


**Range of nonsense-mediated mRNA decay efficiencies detected among
homogenous cell cultures**

A Thesis in Molecular Biology

Katherine Matlin
Colorado College

Submitted: March 16, 2019

First Reader: Dr. Sara Hanson 

Second Reader: Dr. Darrell Killian 

Research Mentor: Dr. Sujatha Jagannathan

Honor Code Upheld: *Katherine Matlin*

Abstract

Nonsense-mediated mRNA decay (NMD) is a eukaryotic quality control mechanism for the dynamic regulation of gene expression. NMD degrades transcripts containing a premature termination codon (PTC) more than 50-55 nucleotides upstream of the final exon-exon junction. Although NMD is a ubiquitous mechanism for degrading RNA transcripts in all eukaryotes, there is great variety in the efficiency and specificity of the degradation mechanism. While most transcripts containing a PTC are degraded via NMD, transcripts containing a PTC can evade NMD and produce truncated or full-length proteins¹. NMD efficiency may also vary based on gene sequence, intracellular location, tissue, or on an individual level. This study aimed to aid the understanding of NMD as an endogenous control for gene expression by evaluating NMD efficiency in homogenous cell cultures. We evaluated NMD efficiencies in human embryonic kidney cells by transfecting cell cultures with dual-fluorescing reporters for NMD. We measured fluorescent levels through flow cytometry, and surprisingly detected varying NMD efficiencies among cells of the same culture. To investigate the possible causes of the range in NMD efficiency, we sorted cell cultures based on NMD efficiency levels and evaluated cell populations for their concentrations of NMD factors through immunoblotting and RT-qPCR. Results revealed that NMD factor expression levels did not correlate with NMD efficiency, which proposes new questions for the role of NMD factors in NMD and other possible intracellular mechanisms affecting NMD efficiency. We hypothesized that cell cycle may be affecting NMD. To study the possible relationship, groups of cells with varying NMD efficiencies were evaluated through immunoblotting for cell cycle stage. Preliminary results did not indicate a relationship; however, the association must be further evaluated. Conclusively, we determined a range in NMD efficiency among individual cells in homogenous human embryonic kidney cell cultures. We aim to progress this research by determining key factors and mechanisms that may influence NMD efficiency. Implications for understanding the specificities of NMD activity are far-reaching in the medical field, as several severe human diseases, such as facioscapulohumeral muscular dystrophy, are strongly tied to NMD inhibition.

Introduction

NMD is critical for global gene regulation

Cells of all types must respond to their surroundings by activating and repressing genes. Through the dynamic control of gene expression around transcription and translation, protein levels can fluctuate appropriately in response to cell needs. Gene expression in DNA is primarily modified through the interaction of transcriptional activators and repressors with cis-acting enhancers and silencers. However, while transcriptional control is well-studied, posttranscriptional methods for gene expression such as RNA degradation are not as well understood. In this study, we analyze nonsense-mediated mRNA decay (NMD), a mechanism used to eliminate aberrant transcripts. NMD is a ubiquitous mechanism for RNA degradation in all eukaryotes, which has vast implications for the control of gene expression¹.

Most transcripts that are targeted for NMD are characterized by having a premature termination codon (PTC) upstream from the final exon-exon junction complex (EJC). Several different cellular mechanisms can result in transcripts gaining a PTC. For example, random mutations in transcription can create a PTC, or PTCs can be produced by programmed splicing mechanisms¹. The inclusion of an alternative exon may introduce a PTC in the open reading frame (ORF), or cause a frameshift mutation which can deposit a PTC downstream from the mutation. Long 3' untranslated regions (UTRs) and upstream open reading frames (uORFs) have also been found to subject the transcript to NMD².

Despite NMD's critical role in clearing aberrant transcripts, NMD also serves an important purpose for the regulation of global gene expression according to cellular conditions, such as cellular stress³ or during embryogenesis⁴. In addition, the NMD pathway has also been shown to be self-regulatory: it degrades factors that promote NMD^{5,6}.

NMD is carried out by a highly dynamic mechanism

NMD is a dynamic pathway and relies on the balanced expression of numerous factors. The core of NMD activity is thought to occur in the cytoplasm, and NMD factors are shuttled between the cytoplasm and the nucleus¹. The NMD pathway relies heavily on RNA helicases, which generally serve to remodel RNA and form complexes to affect translation⁷. In the NMD pathway, RNA helicases use ATP to translocate along the transcript and recruit other NMD factors to appropriate regions on the transcript⁸. Although NMD is ubiquitous among eukaryotes, the precise mechanism varies across different species⁸.

In mammalian cells, transcripts are most commonly targeted for NMD by the presence of a PTC located at least 50-55 nucleotides (nt) upstream from the final EJC. During normal RNA splicing, introns are removed and EJCs are temporarily deposited near the splice junctions and remain there after export from the nucleus. As the ribosome progresses with translation, the ribosome displaces the EJCs. Therefore, when a ribosome is forced to stop at a PTC before it can displace the final EJC, the lingering EJC can recruit NMD factors, initiating the degradation of the transcript⁸ (Figure 1).

An essential NMD factor, UPF1, is a RNA helicase responsible for various RNA-degradation mechanisms, DNA repair and replication, and S phase progression⁹. UPF1 binds to all transcripts nonspecifically before translation and is normally displaced by ribosomes moving along the transcript⁹.

NMD is initiated when a PTC is recognized by the translating ribosome. Specifically eukaryotic release factors, which interact with the ribosome, eRF1 and eRF3 recognize the PTC⁸. eRF3 acts as a GTPase, and hydrolyzes GTP to lead to the dissociation of the nascent polypeptide and the complex¹. During NMD, eRF3 recruits UPF1 to the terminating ribosome to initiate NMD. As a result, UPF1 is phosphorylated, and activated, which causes UPF1 to act as a ribonucleoproteinase (RNPase) to facilitate the NMD mechanism⁸.

UPF1 is phosphorylated by the SMG1c complex; this phosphorylation is considered the rate-limiting step for NMD. The SMG1c complex is composed of protein kinase SMG1 and subunits SMG8 and SMG9¹⁰. The SMG1c complex binds to other NMD factors to form the surveillance (SURF) complex, and is formed alongside the PTC.

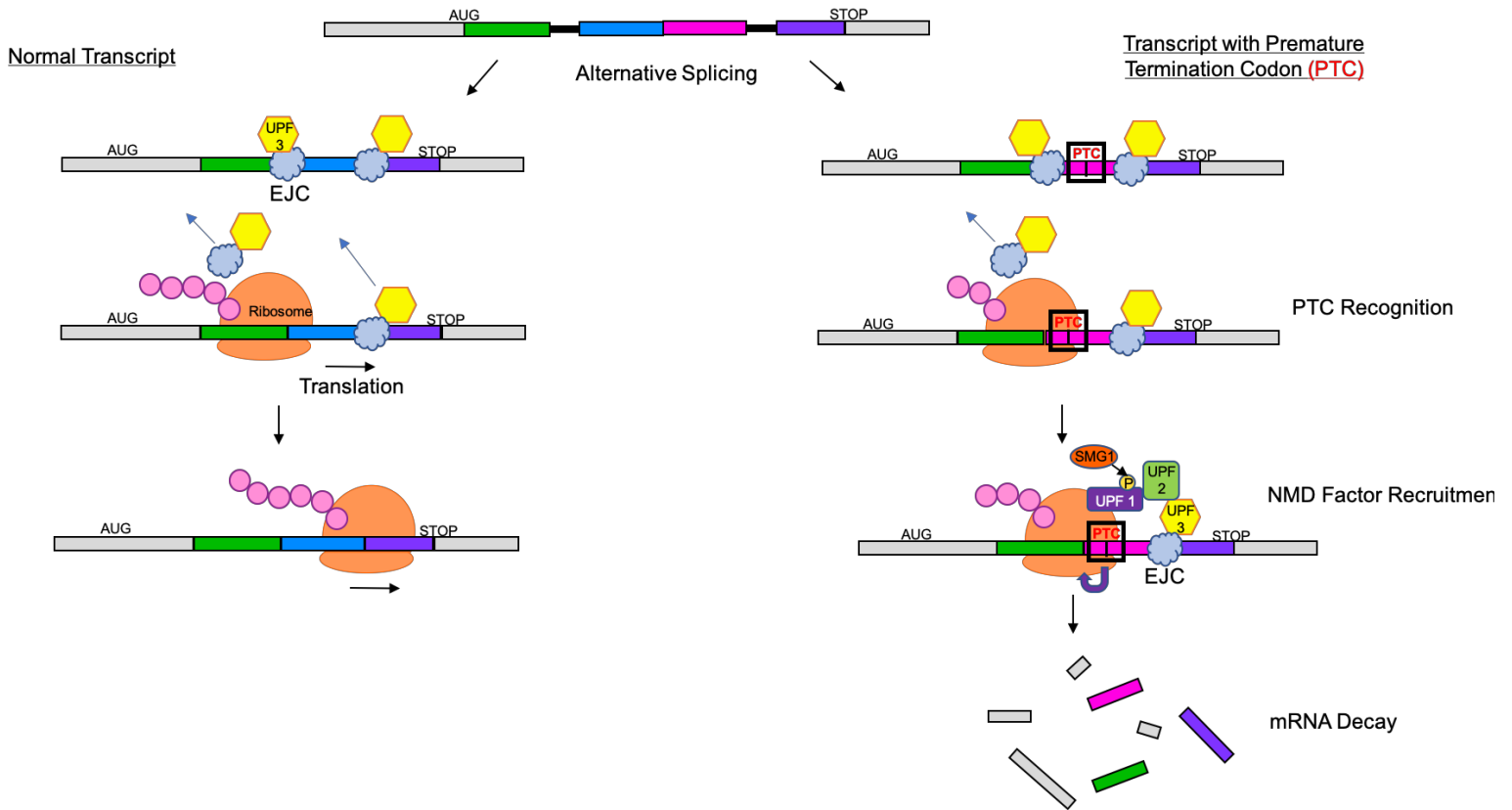


Figure 1. NMD is initiated by the recognition of a PTC at least 50-55 nt upstream from the final EJC. Ribosomes represented in orange; NMD factor UPF3 represented in yellow; EJC represented in pale blue; growing polypeptide represented in pale pink.

SMG1 associates with UPF1, along with eukaryotic release factors eRF1 and eRF3, and RNA helicase DEAH box polypeptide 34 (DHX34)⁸. SMG8 and SMG9 associate with SMG1 and regulate its activity through conformational changes. The SURF complex interacts with UPF2, UPF3b, and the EJC, which are located downstream of the PTC, to form a secondary complex, the decay-inducing complex (DECID)⁸. This ultimately triggers the phosphorylation of UPF1 and the release of Eukaryotic release factors eRF1 and eRF3¹.

In its active helicase conformation, phosphorylated UPF1 translocates along the transcript, causing the dissociation of other proteins interacting with the transcript and in turn, exposing the transcript to nucleases for degradation⁸. In addition, phosphorylated UPF1 associates with phospho-binding proteins SMG5, SMG6, and SMG7, and with other general mRNA degradation factors. SMG6 is an endonuclease that interacts with UPF1 to cleave mRNAs upstream of the PTCs. Meanwhile, SMG5 and SMG7 bind as a heterodimer to phosphorylated UPF1 and the SMG7 subunit recruits Pro-rich nuclear receptor co-activator 2 (PNRC2) to remove the 5' cap of the transcript¹. SMG7 additionally binds to POP2, the catalytic subunit of CCR4-NOT deadenylase to expose the 3' end to degradation¹¹. Consequently, XRN1, a 5'-3' exonuclease, and the exosome, a 3'-5' exoribonuclease, together catalyze the directional degradation of the unprotected pieces of the RNA⁸ (Figure 2).

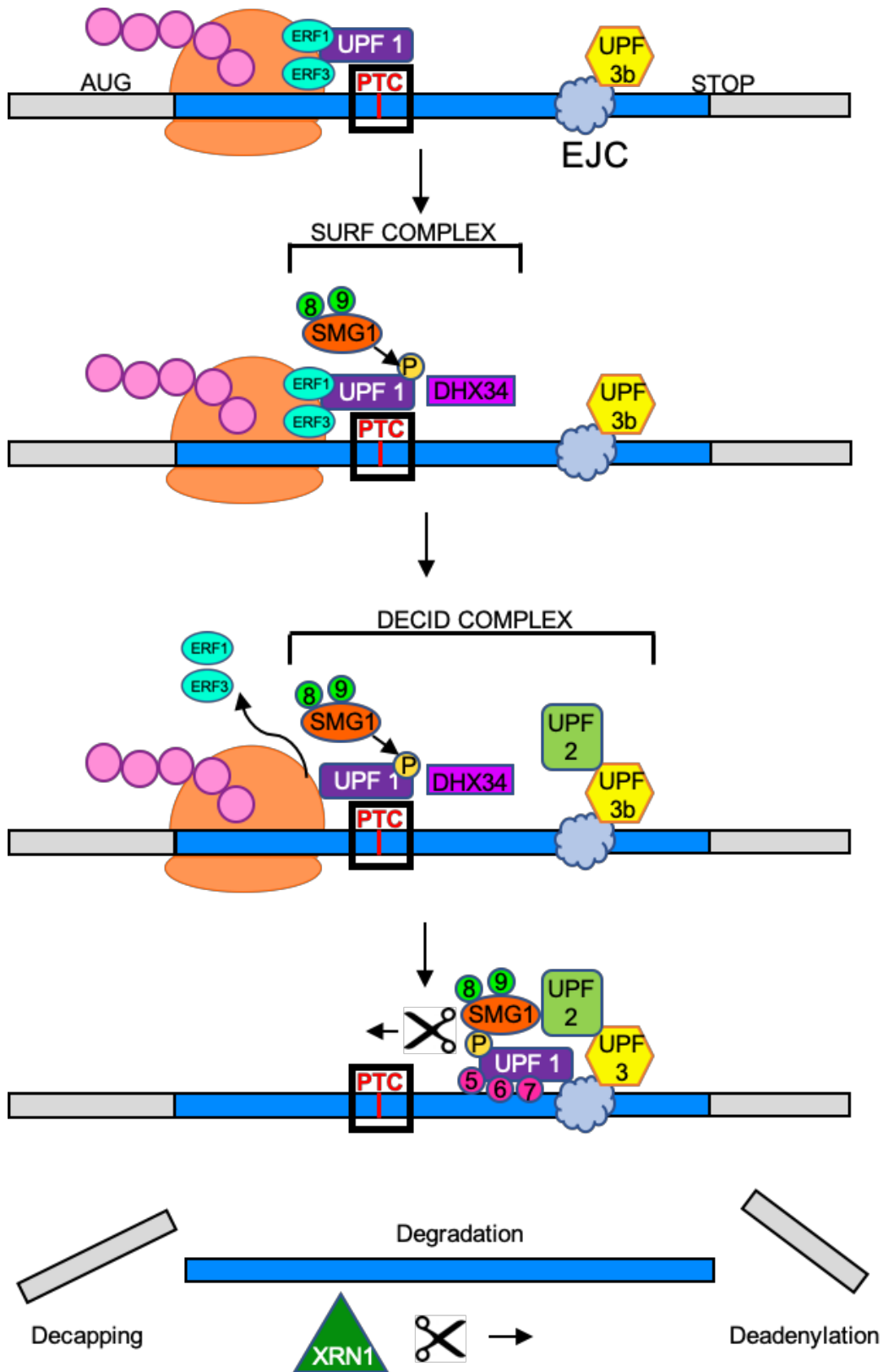


Figure 2: Degradation is collaboratively induced by a series of NMD factors. The general mammalian NMD mechanism.

NMD is disrupted in genetic disorders

The specificity of NMD for target transcripts is not well understood. Interestingly, not all transcripts targeted for NMD contain a PTC; some transcripts are targeted due to structural characteristics, such as having a long 3' UTR². Despite the uncertainty, NMD is known to play an essential role in cell-wide gene expression. For example, about 30% of genes in mice have been found to generate at least one transcript that could be targeted for NMD¹², exemplifying the broad impact NMD has on global gene expression. The importance of the NMD pathway is exemplified by the multitude of diseases which result from NMD-induced loss of function of essential genes. Approximately one third of all human genetic diseases are caused by nonsense and frameshift mutations that generate PTCs; therefore, the research of the cellular responses to PTCs has significant applications to human health¹³.

NMD is inhibited in facioscapulohumeral muscular dystrophy

Because NMD has such varied yet critical roles in gene expression across a variety of cell states, a mutated or inhibited NMD pathway may subject cells and individuals to a variety of genetic disorders¹⁴. Even a partially inactive NMD mechanism has been correlated to disease phenotypes¹⁵. Thus, the clinical application of the NMD pathway is a major focus of NMD research. A major genetic disease linked to mutations in the NMD pathway, which has been studied widely in this context is facioscapulohumeral muscular dystrophy (FSHD). FSHD is an autosomal dominant disease that ultimately causes muscle weakness in the face, shoulders, and upper arms¹⁶.

FSHD is caused by the misexpression of the *D4Z4* macrosatellite array, which causes the expression of the *DUX4* retrogene in skeletal muscles (rather than normal expression in the testis and during early development). *DUX4* inhibits NMD by encoding for a double homeobox transcription factor that results in the proteolytic degradation of UPF1, an essential NMD factor. However, the mechanism involved is incompletely understood. Interestingly, *DUX4* is normally degraded by NMD itself, and this relationship caused by *DUX4* expression causes a double-negative feedback loop that results in an amplification of *DUX4* expression in affected muscle¹⁶. Further research into NMD and

its variability aims to discover the relationship between NMD inhibition and FSHD and provide insight into possible therapies.

NMD is responsive to the cellular environment

NMD has not only been evaluated in its active or inactivated form, but has been shown to vary in efficiency on several different levels. Unfortunately, the variability in NMD efficiency is poorly understood. In investigating NMD variability, it is helpful to understand what peripheral mechanisms may be affecting NMD. Additionally, many NMD factors play roles in generalized responses to cell conditions, including DNA damage repair, stress, related protein degradation, and embryogenesis¹. Many stress-related transcripts contain upstream ORFs or long 3' UTRs, making them prime targets for NMD². Additionally, under stress conditions, the NMD factor *SMG1* is upregulated in response to dsDNA breaks¹⁷. NMD commonly exists in a negative feedback loop with stress response pathways. Specifically, NMD acts in a negative feedback loop with the unfolded protein response (UPR) in response to stress in the endoplasmic reticulum (ER). In unstressed conditions, the NMD pathway degrades components of the UPR. In turn, when the UPR is activated due to stress, the pathway inhibits NMD to allow for an appropriate stress response¹⁸.

In addition to the well-described transcriptional regulation of embryogenesis, NMD has a critical role in regulating gene expression during embryogenesis and development. Certain genes must be silenced until adulthood, while other genes must only be actively expressed during development and the unique balance of gene expression is modulated by NMD. As the embryo develops, NMD is required for stem cell differentiation and neuronal development. For example, NMD has been shown to interact with the TGF- β signaling pathway, which promotes neuronal differentiation. UPF1 aids the destabilization of the substrate for the TGF- β inhibitor, to promote differentiation¹⁹.

Evaluating the variability in the NMD pathway

Although it has been shown that NMD factors play a vital role in DNA and RNA regulation, there is little understanding of how NMD factors fully regulate gene expression. To gain a better understanding of regulation via NMD factors, it is important to analyze

the ways in which NMD efficiency varies among individual cells, different cell types, conditions, and individuals. There is significant evidence of variability in NMD, on a cell-to-cell level and in terms inter-individual variation²⁰. Because many factors that are essential to NMD have tangential roles in other molecular processes, it is possible that this variability arises, in part, from the interaction between the NMD pathway, or NMD factors, with other genome regulation pathways. For example, if an NMD factor is also important for the stress response, then under stress conditions it may not be able to participate in NMD because it is sequestered in stress response protein complexes. However, this relationship has yet to be investigated.

The NMD pathway is complicated by the variability in its efficiency on several levels. While NMD is active in all eukaryotes, the mechanism has been shown to vary in efficiency among tissues and individuals. Furthermore, varying NMD efficiencies have been associated with different disease phenotypes and severities¹³; however, variation in NMD efficiency has not been thoroughly studied, especially on a cellular level. In this study, we analyzed NMD within a homogenous cell culture to determine whether a range in NMD efficiency exists on a cell-to-cell level and investigate factors that may influence NMD efficiency. We expected to see no variety in NMD efficiency, and only cells experiencing high levels of NMD, or no levels of NMD.

We analyzed NMD efficiency in a homogenous human embryonic kidney cell (HEK293) culture, because of their receptiveness to transfection and fast rate of division, to determine if NMD variability existed among cells of the same cell line under identical conditions. To study NMD rates in cell culture, we utilized a dual-fluorescing reporter plasmid construct that reports on the NMD status of a cell, based on a previously described reporter system²¹. We transfected constructs into cells and examined NMD activity using flow cytometry.

We surprisingly found that NMD efficiency was indeed varied, and we further aimed to determine what factors could be contributing to cell-to-cell variability by cross-examining their roles in other molecular pathways. Specifically, we examined the DNA synthesis mechanism during the cell cycle through immunoblotting. We additionally used immunoblotting and RT-qPCR to analyze how NMD factors may correlate to NMD efficiency level.

We found considerable NMD variability in homogenous cell cultures, suggesting other inherent intracellular processes could be effecting NMD rates. However, we did not find a clear association of NMD efficiency to cell cycle stage. We expected NMD would be related to cell cycle due to the overlap in proteins required for both NMD and cell cycle regulation. However, this relationship must be further analyzed. Interestingly, results did not show NMD factor concentrations to be correlated to NMD activity levels, which opens new questions for research into the NMD pathway. Conclusively, we developed an experimental model for understanding NMD efficiency as it relates to NMD factor concentrations and other cellular mechanisms.

Further investigation into the relationship of intracellular molecular pathways and NMD is required for the understanding of NMD regulation. The understanding of NMD variability has the potential to have profound impact on the understanding of the pathology of associated diseases, including FSHD, and the development of groundbreaking therapies.

Results

NMD+ plasmid construct induced NMD in HEK293 cell culture. To test whether NMD could be induced and reported through expression of a fluorescent protein, two plasmid constructs, each including two fluorescent reporter genes with separate promoters were transfected into identical HEK293 cell cultures. GFP acted as an internal control for transfection efficiency, while katushka (KAT) was targeted for NMD due to an additional intron in the 3' UTR in the NMD+ plasmid construct. This intron placement causes the canonical stop codon to be recognized as a PTC (Fig. 3). There is not an additional intron in the wild type (WT) plasmid construct; thus, NMD is not induced in cells carrying that plasmid. Upon visualization, cells transfected with the wild type WT reporter expressed GFP and KAT in cell cultures. Cells transfected with the NMD+ construct expressed GFP and visibly decreased KAT levels (Fig. 4).

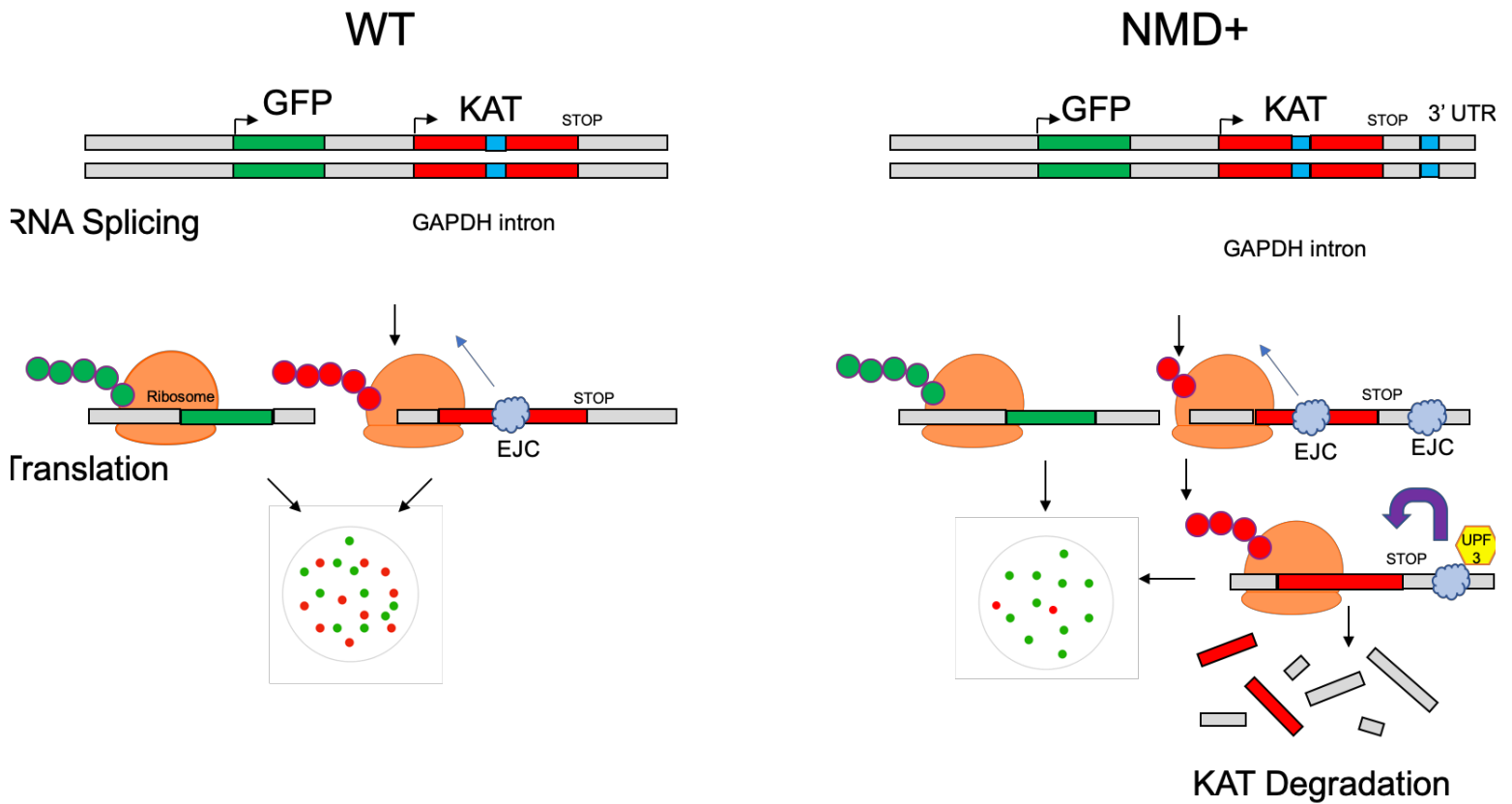


Figure 3. GFP and KAT are used as markers to visualize and quantify NMD efficiency. *GFP* is represented by green regions. *KAT* is represented by red regions. Introns are in blue. Exon junction complexes are represented by light blue shapes. Each circle represents an individual cell and the dots represent GFP or KAT protein.

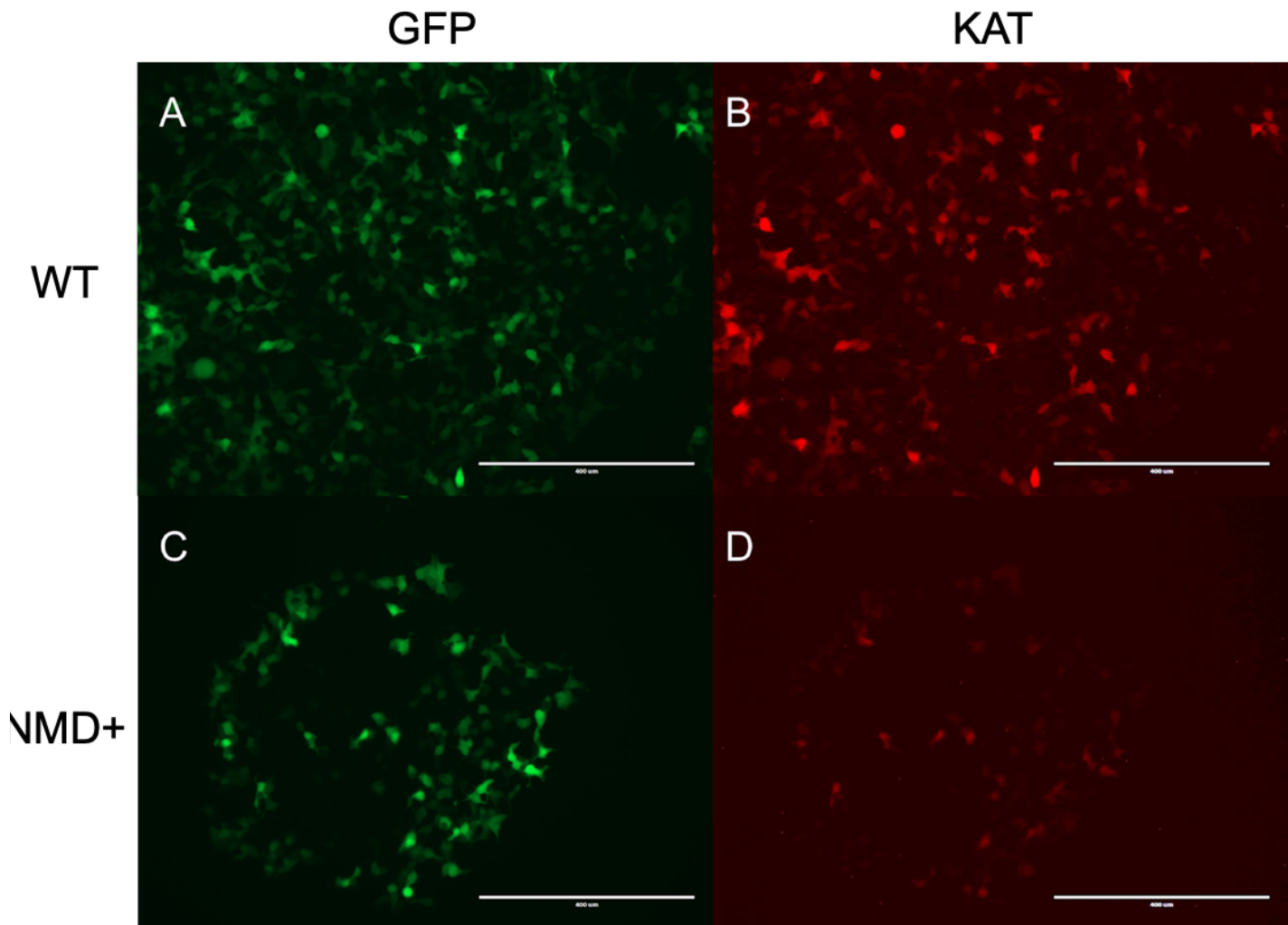


Figure 4. Higher NMD rates were visualized in NMD+ transfected cells than WT transfected cells. WT transfected cells are imaged in A and B for GFP and KAT fluorescence. NMD+ transfected cells are imaged in C and D.

NMD+ transfected cells showed varying levels of NMD efficiency. To determine whether NMD rates were varied among cell cultures, we evaluated GFP and KAT fluorescent levels. We expected to see consistently high levels of GFP across all cells that were transfected, and either very high KAT levels if NMD was inactive or low KAT levels if NMD was active. A range in KAT expression levels among cells with high GFP levels would indicate varying NMD efficiencies, as KAT would be degraded at varying rates among cells. Fluorescence levels were measured through flow cytometry. Results revealed that WT cells had an approximately linear increase of KAT signal corresponding to GFP expression, indicating the lack in overall NMD activity, as expected from the control (Fig. 5A). However, NMD+ transfected cells showed a range in KAT expression associated with a steady, high level of GFP expression, showing that there were varying NMD rates for KAT among cells in the same cell culture (Fig. 5B). When comparing the entire NMD+ and control cell populations, the NMD+ cell population showed significantly decreased KAT expression (as normalized to GFP expression), compared to the control population (Fig. 6A). Differences in KAT to GFP levels were significant. To investigate the cause of varying NMD efficiencies in cultures, cell cultures evaluated with flow cytometry were sorted into groups based on NMD efficiencies for further studies. We collected one population from the control culture, and three populations of varying NMD efficiencies from the from the NMD+ culture. After KAT levels were normalized for GFP levels to control for transfection efficiency, KAT levels were still varied among the 3 NMD+ sorted efficiency groups (Fig. 6B). The three NMD+ sorted populations included all samples with high GFP expression levels, and “high”, “medium”, and “low” NMD rates, (as NMD rates are inversely related to KAT levels). Differences in KAT to GFP levels among the three sort groups were significant.

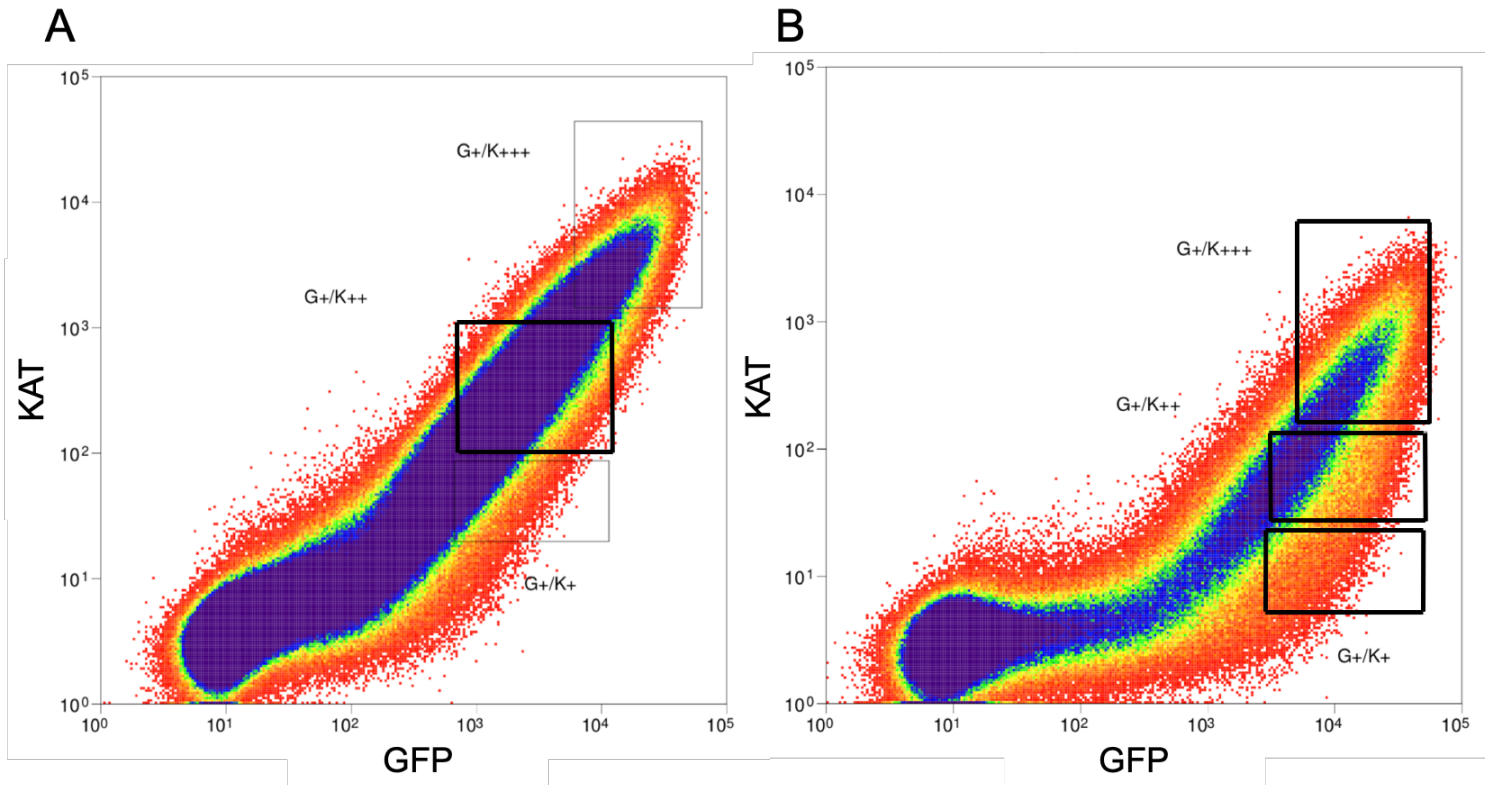
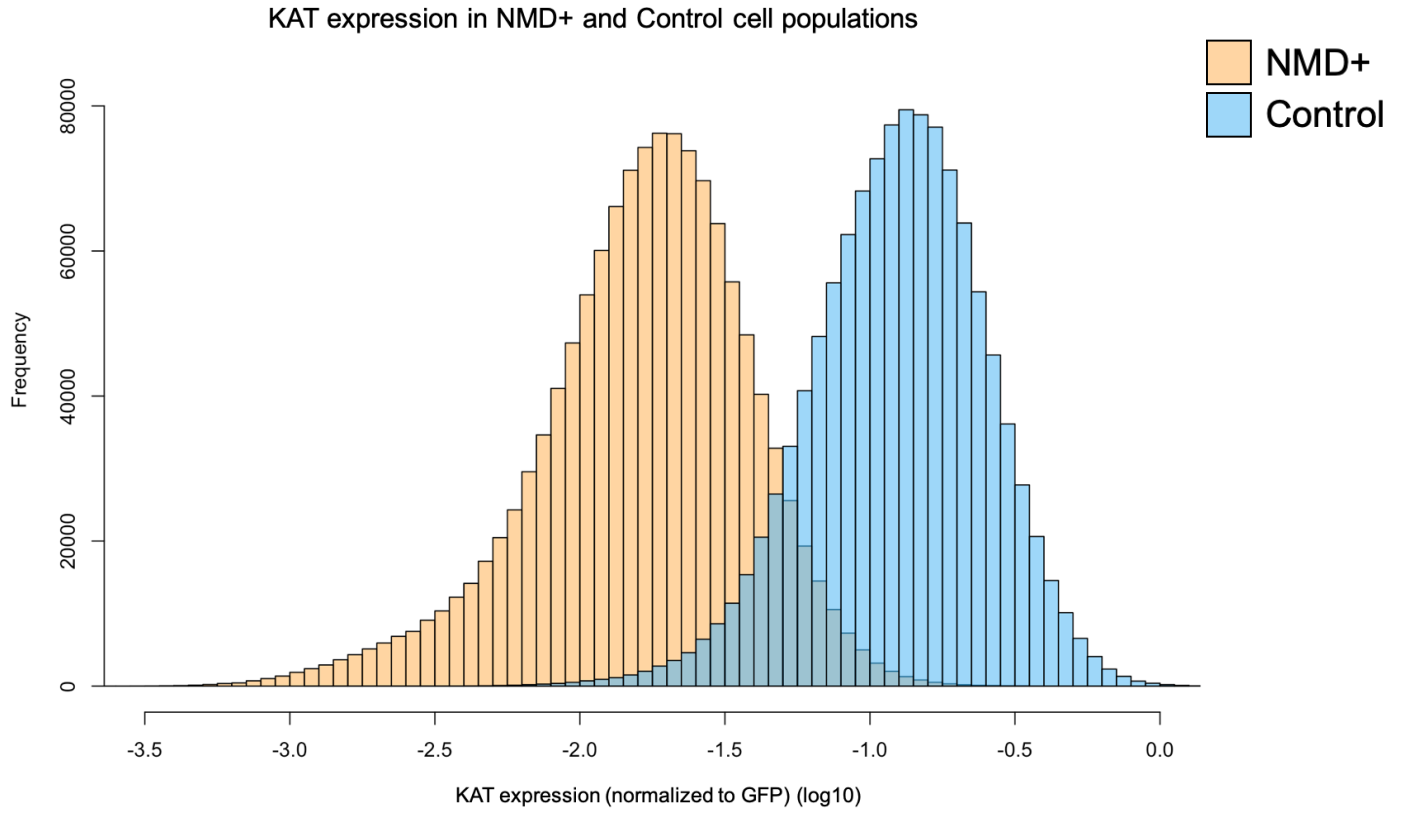


Figure 5. NMD+ cells fluoresced a broader range of KAT compared to WT control. Flow cytometry fluorescent measurement of KAT fluorescence to GFP fluorescence through forward and side scatter measurements. Boxed cells were sorted for analysis. **(A)** Cells transfected with WT reporter. **(B)** Cells transfected with NMD+ reporter.

A



B

KAT expression in gated NMD cell populations

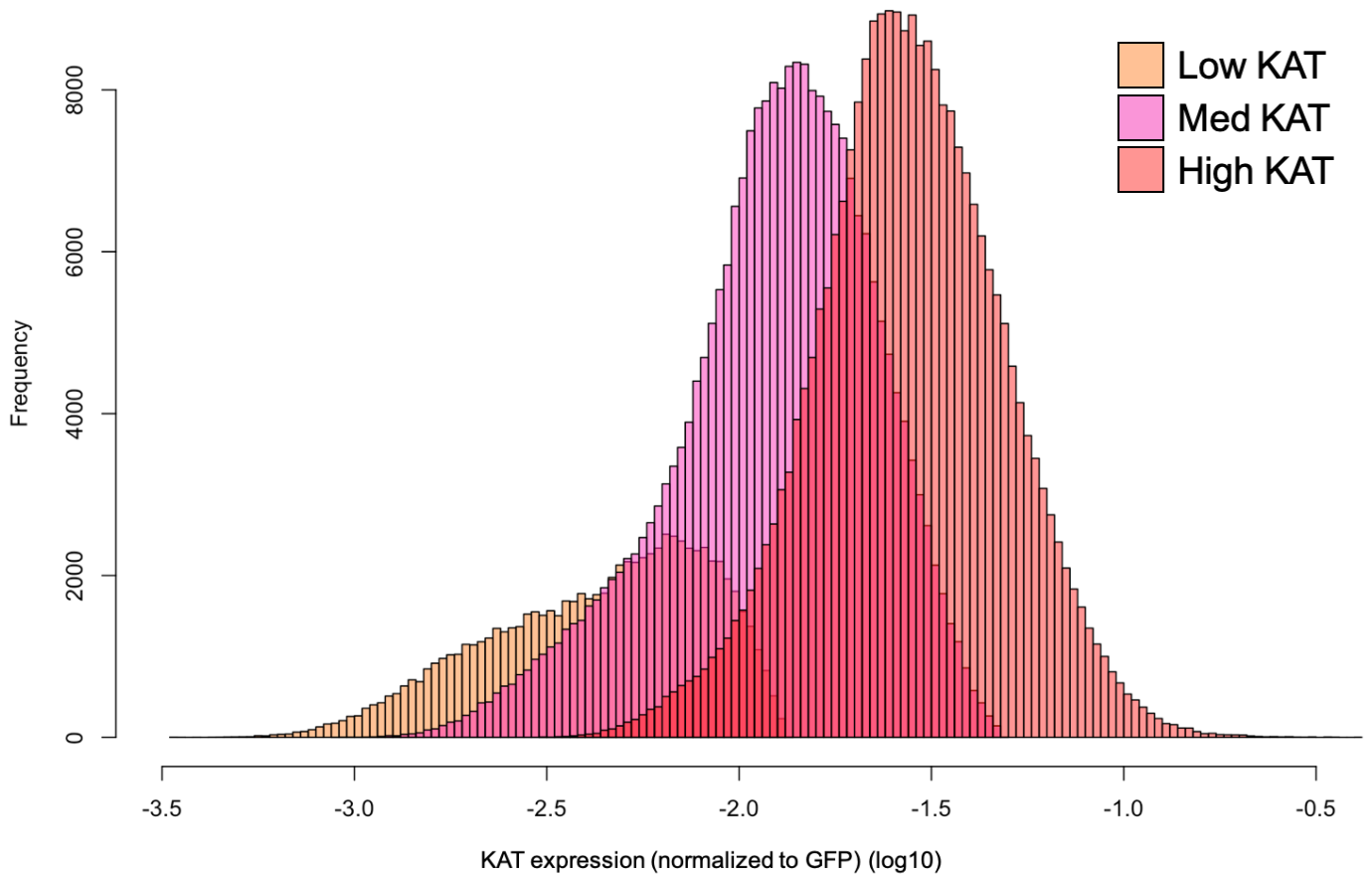


Figure 6. NMD+ sorted groups showed range in NMD activity. Cells fluorescing GFP below $10^{6.75}$ were excluded for analysis and considered noise in both graphs. KAT values for cells were normalized to GFP expression levels. **(A)** NMD+ cell populations are represented in orange. Control cell populations are represented in blue. NMD+ cell population showed lower KAT levels than control. Statistics derived from ks-tests gave a distribution (D) value of 0.83525 and a p-value of $< 2.2e-16$. **(B)** Cells from NMD+ population showed differing rates of NMD efficiency. Populations were gated through single cell sorting with the XDP100. High NMD efficiency population is represented in orange. Medium NMD efficiency population is represented in pink. Low NMD efficiency population is represented in red. D-values among the three curves are as followed: low to medium: 0.61172; medium to high: 0.515; low to high: 0.92071. All p-values were $< 2.2e-16$.

NMD rates showed no correlation to cell cycle stage. In an effort to determine possible intracellular mechanisms affecting NMD efficiency, we set out to evaluate the cell cycle stage of the majority of cells in each sorted group. We hypothesized that cells within each NMD efficiency group would be cells that were currently undergoing the same cell cycle stage. We evaluated cell cycle stage with an immunoblot for cell cycle stage-specific marker. Specifically, G1/S phase was indicated by the presence of phosphorylated Cdk2 (pTyr15) and mitosis was indicated by phosphorylated histone H3 (pSer10). If there was indeed a correlation, we would expect particularly high levels of cell cycle protein combinations to correlate with NMD level. However, none of the samples showed particularly high concentrations of either cell cycle protein, suggesting NMD rate is not correlated to cell cycle (Fig. 7A). We used Histone H3 as a loading control. However, it is important to note that there was a range in general protein concentrations of Histone H3 throughout the samples, indicating unequal loading of protein samples in the gel, which could skew the interpretation of cell cycle marker concentrations and our cell cycle results (Fig. 7A).

GFP and KAT concentrations were detected among all sorted samples. To ensure that GFP and KAT protein expression correlated with fluorescence levels detected, we immunoblotted to determine protein concentration. A FLAG protein tag was embedded in the *GFP* sequence and a HA protein tag was embedded in the *KAT* sequence. FLAG and HA were used as controls to detect protein production of known GFP and KAT expressions. FLAG and HA protein concentrations were confirmed through immunoblotting to ensure protein production of *GFP* and *KAT* regions, supporting that protein production corresponds to fluorescent levels (Fig. 7B).

UPF1 protein concentrations appear inversely related to NMD efficiency rates. Because UPF1 is central to NMD, we aimed to correlate UPF1 protein levels with NMD efficiency levels through immunoblotting. UPF1 protein levels were measured in Control and NMD+ samples. At first pass, UPF1 concentrations seemed to be inversely tied to NMD efficiency level (Fig. 7C). However, our results may be skewed due to a protein-

loading issue, as protein concentrations showed the same concentration patterns in every immunoblot, because the same protein samples were used for each immunoblot.

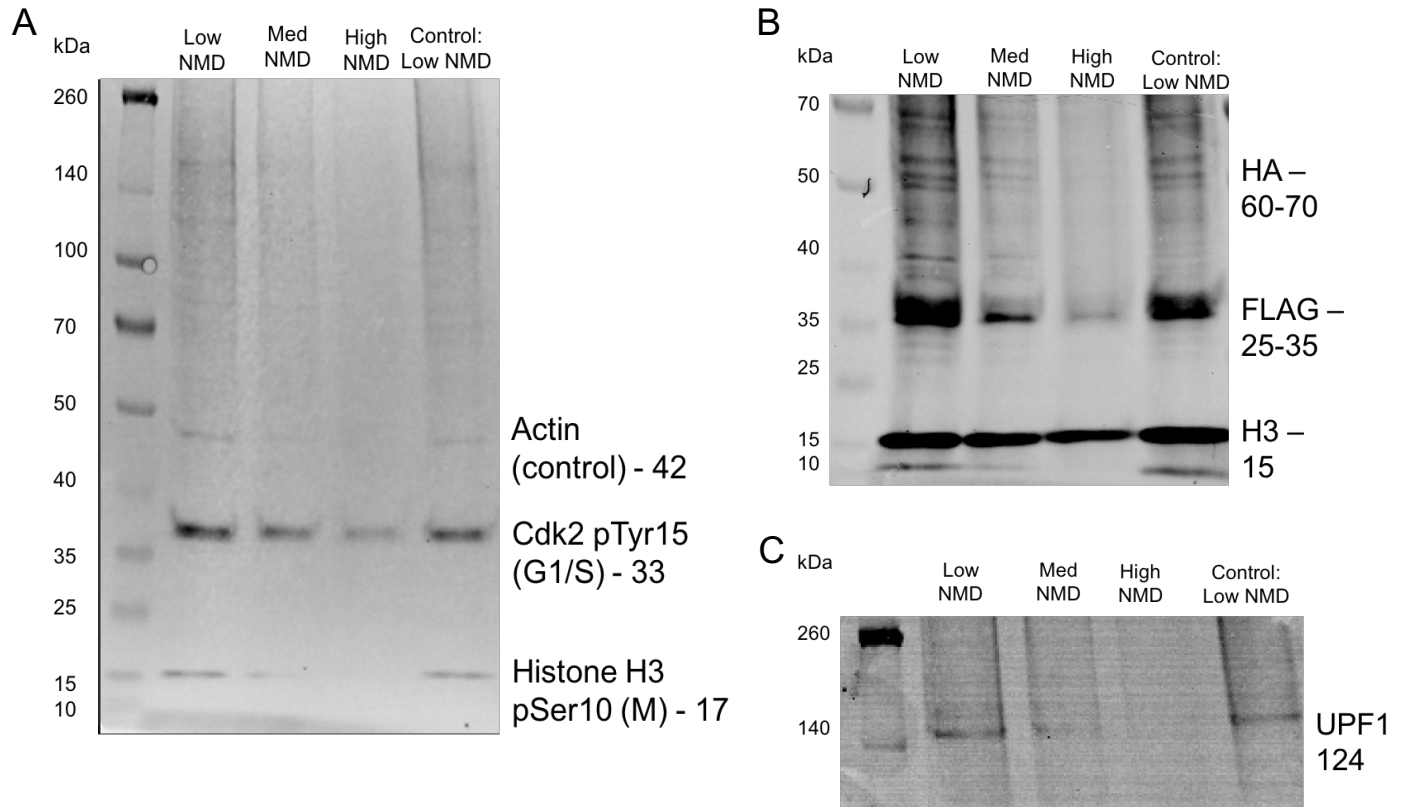


Figure 7. NMD factors and cell cycle proteins did not correlate to NMD efficiency levels. Gated NMD+ samples and control sample were analyzed. **(A)** Antibody cocktail to cell cycle stage markers (Abcam). Actin serves as a control. Cdk2 pTyr15 marked G₁/S arrested cells. Histone H3 pSer10 marked mitotic arrested cells. **(B)** HA tag was embedded in KAT region. FLAG tag was embedded in GFP region. Histone H3 was used as a control. **(C)** UPF1 NMD factor.

UPF and SMG factor mRNA levels appear to be inversely correlated with NMD efficiency. To determine how mRNA levels for NMD factors varied among sorted groups of varied NMD efficiencies, we used RT-qPCR to measure a series of NMD factor RNA levels. We extracted mRNA from sorted cell populations in the control sample as well as low “L”, medium “M”, and high “H” NMD efficiency populations and created cDNA to evaluate gene expression levels. To analyze our data, we normalized gene expression to both *RPL27* gene expression, which is constant among all cells, and to the control sample expression level. We probed samples for the expression of 18 different NMD factors (Fig. 8D). In addition to NMD factor expression, we measured gene expression levels of *GFP* and *KAT*, to confirm mRNA expression levels matched fluorescence levels measured previously to sort NMD+ groups (Fig. 8A). Among all three NMD+ samples, *UPF1* RNA expression remained constant while *UPF2*, *UPF3b*, *SMG1*, *SMG5*, and *SMG6* RNA levels appeared to be inversely tied to NMD efficiency (Fig. 8B and Fig. 8C). In addition, *GNL2*, *NBAS*, *SEC13*, and *UPF3B* also appeared to be inversely correlated to NMD efficiency level (Fig. 8D).

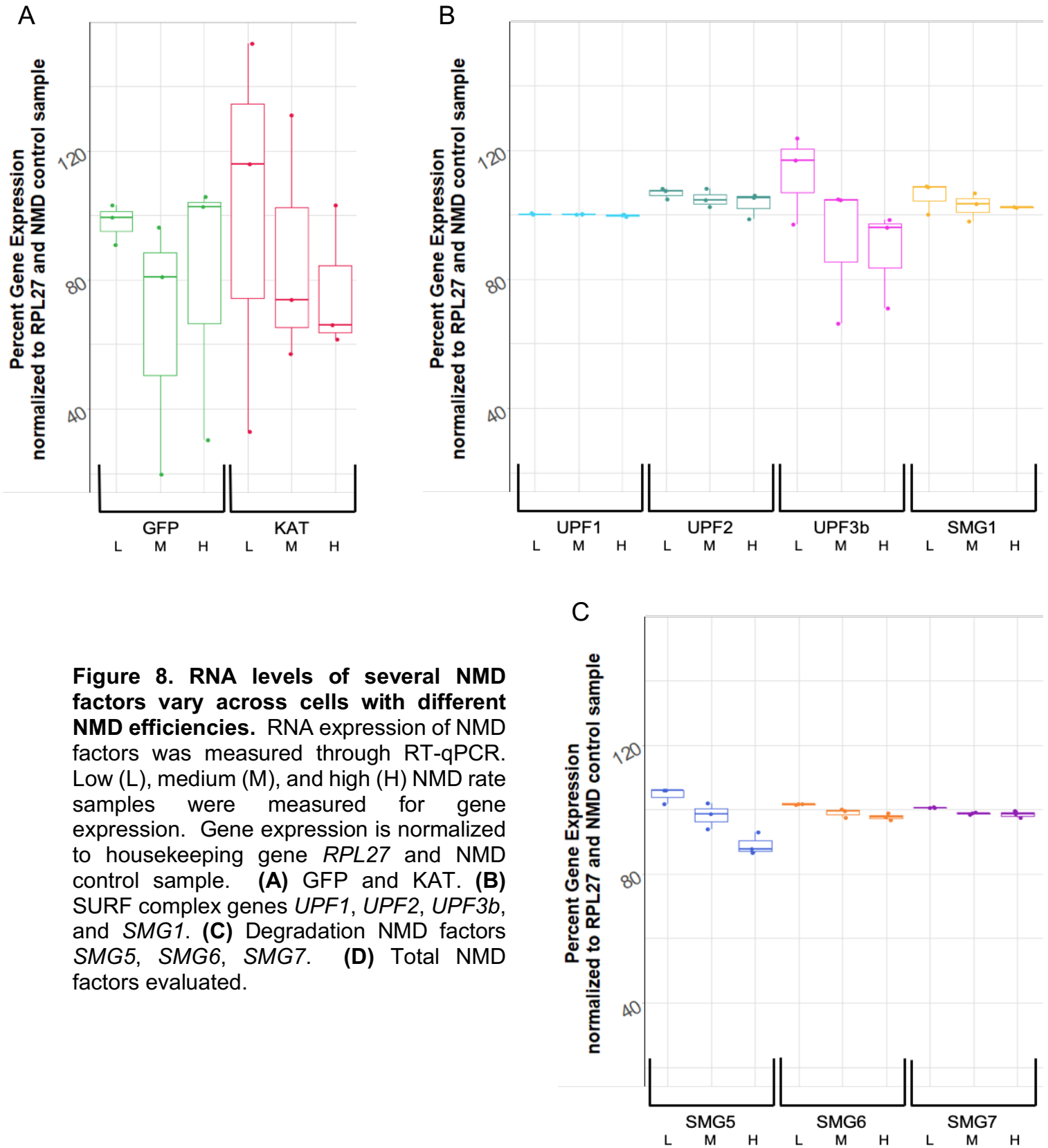
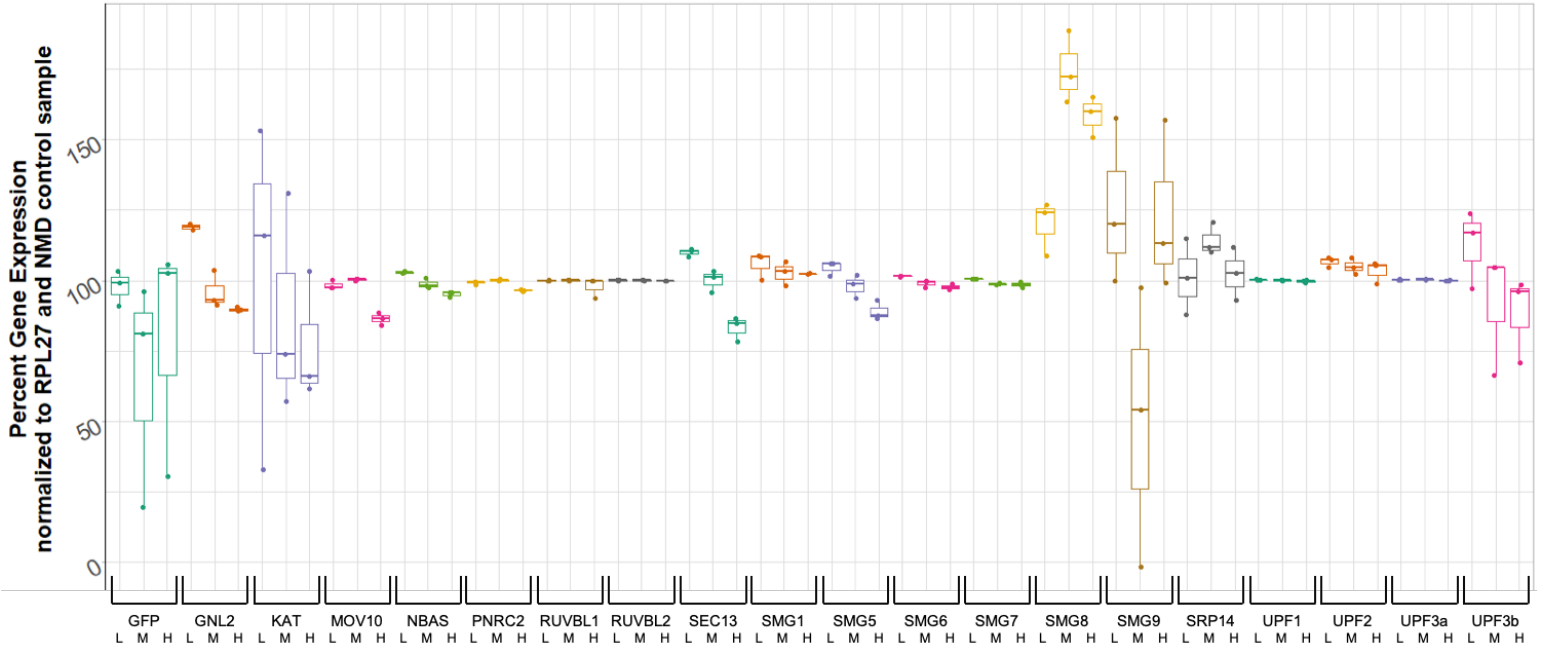


Figure 8. RNA levels of several NMD factors vary across cells with different NMD efficiencies. RNA expression of NMD factors was measured through RT-qPCR. Low (L), medium (M), and high (H) NMD rate samples were measured for gene expression. Gene expression is normalized to housekeeping gene *RPL27* and NMD control sample. **(A)** GFP and KAT. **(B)** SURF complex genes *UPF1*, *UPF2*, *UPF3b*, and *SMG1*. **(C)** Degradation NMD factors *SMG5*, *SMG6*, *SMG7*. **(D)** Total NMD factors evaluated.

D



Discussion

NMD efficiency varies in homogeneous human embryonic kidney cell culture

Variability in NMD efficiency has been recorded on several levels in mammalian cells and among individuals, but few studies have determined the causes of this variability. Thus, in order to investigate the cellular causes of NMD variability, we evaluated NMD efficiency on a cell-to-cell level among homogenous human embryonic kidney cell cultures of the HEK293 cell line. We found considerable variability in NMD efficiency among homogeneous human embryonic kidney cells. Our results are one of the few examples of variability in NMD efficiency on the cell-to-cell level, and serve as the first recorded evidence of variability in the HEK293 cell line.

The novel variation discovered among homogenous human embryonic kidney cells adds another layer of complexity to our understanding of how NMD efficiency is regulated. Although the causes were unclear, variability in NMD efficiency has previously been recorded on several different levels. For example, NMD rates have been shown to vary among different cell types¹³. On multiple occasions, the same PTC-containing transcript has been shown to be degraded at different rates among distinct cell lines. Two different PTC-containing transcripts were tested in both HeLa cervical cancer and MCF7 breast cancer cell lines. Both transcripts were degraded by NMD much more quickly in HeLa cells, showing differences in NMD activity between cell types¹³. Furthermore, a different PTC-containing transcript was tested for NMD efficiency in chondrocytes, lymphoblasts, and bone cells from the same Schmid metaphysical chondrodysplasia patient, and again, the transcript was degraded at different rates in each cell line¹³.

Variation in NMD efficiency was also noted in identical cell types among individuals carrying the same PTC-containing transcript. Specifically, these results were found in nasal epithelial cell lines in cystic fibrosis patients¹³. This phenomenon was further supported by a study analyzing fetuses with Roberts Syndrome, which were spontaneously aborted in the womb. One fetus lived longer, and was also shown to have much higher levels of NMD and more robust clearance of the mutated transcript that was propagating the disease phenotype. Meanwhile, the fetus that died first had a different splicing mutation—one that inhibited NMD. When comparing more patients with different

mutations that cause Roberts Syndrome, mutations that inhibited NMD activity were overall correlated with shorter lifespans²². Thus, NMD efficiency rates have been shown to affect disease phenotypes in several cases, emphasizing the importance of understanding NMD efficiency.

NMD factor concentrations are not positively correlated to NMD efficiency level

Because our results revealed that NMD efficiency varies in homogenous cell cultures, we hypothesized that there could be intracellular mechanisms affecting NMD. To investigate this in more detail, we analyzed how the expression levels of NMD factors corresponded with overall NMD efficiencies. However, we found that in most cases, NMD efficiency did not positively correspond with NMD factor expression level. In fact, many of the NMD factors showed a negative correlation with NMD efficiency. These results suggest that while it is known that a threshold quantity of certain NMD factors are required for NMD functionality, the relationship between NMD factor mRNA expression and NMD efficiency may only go as far as meeting a certain concentration of NMD factors within the cell.

The idea that a threshold expression level of NMD factors is required for NMD activity is supported through the analysis NMD activity in FSHD cell lines. In the case of FSHD, a mutation causing *DUX4* expression causes insufficient levels of NMD factors including UPF1, and results in an increase in NMD substrates¹⁶. Interestingly, however, following *DUX4* expression, and related NMD inhibition, essential NMD factors (*UPF* and *SMG* factors) did not decrease mRNA levels. In fact, *UPF* and *SMG* factors significantly increased in mRNA expression levels in cells, perhaps as a response to the lack of NMD activity¹⁶. A similar example was shown in neurons, when NMD factor UPF3A was shown to compensate for depleted UPF3B levels²³. This creates an interesting comparison to our qPCR data for NMD factor transcript expression.

The balance of NMD factor concentrations for NMD activity is delicate, and not necessarily intuitive from the RT-qPCR results. Several NMD factors, such as UPF1, UPF2, and UPF3B have shown to have roles in normal translation in addition to NMD²⁴. UPF1, SMG1, and SMG6 also have wide ranges of functions outside of their central roles

in NMD²⁵. We must be critical of assigning NMD factors a single function, as further investigation into their global roles could help us to understand NMD in more detail.

Our findings suggest that it is possible that other, currently unknown, intracellular mechanisms are affecting NMD. These could include cell cycle status, levels of NMD factors, the efficiency of translation termination, and the efficiency of splicing and deposition of EJCs, to name a few. In addition, several other mechanisms in RNA processing have been shown to rescue the expression of loss-of-function gene variants in individuals, and could also affect NMD efficiency in cells. One pertinent mechanism is stop codon readthrough, which could enable cells to selectively evade NMD for certain transcripts. Cells may also modify gene expression by varying transcriptional start and stop sites and translational start sites. Alternative splicing and C-terminal truncation have also been shown to play a critical role in RNA processing and gene expression²⁶.

The findings of NMD variability on multiple scales have led to more questions about the NMD pathway and peripheral influences on NMD. Interestingly, however, one study has found that NMD efficiency directly correlates with the concentration of a peripheral EJC protein RNPS1 in cells. The NMD cofactor protein RNPS1 has a role in tethering the EJC to NMD machinery, and low cellular concentrations of RNPS1 has been shown to correlate with NMD activity. This is the first study that shows NMD efficiency directly correlated with NMD cofactor abundance²⁰. This simple correspondence of protein concentration to NMD efficiency suggests factors that are essential to the cell for many purposes, not solely NMD, may influence NMD rates. These correspondences must be investigated further and in many cell types to understand the scope of this finding.

NMD efficiency must be evaluated within a system of other molecular mechanisms

Because many NMD factors have additional roles in other regulatory mechanisms, we aimed to investigate NMD activity in relation to other intracellular mechanisms. Notably, NMD factors have important roles in regulating the cell cycle. Thus, we attempted to analyze NMD efficiency in relation to cell cycle stage by comparing NMD efficiency level to cell cycle status based on the abundance of cell cycle proteins in each NMD efficiency group. Our results did not suggest a clear correlation between NMD and cell cycle stage, prompting more questions on other cellular mechanisms that may be

affecting NMD efficiency. However, while our results did not suggest a relationship, the potential correlation between NMD and cell cycle must be investigated further.

We are still interested in the potential relationship between NMD and cell cycle, as NMD factors, specifically UPF and SMG proteins, have been shown to have numerous roles in DNA maintenance and repair⁹. We expect that these roles could cause NMD factor concentrations to correlate with cell cycle stage.

For example, UPF1 has a role in Staufen 1 (STAU1)-mediated mRNA decay (SMD), a mechanism for mRNA decay and regulation that does not depend on a PTC for target recognition, but is targeted by a specific stem-loop structure and lack of a 3' poly(A) tail²⁶. UPF and SMG factors have also been shown to have regulatory roles in telomere maintenance, and SMG5 and SMG6 have been shown to directly interact with telomerase in the mechanism²⁶. Specifically, UPF1, SMG1, SMG6, and UPF2 have been shown to negatively regulate the binding of telomeric repeat-containing RNA to telomeres, to regulate the facilitation on telomeric heterochromatin assembly²⁷.

UPF1 is additionally involved in several other cellular mechanisms, including a role in histone mRNA degradation during the end of S phase and upon the inhibition of DNA replication²⁶. At the end of S phase, UPF1 is recruited to the histone and SMG1 phosphorylates UPF1, initiating degradation. Through the control of histone mRNA decay, UPF1 could have a direct role in cell cycle regulation. UPF1 binding to chromatin has been shown to increase throughout G1 and is maximized in S-phase, which is suggestive of a regulatory role. UPF1 also has a role in DNA repair, due to its upregulation in response to DNA damage²⁸. For example, in HeLa cells, downregulating UPF1 has been shown to cause S-phase arrest²⁶.

Our results emphasize the broad importance of investigating NMD in relation to other cellular mechanisms. Our experiments, however, had several limitations that prevented us from making strong conclusions. Primarily, we were limited by possible errors of measuring protein concentrations before loading samples for immunoblots. This was revealed through immunoblotting for Histone H3 as a control, which showed a range in general protein concentrations throughout the samples, rather than a controlled concentration. In addition, RT-qPCR samples were pipetted manually, which could have exacerbated standard deviations among replicates. It is also important to note that our

results for immunoblotting and RT-qPCR were limited because of inherent NMD variability within high, medium, and low efficiency cell groups. In future experiments, we would increase the number of NMD efficiency groups for increased specificity.

Our studies should be progressed by applying different methods to analyze the relationship between cell cycle and NMD. In addition, NMD factor expression levels among NMD efficiency groups should be replicated again through RT-qPCR for accuracy and via immunoblotting to measure protein level of the NMD factors. Future analysis of the relationship between NMD efficiency and NMD mRNA and protein level should include measuring protein levels to determine whether protein and mRNA levels correspond.

Furthermore, it would also be of interest to analyze the NMD efficiencies of daughter cells of cells from each sorted NMD efficiency group to understand if NMD efficiency is maintained cross-generationally.

Finally, future studies must investigate the aforementioned cellular mechanisms alongside NMD efficiency, as they are dependent upon NMD factors. This avenue of research has yet to be investigated, and the complex relationship between various intracellular mechanisms must be elucidated to uncover the cellular causes of disparities in NMD efficiencies. The overall importance of this research is evident in its application to a variety of currently untreatable human diseases. In order to move closer to developing cures, we must look closely at complexity of the molecular mechanisms behind diseases. To conclude, we have found a very basic and inherent variability in NMD efficiency among human embryonic kidney cells in homogenous cell culture. This finding gives rise to more questions on how other unknown factors and mechanisms may contribute to NMD efficiency.

Methods

DH5 α *E. coli* Transformation. 1-10ng of NMD inducing plasmid (pRKB208) and control plasmid (pRKB217) (Appendix 1) were taken up by DH5 α cells (Thermo Fisher). Cells were incubated on ice for 30 minutes, heat shocked in a water bath at 42°C for 20 seconds, and incubated on ice for 2 minutes. Super optimal broth with catabolite repression (SOC) media was added to each cell vial. Cells were incubated while rotating at 225RPM at 37°C for 1 hour. 50 μ L cells were spread on carbenicillin (RBI) plates and incubated at 37°C overnight. Single colonies were grown up with 200mg of 1000X carbenicillin in lysogeny broth (LB). For plasmid preparation, Macherey-Nagel NucleoBond Xtra EF Maxiprep plasmid purification procedure was followed.

HEK293 cell culture. HEK293 cells were activated and then passaged every 2-3 days. Cells were activated by warming cells and adding 1mL cells to Dulbecco's Modification of Eagle's Medium (DMEM) (VWR). Cells were spun down and resuspended in DMEM, transferred to a 15cm plate, and incubated at 37°C. To passage cells, media was removed, cells were submerged in phosphate buffered saline (PBS) (Thermo Fisher), PBS was removed, 1X trypsin (TryLE Express Enzyme) (Thermo Fisher) was added to cells, cells were incubated at 37°C for 5 minutes, and DMEM was added to the plate. Cells were then added to fresh DMEM at a 1:7 ratio for 2 days and 1:10 for 3 days in 10cm plates.

HEK293 transfection. Lipofectamine 3000 transfection kit (Thermo Fisher) was used according to manufacturer instructions. Lipofectamine 3000 (Thermo Fisher), Opti-MEM (Thermo Fisher), and Reagent P 3000 (Thermo Fisher) were combined at high dilutions from the manual's instructions and added to cell cultures. 25ug DNA was added to each plate. Cells were incubated at 37°C overnight. Old media was removed and DMEM with 1ug/ μ L doxycycline (Fisher Scientific) was added to each 10cm cell plate. Plates were incubated at 37°C overnight and imaged with a fluorescent microscope.

XDP100 flow cytometry and single-cell sorting. Cells were washed with PBS, trypsin-EDTA (Thermo Fisher) was added to each plate, and cells were incubated at 37°C for 5 minutes. DMEM was added to conical tubes and cells were transferred to tubes. Cell concentrations were measured using a cell counter (Invitrogen) according to the manufacturer's instructions. Cells were spun down, media was removed and cells were resuspended in sorting buffer, composed of 1X PBS (Invitrogen), 1mM Edta, 25mM HEPES and 1% fetal bovine serum (FBS) or bovine serum albumin (BSA) and filter sterilized. XDP100 (Beckman Coulter) was used for cell sorting and flow cytometry. Forward scatter and side scatter were measured. KAT fluorescence was measured against GFP fluorescence. GFP and KAT individual transformants control were used. Sorted cells were collected and stored in PBS. R programming was used for flow cytometry analysis and KS tests were used for statistical analysis.

RNA extraction. Cells were resuspended in trizol and chloroform (Fisher Scientific) was added to each sample. Samples were shaken for 30 seconds and cells were incubated for 2-3 minutes at room temperature. Cells were spun down at 12,000 x g for 10 minutes at 4°C and the aqueous phase was transferred and isopropanol (Fisher Scientific) was added and mixed well. Samples were incubated at room temperature for 10-20 minutes. Samples were spun down at 12,000 x g for 10 minutes at 4°C and 75% ethanol (Fisher Scientific) was added to the pellet. Samples were vortexed and spun down at 7,500 x g for 5 minutes at 4°C (repeated twice) and pellets were air dried for 5-10 minutes and dissolved in nuclease-free water (VWR). RNeasy (Qiagen) clean-up protocol was followed as per manufacturer instructions.

Protein extraction. Trizol (Thermo Fisher) was added to each sample, followed by chloroform. Samples were shaken vigorously for 15 seconds and incubated for 2-3 minutes at room temperature. Samples were spun down at 12,000 x g for 15 minutes at 4°C and the phenolic phase was transferred. Ethanol was added to each sample and samples were mixed well and incubated for 2-3 minutes at room temperature. Samples were spun down at 2,000 x g for 5 minutes at 4°C. The phenol-ethanol supernatant was transferred and isopropanol was added to each sample. Samples were mixed well and

incubated for 10 minutes at 4°C. Supernatant was removed and 0.3M guanidine HCl in 95% EtOH was added to each pellet. Samples were mixed and incubated at room temperature for 20 minutes and spun down at 7,500 x g for 5 minutes at 4°C (repeated 3 times). Ethanol was added to each pellet. Samples were mixed, incubated at room temperature for 20 minutes, and spun down at 7,500 x g for 5 minutes at 4°C. Pellets were air dried for 5-10 minutes and dissolved in sample buffer (0.5M Tris, 5% SDS) on a heat block followed by an electrobath. Protein levels were calculated using a Pierce BCA protein assay kit (Thermo Fisher) according to manufacturer instructions.

Immunoblots. Immunoblot protocol was followed as outlined in the Thermo Fisher Scientific manual. Sample buffer (Thermo Fisher Scientific) was added to each protein sample. 10ug protein was added to each gel well. The same protein samples were used for each immunoblot. Blots were stripped and reblotted using the Thermo Fisher Scientific stripping protocol. 10µL 1,000X FLAG antibody (Novus) was used to detect the FLAG tag embedded in the GFP construct. 10µL 1,000X HA antibody (Cell Signaling Technology) was used to detect the HA embedded in the KAT construct. 0.5µL 20,000X H3 antibody (Abcam) was used as a control. 1µL 800CW goat anti-rabbit (LI-COR Biosciences) secondary antibody was used. The blot was stripped and 0.4µL 10,000X UPF1 antibody (IDT) was used. 1µL 800CW goat anti-rabbit secondary antibody was used. 40µL cell cycle antibody cocktail (Abcam) was added. 1µL 800CW goat anti-rabbit secondary antibody was used for probing. Immunoblots were imaged.

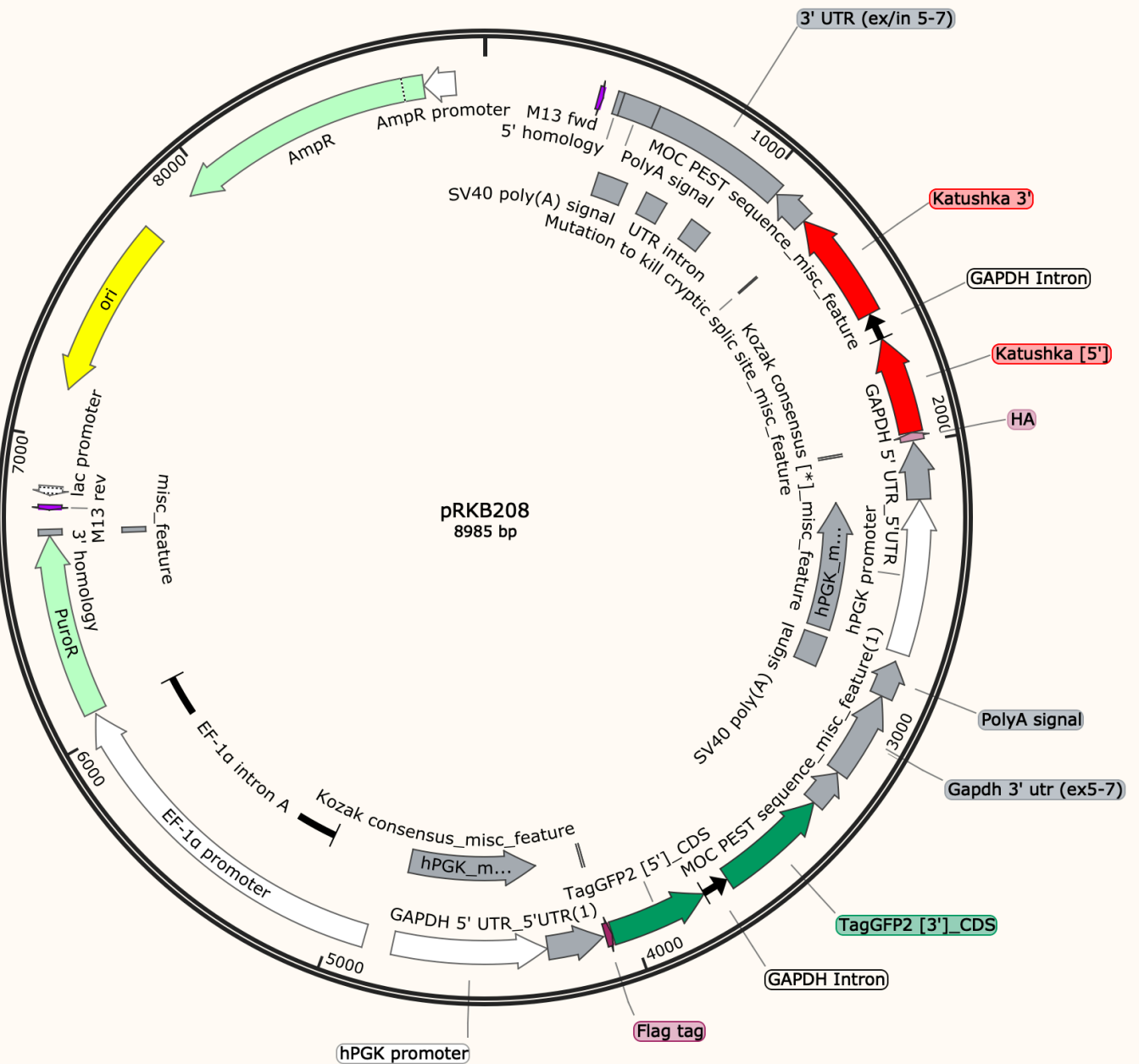
RT-qPCR. SuperScript IV VILO Thermo Fisher protocol was followed to generate cDNA. A standard curve was created with 1:4 to 1:1024 dilutions. Primers used are outlined in Appendix 2. Primer master mixes were made with 0.1X primers in water. SYBER green (Bio-Rad) + primer master mixes were created at a 12.5:1 ratio. A 384 well plate was used for qPCR. SYBER + primer master mix was added to each well and 0.4X cDNA sample was added to each appropriate well. Each cDNA sample was tested in triplicates for accuracy. Samples were run at the following parameters with a Bio-Rad CFX384 qPCR reader: polymerase activation and DNA denaturation at 95°C for 30 seconds;

denaturation at 95°C for 5 seconds; annealing/extension at 60°C for 30 seconds; default melt curve analysis; 40 cycles.

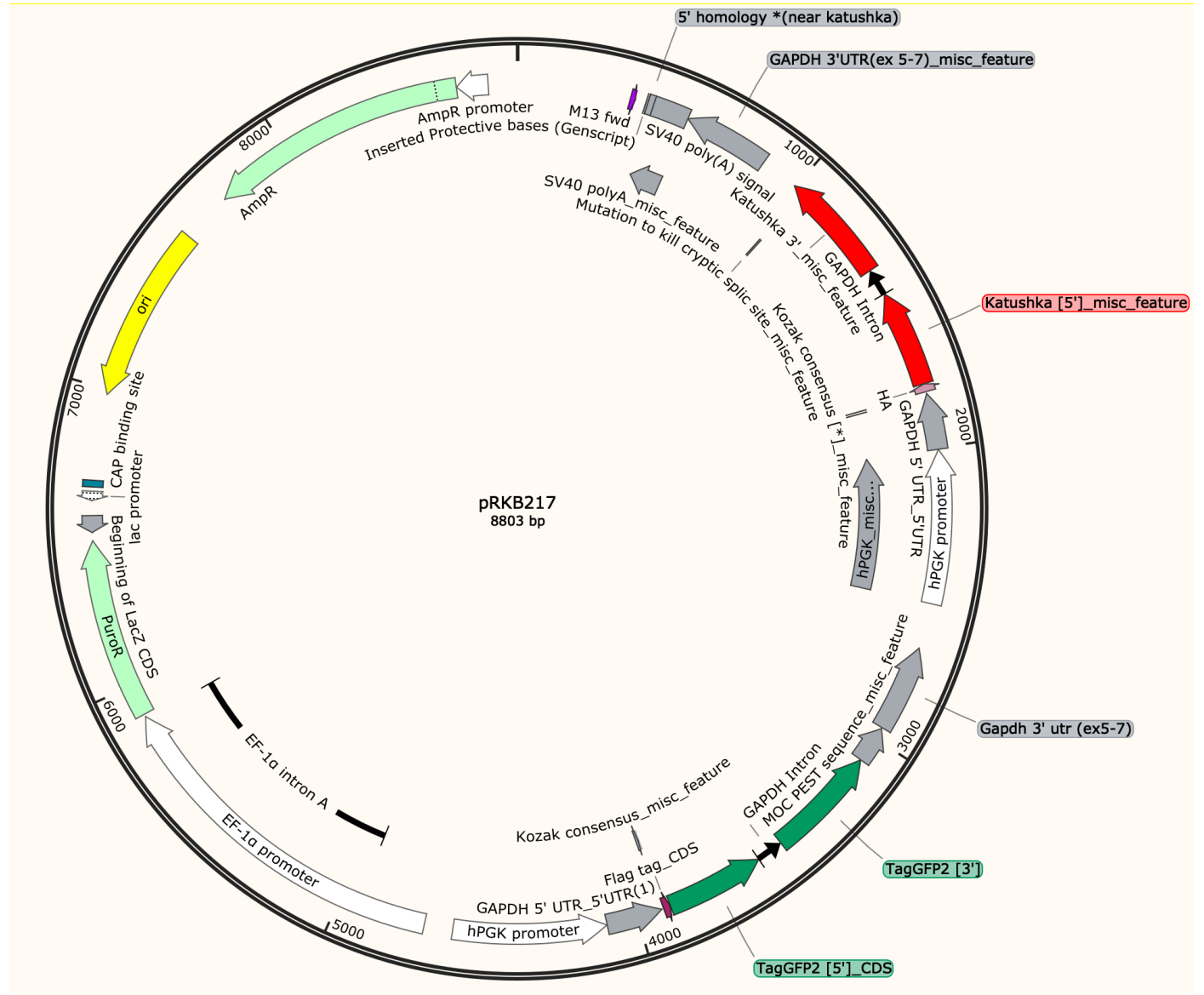
Appendix

Appendix 1: Plasmid Constructs.

pRKB208



pRKB217



Appendix 2. Primer Sequences (Integrated DNA Technologies).

F= Forward Primer

R= Reverse Primer

Primer Name	Sequence (5'-3')
RPL27-F	GCAAGAAGAAGATCGCCAAG
RPL27-R	TCCAAGGGGATATCCACAGA
SRP14-F	GAGAGCGAGCAGTTCCTGAC
SRP14-R	GTTTGGTTCGACCGTCATACT
HNRNPD-F	GCAGAGTGGTTATGGGAAGG
HNRNPD-R1	ATGAAGTCCCGCTGTTGG
HNRNPD-R2	CACCACCTGTTGGGGATAAG
SRSF2-F	GTGTCCAAGAGGGAATCCAA
SRSF2-R1	AGGAGACCGCAGCATTTTCT
SRSF2-R2	TGCTTGCCGATACATCATTT
SRSF3-F	TGGAAGTGTGCAATGGTGAA
SRSF3-R1	GGGTGGTGAGAAGAGACATGA
SRSF3-R2	CTTGGAGATCTGCGACGAG
DUX4-F	GGCCCGGTGAGAGACTCCACAC
DUX4-R	CCAGGAGATGTAAGTCTAATCCAGGTTTGC
ZSCAN4-F	TGGAAATCAAGTGGCAAAAA
ZSCAN4-R	CTGCATGTGGACGTGGAC
KHDC1L-F	CACCAATGGCAAAGCAGTGG
KHDC1L-R	TCAGTCTCCGGTGTACGGTG
UPF1-F1	GCTCGGCACTGTTACCTCTC
UPF1-R1	GTACGCCTCCACGCTCAT
UPF1-F2	CAGCTCGCAGACTCTCACTTT
UPF1-R2	TGCGTCTGGCTAGGAAGAGT
LEUTX-F	CTGCAGCACACAGCTGATCG
LEUTX-R	CTTGCCCTCGCCCAACTTAC
DUX4-CA-F	CACCACCACCACCAAGG
DUX4-CA-R	GAACGGACGTGAAGAATGTG
SMG1-F	TGGGAAAGACCACCACTGCACA
SMG1-R	TGCATGTGTTGACTGGCCTGCT
UPF2-F	CCCAGCTCCAGCAAACACCAAT
UPF2-R	ATCAACGTCTCCTCCACCAGT

UPF3a-F	AGAAGCTGTCGGCCCTAGAA
UPF3a-R	GGATGACCACCTTGCTCAGG
UPF3b-F	AAACAAGGATCGTCCAGCGA
UPF3b-R	TGGCTAATACCACTTTCCTGCT
SMG5-F	TGGAGGCTGTGCATCGACTTGA
SMG5-R	AGCTCACGCAGCTTGTTCTCA
SMG6-F	TGGCCAGCTGGGTAACAACGAT
SMG6-R	TTTGCTGGCGGGCATGAAGT
SMG7-F	CAAACATAGACCGCAGGGGCAA
SMG7-R	TGGGGTCCTCAAACGGCATTCT
NBAS-F	TGCAAGTGCGATTGTGCCCTGA
NBAS-R	TTCTCACCTGCAACCCTCAGCA
DHX34-F	TTTTCCACACGCAGGCCAAGCA
DHX34-R	TCTTGTCTTGTCGTCTCGGCT
GNL2-F	TTTCACCACGTAAGCCGGACCT
GNL2-R	TGCACTCGATCTGGGTTTGTGC
SEC13-F	TGGAGAGAGGAAAACGGCAC
SEC13-R	AGCACACCGAGTTCCTGAG
SMG8-F	TCCTCTCAAGGTAGAGGGCT
SMG8-R	CTGGCTGAACCTGAGGCATT
SMG9-F	TGCCAGCGAAGAGACAAGCACT
SMG9-R	TGGCTGTTTTGACCGCTCTGCT
PNRC2-F	ATCGGTGTTTCGTGGGCTTT
PNRC2-R	AGTCCTAGTGACTTCAAGCTCGG
RUVBL1-F	TGCATCCAACCGAGGCAACTGT
RUVBL1-R	GCAGCTGCACTGAGTACCTGTT
RUVBL2-F	AAGGATTGAGCGAATCGGTGCC
RUVBL2-R	GGAGGCGCACATGAGGTGATTT
MOV10-F	AGGCACATTGTTACGGGCACCA
MOV10-R	GCAAGTGCTTCACCACCTGCTT

Acknowledgements

I thank Dr. Sujatha Jagannathan, for designing and troubleshooting this project and welcoming me into the lab. This project was funded by the RNA Bioscience Initiative at the University of Colorado Denver Anschutz Medical School. I thank Dr. Sara Hanson, for her constant advising and support of the writing, analyzing, and presenting of this thesis. I would also like to thank Dr. Jennifer Garcia for assisting with data analysis, and Dr. Darrell Killian for reviewing this thesis. Finally, I thank Heather Johns and the Bradley Laboratory from Fred Hutchinson Cancer Research Center in Seattle, WA for helping to design the NMD reporter plasmid constructs.

References

1. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **16**, 665–677 (2015).
2. Karam, R. *et al.* The unfolded protein response is shaped by the NMD pathway. *EMBO Rep.* **16**, 599–609 (2015).
3. Gardner, L. B. Hypoxic Inhibition of Nonsense-Mediated RNA Decay Regulates Gene Expression and the Integrated Stress Response. *Mol. Cell. Biol.* **28**, 3729–3741 (2008).
4. McIlwain, D. R. *et al.* Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci.* **107**, 12186–12191 (2010).
5. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
6. Huang, L. *et al.* RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD. *Mol. Cell* **43**, 950–961 (2011).
7. Jankowsky, E. RNA Helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36**, 19–29 (2011).
8. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* **44**, 1483–1495 (2016).
9. Azzalin, C. M. & Lingner, J. The Double Life of UPF1 in RNA and DNA Stability Pathways. *Cell Cycle* **5**, 1496–1498 (2006).

10. Yamashita, A. *et al.* SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev.* **23**, 1091–1105 (2009).
11. Loh, B., Jonas, S. & Izaurralde, E. 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 (2013).
12. Weischenfeldt, J. *et al.* Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* **13**, R35 (2012).
13. Linde, L., Boelz, S., Neu-Yilik, G., Kulozik, A. E. & Kerem, B. The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells. *Eur. J. Hum. Genet.* **15**, 1156–1162 (2007).
14. Frischmeyer, P. A. & Dietz, H. C. Nonsense-Mediated mRNA Decay in Health and Disease. *Hum. Mol. Genet.* **8**, 1893–1900 (1999).
15. Nguyen, L. S., Wilkinson, M. F. & Gecz, J. Nonsense-mediated mRNA decay: Inter-individual variability and human disease. *Neurosci. Biobehav. Rev.* **46**, 175–186 (2014).
16. Feng, Q. *et al.* A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *eLife* **4**, (2015).
17. Isken, O. & Maquat, L. E. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.* **9**, 699–712 (2008).
18. Nasif, S., Contu, L. & Mühlemann, O. Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Semin. Cell Dev. Biol.* **75**, 78–87 (2018).

19. Lou, C. H. *et al.* Posttranscriptional Control of the Stem Cell and Neurogenic Programs by the Nonsense-Mediated RNA Decay Pathway. *Cell Rep.* **6**, 748–764 (2014).
20. Viegas, M. H., Gehring, N. H., Breit, S., Hentze, M. W. & Kulozik, A. E. 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 (2007).
21. Pereverzev, A. P. *et al.* Method for quantitative analysis of nonsense-mediated mRNA decay at the single cell level. *Sci. Rep.* **5**, (2015).
22. Resta, N. *et al.* A homozygous frameshift mutation in the ESCO2 gene: Evidence of intertissue and interindividual variation in Nmd efficiency. *J. Cell. Physiol.* **209**, 67–73 (2006).
23. Nguyen, L. *et al.* Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. *Mol. Psychiatry* **17**, 1103–1115 (2012).
24. Wilkinson, M. F. A new function for nonsense-mediated mRNA-decay factors. *Trends Genet.* **21**, 143–148 (2005).
25. Nicholson, P. *et al.* Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell. Mol. Life Sci.* **67**, 677–700 (2010).
26. Jagannathan, S. & Bradley, R. K. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Res.* **26**, 1639–1650 (2016).

27. Azzalin, C. M., Reichenbach, P., Khorიაuli, L., Giulotto, E. & Lingner, J. Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science* **318**, 798–801 (2007).
28. Azzalin, C. M. & Lingner, J. The Human RNA Surveillance Factor UPF1 Is Required for S Phase Progression and Genome Stability. *Curr. Biol.* **16**, 433–439 (2006).