyses were performed in the same way as with the RNA-seq data, which is described in the following paragraphs.

**Gene counting and differential expression analysis**

For the gene-level analysis, sequencing reads were processed with Trimmomatic (Bolger et al. 2014) to remove low quality regions, poly(A) tails and adapter sequences. The reads were then mapped to the human genome (GRCh38) with TopHat (Kim et al. 2013), version 2.0.13. This step was aided by the use of the Ensembl gene annotation, release 81. The gene counting was performed with the program featureCounts (Liao et al. 2014), version 1.4.6. DESeq2 (Love et al. 2014) was employed for the differential expression computation, version 1.6.3. All comparisons were corrected with the sva package, version 3.12.0, in order to compensate for secondary biases in the data. The transcript-level analysis was based on the isoform abundance

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



estimation provided by RSEM (Li and Dewey 2011), version 1.2.19. The annotation used was Ensembl, release 84.

**Meta-analysis of the data**

Our final goal is to provide a unique score for every gene, to estimate its likelihood of being an NMD target. The first step to achieve such synthetic result was to combine KD and rescue conditions for every factor. A joint gene-specific log2FC value was generated by computing the average between the KD/Ctrl log2FC and the inverse of the rescue/KD log2FC. So if a gene is upregulated in UPF1 KD compared to Ctrl and it's downregulated in the rescue compared to the KD will have a high positive combined log2FC. We called this quantity KD-rescue log2FC. The significance of this combined log2FC was computed by a technique called sum of *p*-value (Edgington 1972). All genes downregulated in a KD or upregulated in a rescue, were assigned a p-value of 1 before applying this algorithm. We called this quantity KD-rescue *p*-value. In all cases in which we will refer to the significance or the log2FC of a single condition, like UPF1 or dKD_SMG6, we refer to these meta-analysis computations. Next, we aimed at finding the genes that complied with our definition of NMD target: gene reacting to UPF1 and at least one between SMG6 and SMG7. We therefore combined all conditions in a single list of significant results, using a set of *p*-value meta-analysis methods (Supplemental Table S2). The results from SMG6 and dKD_SMG6 were combined with Fisher's method in a single meta_SMG6 score. The same comparison was done for SMG7. A meta_SMGs significance score was then computed with a sum of *p*-value from meta_SMG6 and meta_SMG7. The final significance parameter used to determine the list of most significant NMD targets was calculated with a Fisher's method from meta_SMGs and UPF1_FDR (meta_meta).

**Differential expression simulations**

To rigorously compare two perturbed transcriptional conditions we performed a number of simulations. First, we recreated a theoretical dataset in which the transcriptional changes were based on the differences observed between

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



SMG6 rescue and the dKD. We generated gene counts by sampling from a negative binomial distribution whose parameters were estimated with DESeq2. This simulation was meant to produce a hypothetical SMG7 rescue condition that behaved in the same exact way as SMG6 rescue. The only exception was that the intensity of the log2FC was decreased by a factor of 0.57, which was the difference we estimated in SMG6 and SMG7 intensity from a linear model of the real data. Since this initial simulation showed a correlation with SMG6 rescue log2FC slightly higher than the observed one for SMG7, we executed a series of additional simulations to which we added more variability. The expected log2FC were multiplied by a confounding factor of different intensities. This method allowed us to estimate the percent level of specificity between the transcriptional effects of SMG6 and SMG7.

## DATA AVAILABILITY

All sequencing data from this study are available on the Gene Expression Omnibus (GEO) with the id: GSE86148. In an attempt to allow complete reproducible research, all scripts used in this work are available online on GitHub at the address: https://github.com/Martombo/NMDseq.

## ACKNOWLEDGMENTS

We are grateful to David Zünd and Laurent Gillioz for their contributions during the initial phase of the project and to Michèle Ackermann and Muriel Fragnière of the NGS platform of the University of Bern for cDNA library preparations and sequencing. This work has been supported by the NCCR RNA & Disease funded by the Swiss National Science Foundation (SNSF), by SNFS grants 31003A-143717 and 31003A-162986 to O.M., and by the canton of Bern (University intramural funding to O.M.). M.C. was supported by a grant from the NOMIS Foundation.

## REFERENCES

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.

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



Balistreri G, Horvath P, Schweingruber C, Zünd D, McInerney G, Merits A, Mühlemann O, Azzalin C, Helenius A. 2014. The host nonsense-mediated mRNA decay pathway restricts mammalian RNA virus replication. *Cell Host Microbe* **16**: 403–411.

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.

Bhattacharya A, Czaplinski K, Trifillis P, He F, Jacobson A, Peltz SW. 2000. Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* **6**: 1226–1235.

Boehm V, Haberman N, Ottens F, Ule J, Gehring NH. 2014. 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep* **9**: 555–568.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

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.

Burroughs a M, Ando Y, de Hoon MJL, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y, et al. 2010. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res* **20**: 1398–410.

Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U. 2015. Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods* **13**: 1–9.

Carlevaro-Fita J, Rahim A, Vardy LA, Johnson R, Guigó R, Vardy LA, Johnson R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22**: 867-821.

Carter MS, Doskow J, Morris P, Li SL, Nhim RP, Sandstedt S, Wilkinson MF. 1995. a Regulatory Mechanism That Detects Premature Nonsense Codons in T-Cell Receptor Transcripts in-Vivo Is Reversed By Protein-Synthesis Inhibitors in-Vitro. *J Biol Chem* **270**: 28995–29003.

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



Chakrabarti S, Bonneau F, Schüssler S, Eppinger E, Conti E. 2014. Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. *Nucleic Acids Res* **42**: 9447–9460.

Chan W-K, Huang L, Gudikote JP, Chang Y-F, Imam JS, MacLean J a, Wilkinson MF. 2007. An alternative branch of the nonsense-mediated decay pathway. *EMBO J* **26**: 1820–1830.

Eberle AB, Lykke-Andersen S, Mühlemann O, Jensen TH. 2009. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* **16**: 49–55.

Eberle AB, Stalder L, Mathys H, Orozco RZ, Mühlemann O. 2008. Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. *PLoS Biol* **6**: 849–859.

Edgington ES. 1972. An Additive Method for Combining Probability Values from Independent Experiments. *J Psychol* **80**: 351–363.

Garcia D, Garcia S, Voinnet O. 2014. Nonsense-mediated decay serves as a general viral restriction mechanism in plants. *Cell Host Microbe* **16**: 391–402.

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**: 1–25.

Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE. 2005. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell* **20**: 65–75.

Hansen KD, Lareau LF, Blanchette M, Green RE, Meng Q, Rehwinkel J, Gallusser FL, Izaurralde E, Rio DC, Dudoit S, et al. 2009. Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in Drosophila. *PLoS Genet* **5:** e1000525.

He F, Jacobson A. 2015. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu Rev Genet* **49**: 339–366.

Hogg JR, Goff SP. 2010. Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* **143**: 379–389.

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



Huntzinger E, Kashima I, Fauser M, Saulière J, Izaurralde E. 2008. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14**: 2609–17.

Hurt J a, Robertson AD, Burge CB. 2013. Global analyses of UPF1 binding and function reveals expanded scope of nonsense-mediated mRNA decay. *Genome Res* **23**: 1636–1650.

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. *Mol Cell* **39**: 396–409.

Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJS, Jackson SE, Wills MR, Weissman JS. 2014. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep* **8**: 1365–1379.

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.

Isken O, Maquat LE. 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet* **9**: 699–712.

Ivanov P V, Gehring NH, Kunz JB, Hentze MW, Kulozik AE. 2008. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J* **27**: 736–747.

Ivanov A, Mikhailova T, Eliseev B, Yeramala L, Sokolova E, Susorov D, Shuvalov A, Schaffitzel C, Alkalaeva E. 2016. PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Res*. **44**: 7766-76.

Johansson MJO, He F, Spatrick P, Li C, Jacobson A. 2007. Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc Natl Acad Sci U S A* **104**: 20872–7.

Johns L, Grimson A, Kuchma SL, Newman CL, Anderson P. 2007. Caenorhabditis elegans SMG-2 Selectively Marks mRNAs Containing Premature Translation Termination Codons. *Mol Cell Biol* **27**: 5630–5638.

Jonas S, Weichenrieder O, Izaurralde E. 2013. An unusual arrangement of two 14-3-3- like domains in the SMG5-SMG7 heterodimer is required for

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



efficient nonsense-mediated mRNA decay. *Genes Dev* **27**: 211–225.

Karousis ED, Mühlemann O. 2016. Nonsense-mediated mRNA decay□: novel mechanistic insights and biological impact. *Wiley Interdiscip Rev RNA* **7**: 661–682.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Kurosaki T, Li W, Hoque M, Popp MWL, Ermolenko DN, Tian B, Maquat LE. 2014. A Post-Translational regulatory switch on UPF1 controls targeted mRNA degradation. *Genes Dev* **28**: 1900–1916.

Kurosaki T, Maquat LE. 2013. Rules that govern UPF1 binding to mRNA 3' UTRs. *Proc Natl Acad Sci U S A* **110**: 3357–62.

Lee SR, Pratt GA, Martinez FJ, Yeo GW, Lykke-Andersen J. 2015. Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. *Mol Cell* **59**: 413–425.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

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

Losson R, Lacroute F. 1979. Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc Natl Acad Sci U S A* **76**: 5134–5137.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Luke B, Azzalin CM, Hug N, Deplazes A, Peter M, Lingner J. 2007. Saccharomyces cerevisiae Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. *Nucleic Acids Res* **35**: 7688–7697.

Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, Jensen

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



TH. 2014. Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* **28**: 2498–2517.

Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**: 665–677.

Malabat C, Feuerbach F, Ma L, Saveanu C, Jacquier A. 2015. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* **4**: e06722.

Maquat LE, Kinniburgh a J, Rachmilewitz E a, Ross J. 1981. Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. *Cell* **27**: 543–553.

Mendell JT, Sharifi N a, Meyers JL, Martinez-Murillo F, Dietz HC. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* **36**: 1073–1078.

Metze S, Herzog VA, Ruepp M-D, Mühlemann O. 2013. Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. *RNA* **19**: 1432–48.

Mitrovich QM, Anderson P. 2005. mRNA surveillance of expressed pseudogenes in C. elegans. *Curr Biol* **15**: 963–967.

Mühlemann O, Jensen TH. 2012. mRNP quality control goes regulatory. *Trends Genet* **28**: 70–77.

Mühlemann O, Lykke-Andersen J. 2010. How and where are nonsense mRNAs degraded in mammalian cells? *RNA Biol* **7**: 28–32.

Nicholson P, Joncourt R, Mühlemann O. 2012. Analysis of nonsense-mediated mRNA decay in mammalian cells. *Curr Protoc Cell Biol* **55**: 4–61.

Nicholson P, Josi C, Kurosawa H, Yamashita A, Mühlemann O. 2014. A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res* **42**: 9217–9235.

Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, Ohno S. 2012. N-and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res* **40**: 1251–1266.

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



Paillusson A, Hirschi N, Vallan C, Azzalin CM, Mühlemann O. 2005. A GFP-based reporter system to monitor nonsense-mediated mRNA decay. *Nucleic Acids Res* **33**: 1–12.

Peixeiro I, Inácio Â, Barbosa C, Silva AL, Liebhaber SA, Romão L. 2012. Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. *Nucleic Acids Res* **40**: 1160–1173.

Pollard KS, Hubisz MJ, Siepel A. 2010. Detection of Non-neutral Substitution Rates on Mammalian Phylogenies. *Genome Res* **20**: 110–121.

Rehwinkel J, Letunic I, Raes J, Bork P, Izaurralde E. 2005. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA* **11**: 1530–1544.

Schmidt SA, Foley PL, Jeong DH, Rymarquis LA, Doyle F, Tenenbaum SA, Belasco JG, Green PJ. 2015. Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. *Nucleic Acids Res* **43**: 309–323.

Schweingruber C, Rufener SC, Zünd D, Yamashita A, Mühlemann O. 2013. Nonsense-mediated mRNA decay - Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta - Gene Regul Mech* **1829**: 612–623.

Silva AL, Ribeiro P, Inácio Â, Ina N, Roma SA. 2008. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. **14**: 563–576.

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 Biol* **6**: 860–871.

Soneson C, Delorenzi M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**: 91.

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**: 1521.

Stalder L, Mühlemann O. 2008. The meaning of nonsense. *Trends Cell Biol* **18**: 315–321.

Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y,

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



Akimitsu N. 2012. Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol* **9**: 1370–9.

Tani H, Torimura M, Akimitsu N. 2013. The RNA Degradation Pathway Regulates the Function of GAS5 a Non-Coding RNA in Mammalian Cells. *PLoS One* **8**: 1–9.

Thermann R, Neu-Yilik G, Deters A, Frede U, Wehr K, Hagemeier C, Hentze MW, Kulozik AE. 1998. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* **17**: 3484–3494.

Toma KG, Rebbapragada I, Durand S, Lykke-Andersen J. 2015. Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *RNA* **21**: 887–97.

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.

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 Dev* **22**: 1381–1396.

Wittmann J, Hol EM, Jäck H-M. 2006. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol* **26**: 1272–87.

Yamashita A, Ohnishi T, Kashima I, Taya Y, Ohno S. 2001. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev* **15**: 2215–2228.

Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Mühlemann O. 2011. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *Rna* **17**: 2108–2118.

Zünd D, Gruber AR, Zavolan M, Mühlemann O. 2013. Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3'

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



UTRs. *Nat Struct Mol Biol* **20**: 936–943.

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



# FIGURE LEGENDS

**FIGURE 1.**

Monitoring of UPF1, SMG6 and SMG7 knockdown (KD), double knockdown (dKD) and rescue experiments compared to a control knockdown (Ctrl).

A) Lysates corresponding to $2x10^5$ cell equivalents of HeLa cells transiently transfected with the indicated knockdown and rescue constructs were analyzed by western blotting. After electrophoretic separation of the proteins on 10% SDS-PAGE and transfer to nitrocellulose membranes, membrane sections were incubated with antibodies against UPF1, SMG6, SMG7 and CPSF73, the latter serving as loading control. The anti-SMG7 antibody gives a double band of which only the upper band (arrow) corresponds to SMG7. B) Relative mRNA levels of known endogenous NMD-targeted mRNAs (GAS5, RP9P, SMG5, ATF4), normalized to β-actin mRNA (ACTB), were determined for all conditions 72 hours post transfection by RT-qPCR. Mean values and standard deviations of three independent experiments are shown, with the samples in the control knockdown (Ctrl) set to 1.0.

**FIGURE 2.**

High overlap of putative NMD targets identified in the different conditions.

A) Bar plot displaying the results of a k-means clustering procedure performed on the log2FC measured for each gene in the indicated conditions. The number of clusters was set to two. The y-axis shows the average log2FC of all the genes present in cluster 1 (comprising 40’000 genes) and cluster 2 (4’000 genes). KD refers to the log2FC (KD/Ctrl) and rescue to the log2FC (rescue/KD), respectively. The double knockdown of SMG6 and SMG7 (dKD) was rescued either with SMG6 (dKD SMG6 rescue) or with SMG7 (dKD SMG7 rescue). B) Histogram of genes with positive KD-rescue log2FC, which is expected from NMD targets. The top 1000 targets in each of our conditions were determined. Each of these sets corresponds to a cluster of columns in the plot. The y-axis shows the fraction of these targets that have a positive log2FC in the other conditions. The log2FC used for this analysis is the average between the log2FC (KD/Ctrl) and the inverse of the log2FC

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press



(rescue/KD) (see methods). C) Bar plots showing the RNA levels of the three factors under study upon each single KD.

**FIGURE 3.**

SMG6 and SMG7 dKD and individual rescues reveal the highly redundant activity of these two NMD factors.

A) Scatter plot comparing SMG6 and SMG7 rescues from the dKD. The picture shows the log2FC of the analysis of SMG6 rescue vs dKD (x-axis) and SMG7 rescue vs dKD (y-axis). Coloured in red are genes significantly downregulated in *either* of the two conditions. The histograms on the y-axis (on the right side) and x-axis (on the top) show the SMG7 log2FC distribution of all the significant downregulated targets in SMG6 rescue and *vice versa*, respectively. B) Simulation to estimate the variation between SMG6 and SMG7 results. Based on the negative binomial parameters computed from our data, new counts datasets were simulated. Additional variation was added to provide an accurate estimate of the individual difference between dKD_SMG6 and dKD_SMG7 rescues. The picture compares the correlation scores (y-axis) found in the simulations at different levels of variation (x-axis) with the observed one in our dataset (black horizontal line).

**FIGURE 4.**

NMD targets transcripts classified as non-coding, small-RNA host RNAs and products of pervasive transcription.

A) Pie chart illustrating the top 1000 NMD targets categorized according to their biotype. 78% of these NMD targets code for protein, 9% are pseudogenes, 6% lincRNAs, and 4% antisense transcripts. B) The number of possible ORFs in non-coding RNAs correlates with the likelihood of undergoing NMD. The expressed non-coding genes are partitioned in different bins depending on how many theoretical ORFs can be predicted on their sequence. The y-axis reports the percentage of NMD targets (top 1000) of all the genes in each bin (e.g. 8% of genes with 3 ORFs are NMD targets). C) Top NMD targets are enriched in snoRNA and miRNA host genes. Each bar shows the percentage of genes that are among the NMD targets (top 1000) in every class of genes. D) Transcripts initiating upstream of the canonical

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press

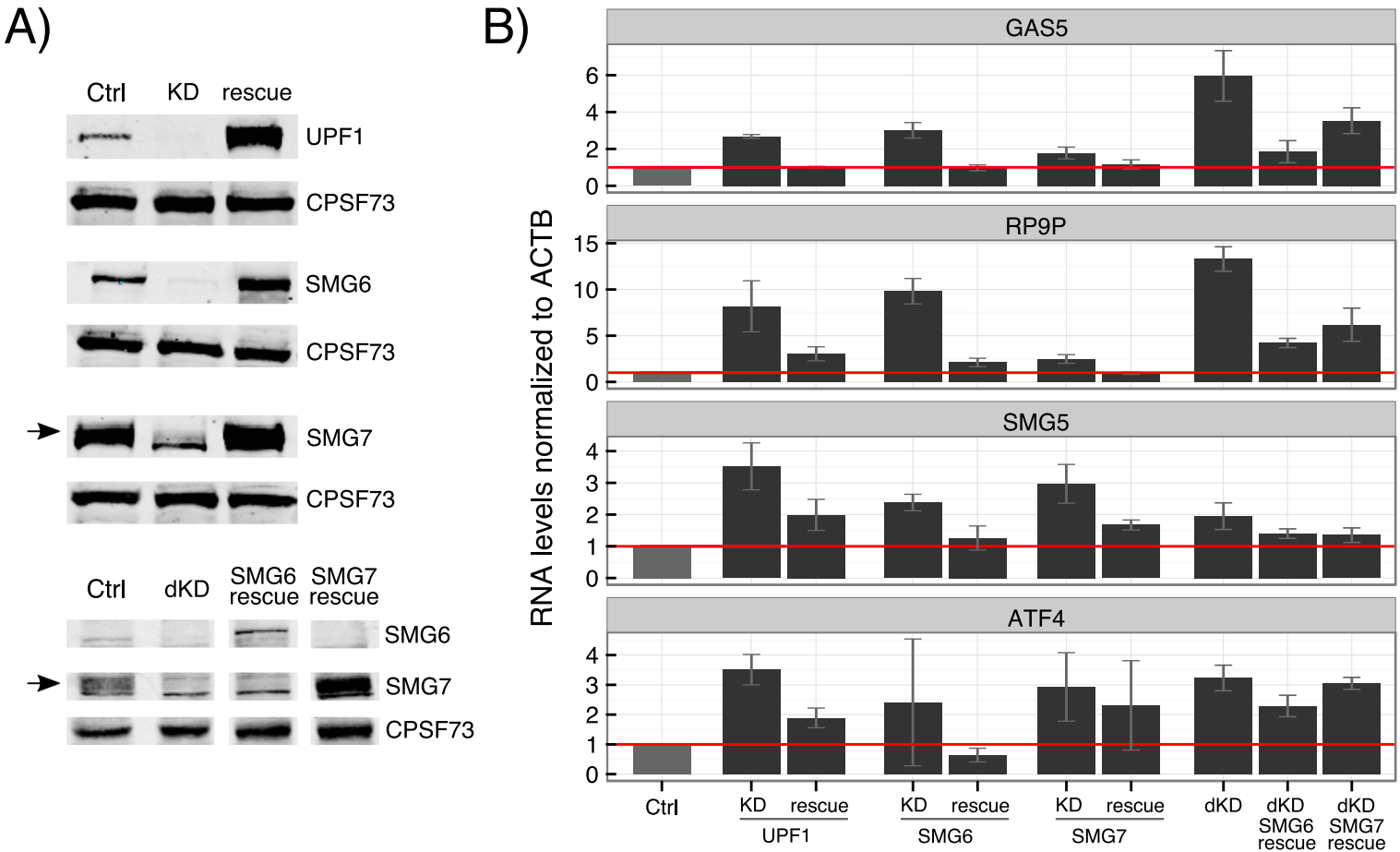


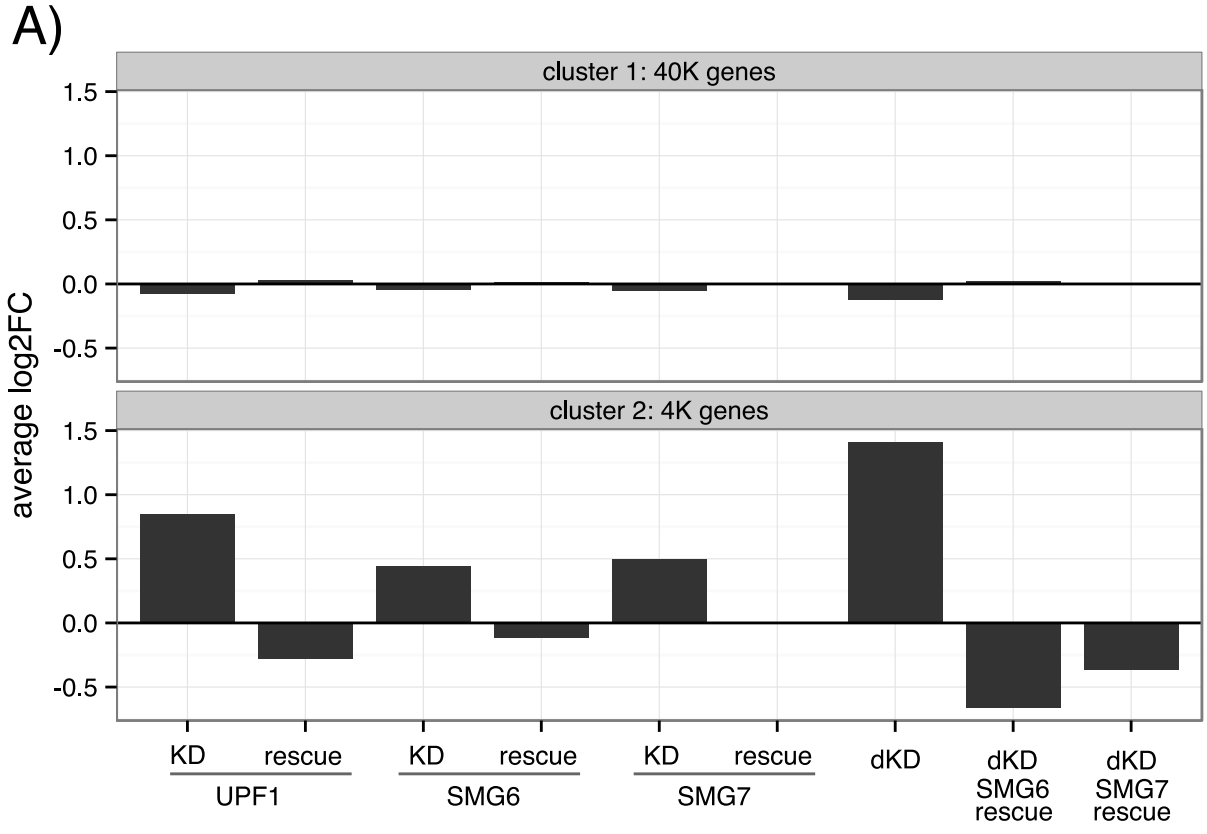
transcription start site (TSS) are partially cleared by NMD. The number of reads upstream of every annotated TSS has been computed in the indicated conditions. This quantity was divided by the total counts of every gene and every condition was analyzed comparing KDs to Ctrl and rescues to KDs, as in the normal analysis (see methods). A box plot showing the log2FC of these quantities is displayed.

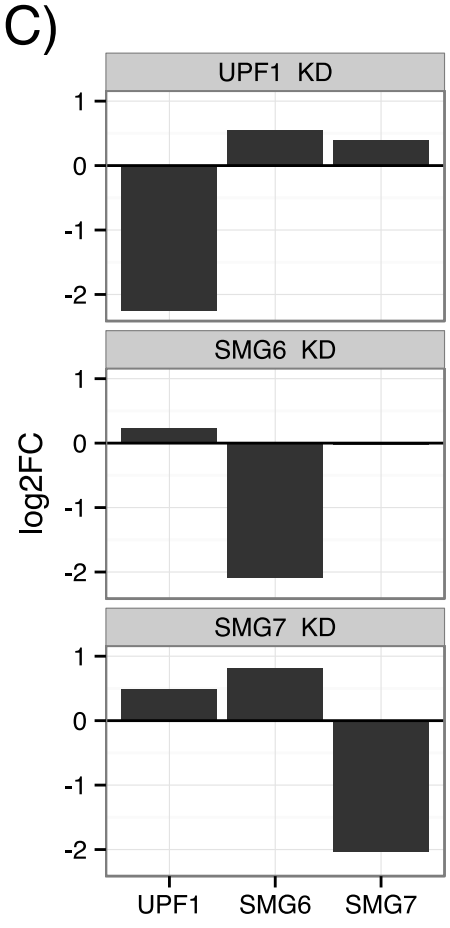
**FIGURE 5.**

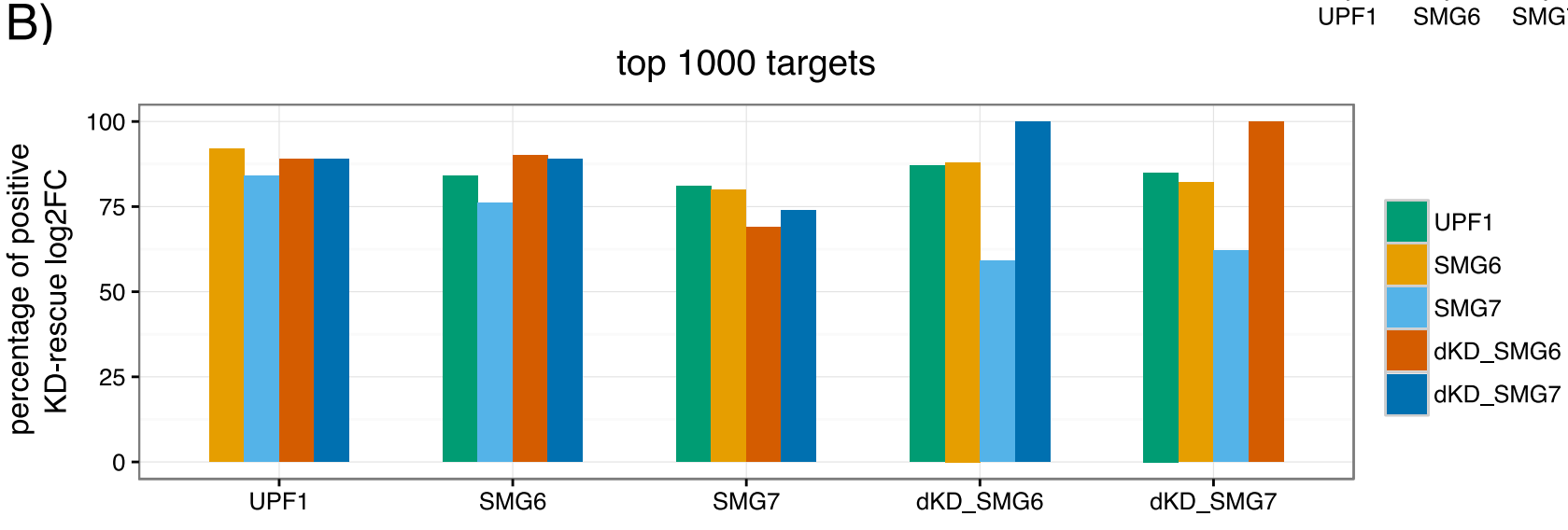
NMD targets are enriched in known and novel characteristic features.

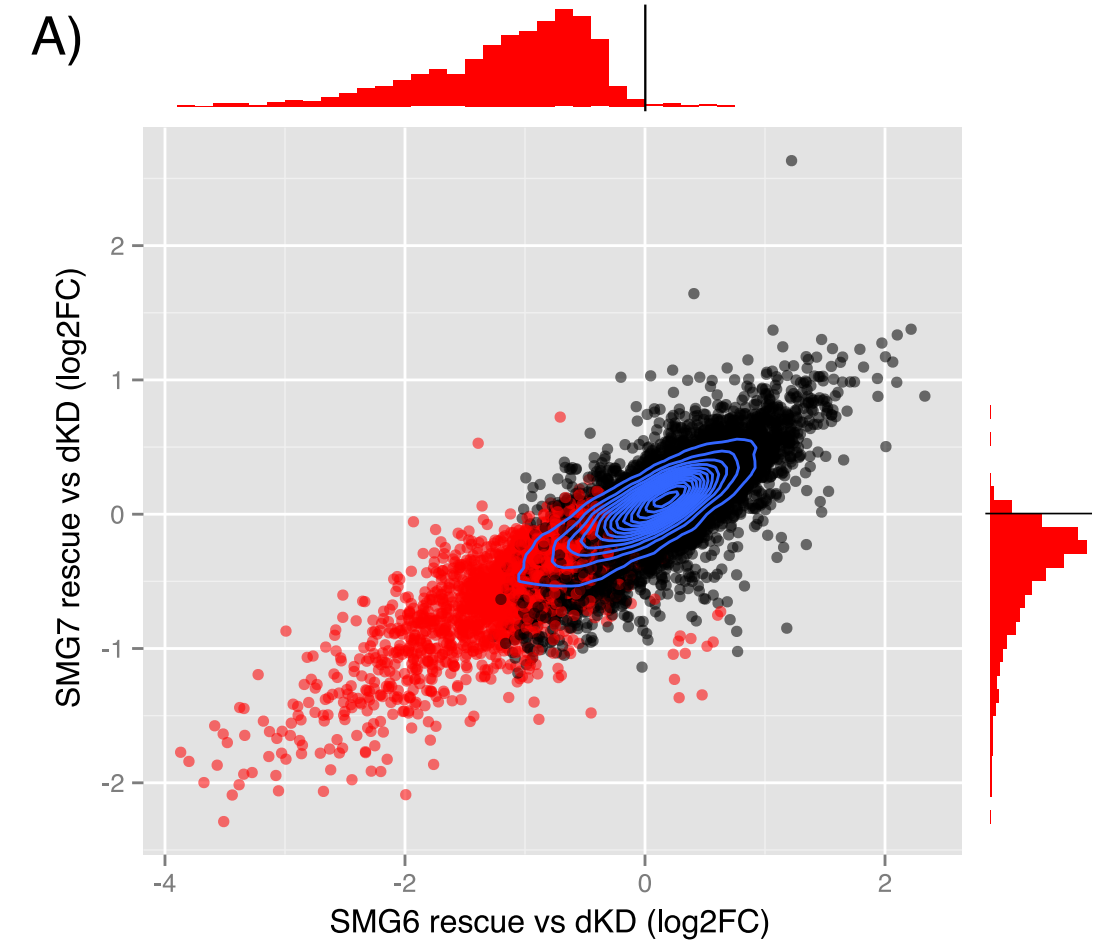
A) Bar plot showing the enrichment of mRNAs with introns in the 3' UTR and uORFs among NMD targets. All mRNAs with introns in the 3' UTR have been removed from the analysis of uORFs. B) Box plots comparing 3' UTR features between NMD targets and a matched control group. All mRNAs with 3' UTR introns or uORFs have been removed from the analysis. This resulted in a total of 660 significant isoforms. The control set is always a set of mRNAs that have the same expression levels of the NMD mRNAs in the Ctrl condition.

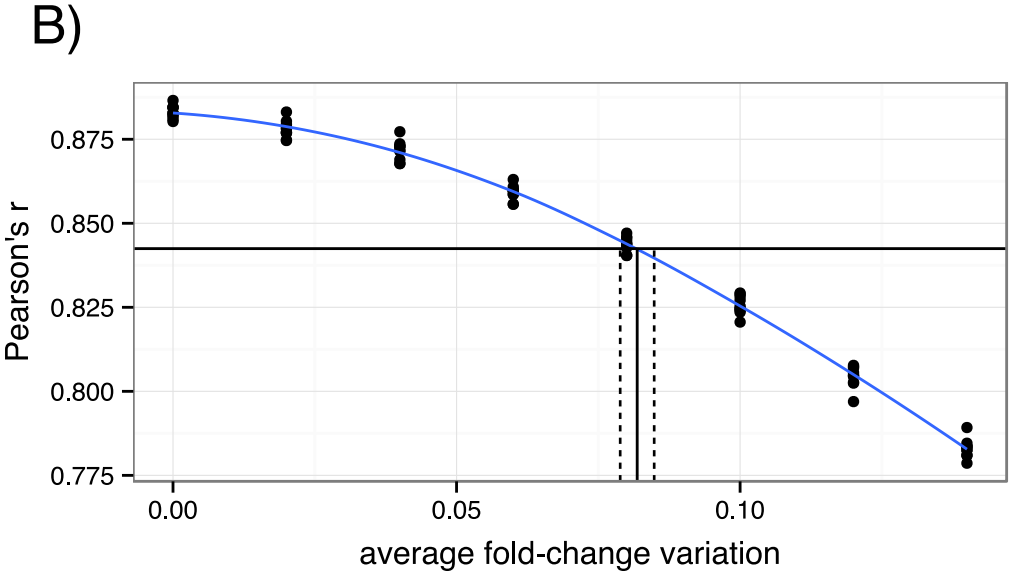

A)
Ctrl
KD
rescue
UPF1
CPSF73
SMG6
CPSF73
SMG7
CPSF73
Ctrl
dKD
SMG6 rescue
SMG7 rescue
SMG6
SMG7
CPSF73
B)
GAS5
RP9P
SMG5
ATF4
RNA levels normalized to ACTB
Ctrl
KD
rescue
UPF1
KD
rescue
SMG6
KD
rescue
SMG7
dKD
dKD SMG6 rescue
dKD SMG7 rescue

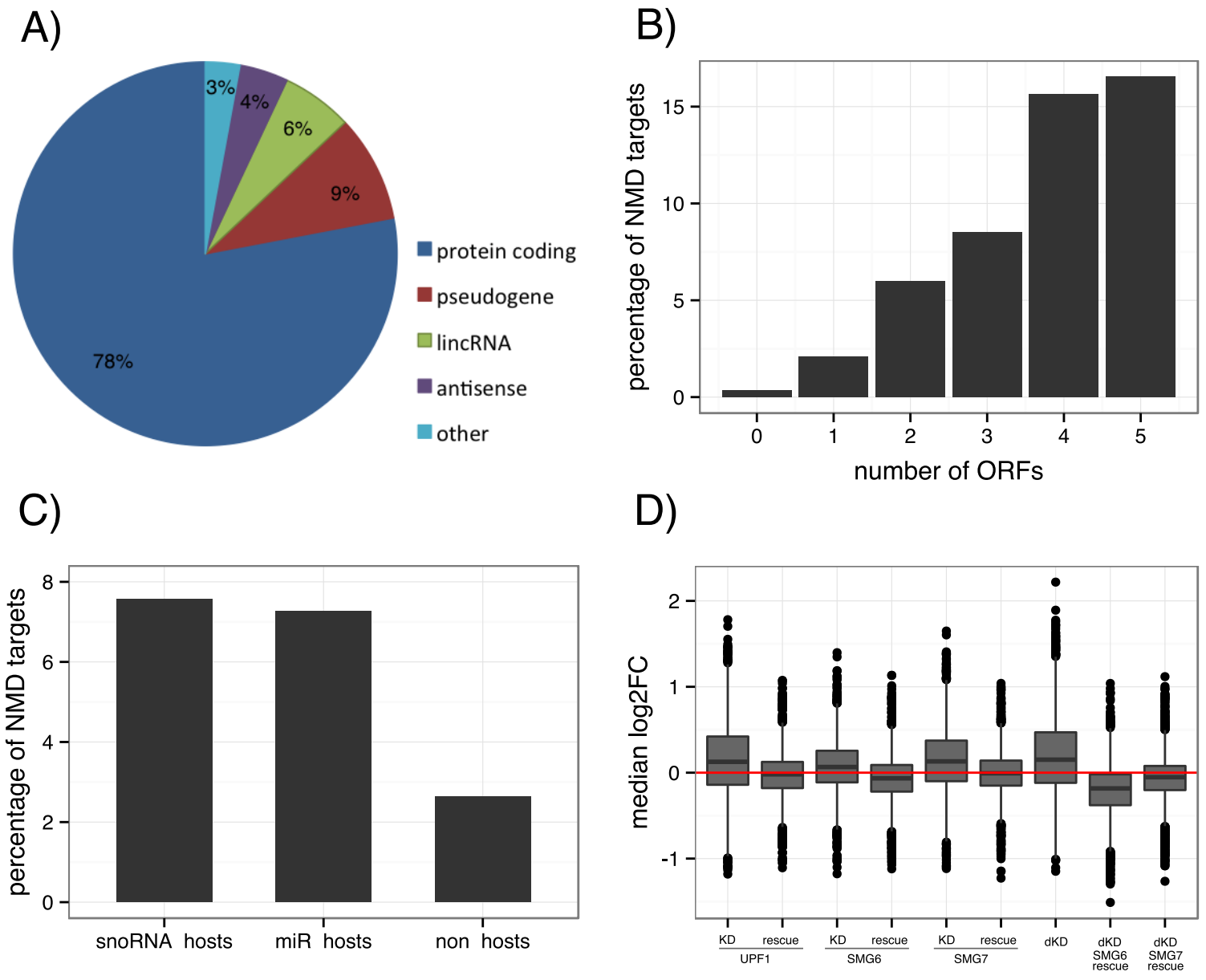

A)
cluster 1: 40K genes
cluster 2: 4K genes
average log2FC
1.5
1.0
0.5
0.0
-0.5
KD
rescue
UPF1
KD
rescue
SMG6
KD
rescue
SMG7
dKD
dKD SMG6 rescue
dKD SMG7 rescue

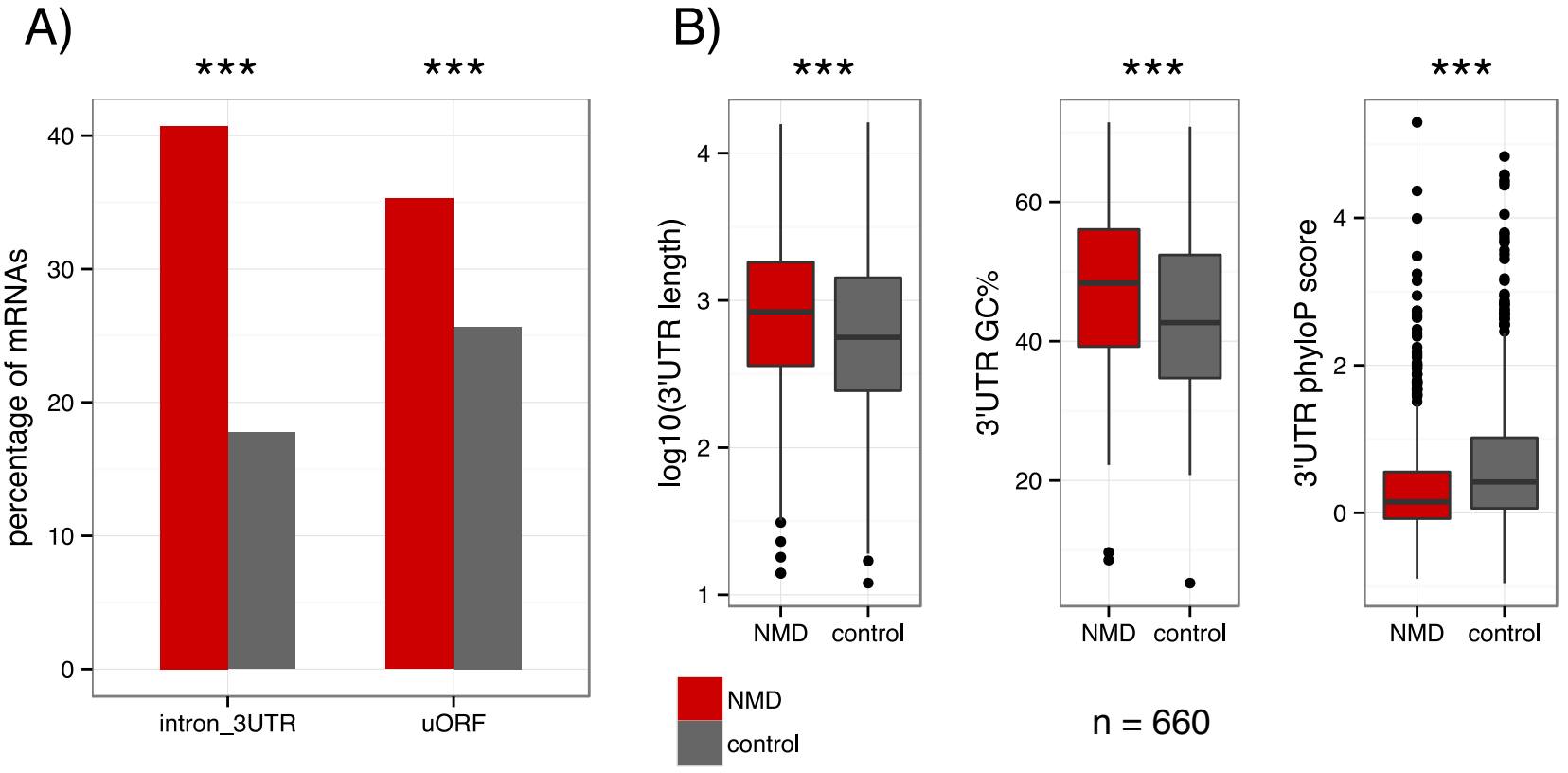

C)
UPF1 KD
SMG6 KD
SMG7 KD
log2FC
1
0
-1
-2
UPF1
SMG6
SMG7


B)
top 1000 targets
percentage of positive KD-rescue log2FC
100
75
50
25
0
UPF1
SMG6
SMG7
dKD_SMG6
dKD_SMG7
UPF1
SMG6
SMG7
dKD_SMG6
dKD_SMG7


A)
SMG7 rescue vs dKD (log2FC)
SMG6 rescue vs dKD (log2FC)
2
1
0
-1
-2
-4
-2
0
2


B)
Pearson's r
average fold-change variation
0.875
0.850
0.825
0.800
0.775
0.00
0.05
0.10


A)
3%
4%
6%
9%
78%
protein coding
pseudogene
lincRNA
antisense
other
B)
percentage of NMD targets
15
10
5
0
0
1
2
3
4
5
number of ORFs
C)
percentage of NMD targets
8
6
4
2
0
snoRNA hosts
miR hosts
non hosts
D)
median log2FC
2
1
0
-1
KD
rescue
UPF1
KD
rescue
SMG6
KD
rescue
SMG7
dKD
dKD SMG6 rescue
dKD SMG7 rescue


A)
***
***
percentage of mRNAs
40
30
20
10
0
intron_3UTR
uORF
B)
***
***
***
log10(3'UTR length)
4
3
2
1
3'UTR GC%
60
40
20
3'UTR phyloP score
4
2
0
NMD
control
NMD
control
NMD
control
NMD
control
n = 660

Downloaded from rnajournal.cshlp.org on November 27, 2016 - Published by Cold Spring Harbor Laboratory Press

RNA
A PUBLICATION OF THE RNA SOCIETY

# Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways

Martino Colombo, Evangelos D. Karousis, Joël Bourquin, et al.

*RNA* published online November 18, 2016

**Supplemental Material** http://rnajournal.cshlp.org/content/suppl/2016/11/18/rna.059055.116.DC1.html

**P<P** Published online November 18, 2016 in advance of the print journal.

**Open Access** Freely available online through the *RNA* Open Access option.

**Creative Commons License** This article, published in *RNA*, 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.**

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

Published by Cold Spring Harbor Laboratory Press for the RNA Society