Peptidyl-tRNA hydrolysis rate influences the efficiency of nonsense-mediated mRNA decay

Divya Kolakada[1,2], Rui Fu[3#], Nikita Biziaev[4], Alexey Shuvalov[4], Mlana Lore[2], Amy E. Campbell[1], Michael A. Cortázar[1], Marcin P. Sajek[1], Jay R. Hesselberth[1,3], Neelanjan Mukherjee[1,3], Elena Alkaeva[4], and Sujatha Jagannathan[1,3,5*]

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA.

[2]Molecular Biology Graduate Program, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA.

[3]RNA Bioscience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA.

[4]Engelhardt Institute of Molecular Biology, The Russian Academy of Sciences, 119991 Moscow, Russia.

[5]Lead contact

*Corresponding Author: sujatha.jagannathan@cuanschutz.edu

[#]Current Address: New York Genome Center, New York, NY

Keywords: NMD, PTC, translation, termination, splicing

SUMMARY

Nonsense-mediated mRNA decay (NMD) is a quality control mechanism that prevents the accumulation of harmful truncated proteins by degrading transcripts with premature termination codons (PTCs). NMD efficiency varies across many contexts, but the factors that influence this variability remain poorly understood. Here, we find an enrichment of glycine (Gly) codons preceding a PTC in common nonsense variants in contrast with a depletion of Gly codons preceding a normal termination codon (NTC). Gly-PTC contexts have higher NMD activity compared to an alanine-PTC context, and this effect is stronger on NMD substrates with long 3'UTRs. We used a massively parallel reporter assay to test all possible combinations of -2 and -1 codons, the PTC, and the +4 nucleotide to assess comprehensively how PTC sequence context affects NMD efficiency. A random forest classifier revealed that peptidyl-tRNA hydrolysis rate during translation termination was the most important feature in discriminating high and low NMD activity. We show with *in vitro* biochemical assays that Gly-TC contexts have the slowest termination rate compared to other codons. Furthermore, Gly-PTC enrichment is most pronounced in genes that tolerate loss-of-function variants, suggesting that enhanced NMD of Gly-PTC context has shaped the evolution of PTCs. Based on these findings, we propose that NMD efficiency is modulated by the "window of opportunity" offered by peptidyl tRNA hydrolysis rate and thus, translation termination kinetics.

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) eliminates transcripts that contain protein-truncating variants (PTVs), often causing loss of gene function. As a result, PTVs underlie Mendelian diseases, including cystic fibrosis, Duchenne muscular dystrophy, and β-thalassemia (1,2). Much of our understanding of NMD comes from studying disease-causing PTVs, including the "50-55 nt rule", which posits that premature termination codons (PTCs) trigger NMD >50-55 nucleotides (nts) upstream of the 3'-most exon-exon junction (3). Here, the terminating ribosome fails to evict the exon-junction complex (EJC), which is a potent trigger for NMD (4,5). An alternative mode of NMD is triggered when a PTV creates a long 3′ untranslated region (UTR) where, in the absence of EJCs, RNA binding proteins (RBPs) such as UPF1 help to recruit decay machinery (6,7). In both cases, the terminating ribosome, along with downstream RBPs in the 3′ UTR, serves as the platform for the assembly of NMD factors. Thus, the terminating ribosome at the stop codon is essential to trigger NMD.

Studies over the years have shown that the "50-55 nt rule" does not encompass the breadth of NMD efficiencies observed across transcripts and even across different PTCs within the same transcript (8-10). For example, the 50-55 nt rule only predicted the fate of ~50% of PTVs in The Cancer Genome Atlas database; while additional factors such as exon length could account for another ~20% of NMD variability, that still leaves ~30% of NMD variability that could not be explained by any known rule (11). Another study that investigated nonsense variants that are highly frequent in the healthy human population found that ~50% escape NMD via mechanisms including alternative splicing, stop codon readthrough, and alternative translation initiation (12). In contrast, ultrarare nonsense variants tend to show high NMD efficiency (13). These disparate observations underscore the ascertainment bias, based on the substrates investigated and their phenotypic consequences, that has influenced our view of NMD.

If every premature termination event that satisfies the 50-55 nt rule does not lead to NMD, what other factors influence the probability and efficiency of NMD? One potential factor is translation termination efficiency, which varies and likely influences NMD. Translation termination efficiency can be conceptualized in two ways: (1) the fidelity of termination, reflecting the likelihood of stop codon readthrough, and (2) the kinetics of termination, reflecting the dwell time of the terminating ribosome at the stop codon. Both these aspects are influenced by cis-acting elements (14-18). For example, specific elements within the stop codon sequence context, as well as certain viral structures, can reduce termination fidelity (14-16,19-21). Additionally, the identity of the nucleotide immediately after the stop codon and the peptide-

tRNA hydrolysis rates of the amino acid preceding the stop codon impact the rate of peptide release (17,18). However, most of these studies focus on normal termination, which has been proposed to be distinct from premature termination (22-25), and thus, how these factors influence NMD is unknown.

In this study, we hypothesized that investigating sequence contexts of PTVs that occur in healthy individuals may offer novel insight into how these sequences influence NMD. Indeed, we found that Gly-PTC, a context highly enriched in healthy individuals, enhances NMD activity. To systematically assess the impact of PTC sequence context on NMD activity, we performed a massively parallel reporter assay and found a broad range of NMD efficiencies associated with different sequences. Finally, we show that translation termination kinetics could explain sequence-based differences in NMD efficiency. Taken together, our results provide a framework to understand how factors beyond the central NMD triggers can profoundly influence NMD efficiency. We propose PTC sequence can alter NMD efficiency by modulating the kinetic window of opportunity for NMD offered by the terminating ribosome.

RESULTS

Glycine codons are enriched preceding PTCs and depleted preceding NTCs. Healthy individuals harbor ~100 putative loss of function genetic variants, most of which are protein-truncating (26). In previous work, we showed that highly frequent PTVs (minor allele frequency (MAF) > 5%) are enriched for NMD escape events (12). Here, we focused on rare variants with MAF ≤ 1% among healthy individuals that would be expected to undergo NMD but perhaps with varying efficiencies. We analyzed sequence enrichment in PTC contexts of these variants within 10 amino acids preceding PTCs found in the gnomAD database (n=151,332, v2.1.1), which contains genetic variant data for >700,000 healthy humans (27) (**Fig. 1A)**. We compared the enrichment of these amino acids to that of normal termination codon (NTC) contexts in the reference genome (n=23,066, **Fig. 1B**). Gly codons stood out as the most enriched at the -1 position of PTCs (**Fig. 1A** and **C**) and least enriched before the NTC (**Fig. 1B** and **D**). Threonine (Thr) and proline (Pro) were the next most enriched in PTC compared to NTC (**Fig. 1E**). Other amino acids also show the opposite trend – most enriched before an NTC and least enriched before a PTC – including asparagine (Asn), valine (Val), and aspartic acid (Asp). Nevertheless, the strongest effect was seen for Gly-NTC versus Gly-PTC contexts (**Fig. 1E**). These data suggest that the enrichment of Gly immediately upstream of a PTC (Gly-PTC) has a functional impact that allowed it to rise to high frequency in healthy individuals.

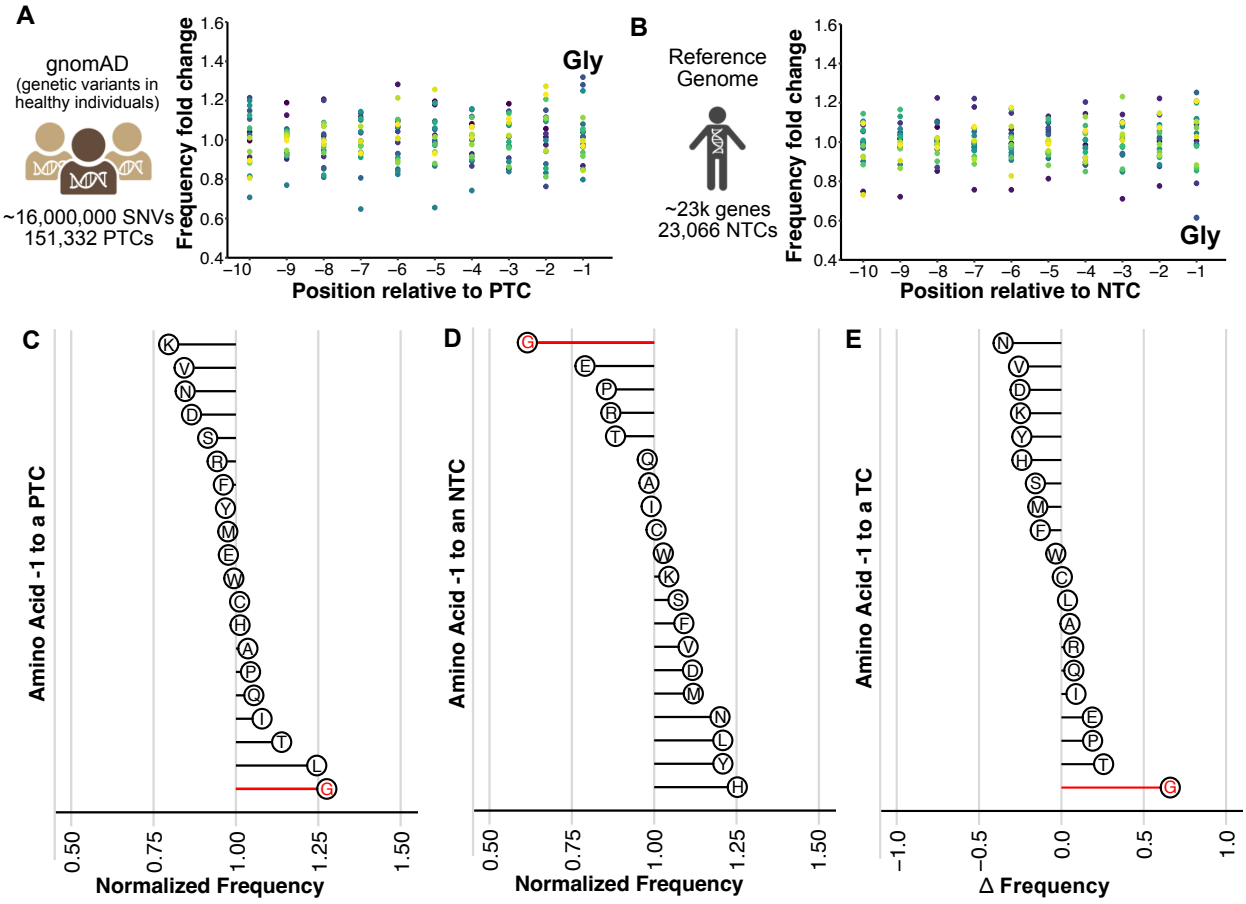


Figure 1. Gly is enriched preceding a PTC and depleted preceding an NTC. (**A-B**) The enrichment of amino acids in the 10 positions before a PTC (A) or the NTC (B) (**C-D**) Amino acid enrichment -1 to a PTC (C) or an NTC (D). (**E**) The frequency delta (enrichment before a PTC-NTC) for each amino acid at the -1 position.

Gly codons that precede a PTC enhance NMD efficiency. A previous study showed that Gly acts as a C-terminal degron and is rarely found at the C-terminal end of the normal proteome, suggesting one possible biological basis for Gly-PTC enrichment (28,29). It is also possible that Gly-PTC stimulates PTC readthrough, leading to NMD escape and some functionality of the protein product (30), or that Gly-PTC enhances NMD and elimination of the mRNA transcript. Readthrough would confer a benefit to the organism as it would preserve gene function, while elimination of the truncated protein or encoding transcript might have a neutral effect if the gene is otherwise non-essential.

To distinguish between these hypotheses, we used a set of NMD reporters (31) that enabled the study of EJC-independent and EJC-enhanced forms of NMD, as well as NMD-associated mechanisms such as PTC readthrough and truncated protein degradation. The "NMD**+**" reporter consists of a bi-directional promoter that drives BFP as a transfection control, in addition to an RFP and GFP containing transcript (**Fig 2A**, left). A stop codon between RFP and GFP creates a long 3′ UTR on GFP, making the transcript a target of the EJC-independent NMD. In another version of the reporter, an intron within the GFP sequence renders the transcript an ideal EJC-enhanced NMD substrate (**Supp. Fig 1A**). A T2A signal after the stop ensures independent expression of RFP and GFP proteins in the event of a readthrough of the stop codon. Using this reporter, we varied the codon at the -1 position of the PTC (TAA), testing two different glycine codons or an alanine codon. We also tested these sequences at the NTC of an "NMD–" reporter, which lacks the GFP 3′ UTR and so behaves as a normal transcript, allowing us to test the effect of glycine as a C-end degron independent of NMD. (**Fig 2A**, middle).

If Gly-PTC promotes readthrough, we would expect to see production of GFP protein. If Gly-PTC promotes efficient protein degradation, then we would expect to see much less RFP protein accumulation in both NMD**+** and NMD**–** contexts compared to the Ala-PTC context. If Gly-PTC promotes efficient NMD, we would expect to see less RFP protein levels compared to Ala-PTC for the NMD**+** contexts but no difference for the NMD**–** contexts (**Fig 2A**, right). We transfected HEK293T cells with these reporters, treating them with either DMSO vehicle control or SMG1i, an inhibitor of the SMG1 kinase activity, and thus NMD. Transfected cells were imaged for BFP, RFP, and GFP. We also included a control construct that lacks the stop codon between the RFP and GFP, and therefore robustly expresses all three fluorescent proteins.

As shown in Fig 2B, all reporters yielded robust BFP expression. While the no PTC control construct produced GFP, there was no GFP protein detected with either the glycine or alanine PTC contexts (**Fig 2B**), ruling out the first hypothesis that a Gly-PTC context promotes readthrough. Notably, both EJC-independent NMD**+** reporters with Gly-PTC resulted in less RFP protein compared to the Ala-PTC reporter, a difference which recovered upon treatment with SMG1i (quantified in **Fig 2C**). In comparison, only a modest reduction in in RFP protein was observed for the Gly-PTC NMD**–** reporters compared to the Ala-PTC reporter (**Fig 2B**), suggesting that a C-end glycine degron was not sufficient for the protein depletion we observed in the NMD+ reporters. No RFP was detected from any of the EJC-enhanced NMD reporters, consistent with their high NMD activity (**Supp. Fig 1B-C**). These data suggest that Gly-PTC inherently influences NMD efficiency through a hitherto unknown mechanism.

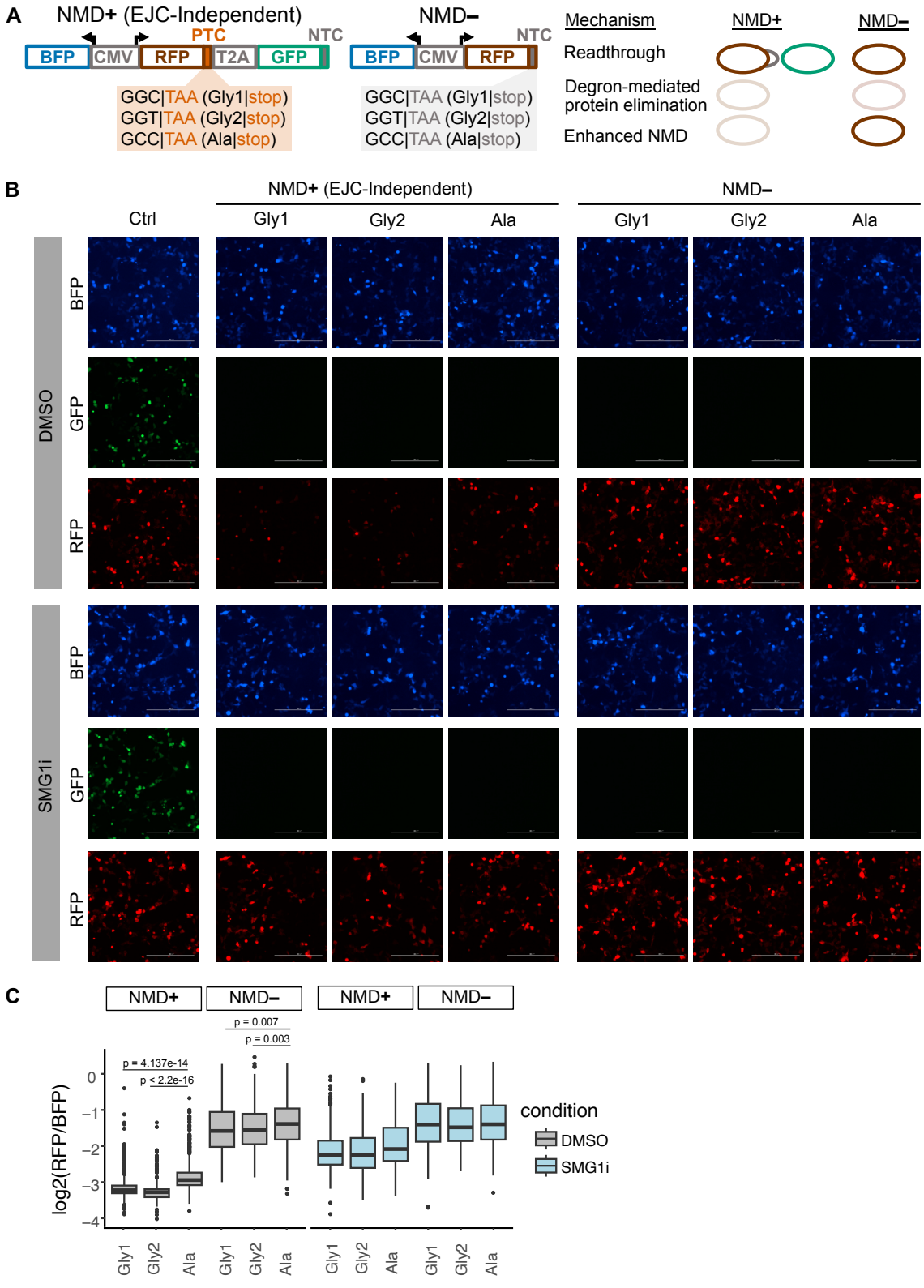


Figure 2. Gly-PTC contexts enhance NMD. (A) Schematic of NMD reporter system used. Left: EJC-independent NMD+ reporter, middle: NMD– reporter; right: expected outcomes. (B)

Fluorescence imaging of BFP, GFP, and RFP from cells transfected with NMD reporters containing Gly-PTC or Ala-PTC reporters, treated with a DMSO as a control or SMG1i. Ctrl refers to a construct that robustly expresses all three fluorescent proteins. Gly1 and Gly2 are the GGC and GGT codons for Gly. The codon used for Ala was GCC. (C) Quantification of fluorescence from panel (B).

Massively parallel reporter assay (MPRA) captures widespread sequence-dependent variation in NMD efficiency. Having established that Gly-PTC causes enhanced NMD, we systematically examined the relationship between NMD efficiency and the PTC sequence context. We used a massively parallel reporter assay wherein we varied the 6 nucleotides before the PTC, the PTC itself, and 1 nucleotide after the PTC in an unbiased fashion (N6TRRN; corresponding to 65,536 unique sequences; **Fig. 3A**) and asked how these sequences impacted NMD efficiency in both EJC-independent and EJC-enhanced reporters. Because we employed stop codons with "TRR" for array-based oligonucleotide synthesis, 25% of our sequence library consists of a tryptophan (Trp) codon (UGG) instead of a PTC and serve as control sequences that do not undergo NMD. All sequences were incorporated into both EJC-independent and EJC-enhanced fluorescent reporter backbones (**Supp. Fig. 2A-B**) and integrated into HEK293T cells via Cre-Lox mediated recombination (32). These cells were then treated with either DMSO or SMG1i, harvested for RNA, and subjected to targeted sequencing. Sequencing reads were analyzed by counting the occurrences of each reporter within a sample. The NMD activity of an individual reporter sequence was the log2-fold change in between DMSO-treated cells to SMG1i-treated cells calculated with DESeq2 (33) using the ~1/4th of control sequences containing the Trp codon rather than a stop codon for normalization. In this context, low transcript fold-change values indicate high NMD activity, and vice versa.

The NMD activity as measured by transcript fold changes for both the EJC-independent and EJC-enhanced reporters had a bimodal distribution, corresponding to the stop codon identity of the sequences in each population (**Fig. 3B**): sequences containing a stop codon were left-shifted compared to the sequences containing the Trp control, indicating that the former were targets of NMD. In the population of sequences undergoing NMD (i.e. those containing a stop codon) there was a wide range of NMD activity for both the EJC-independent and the EJC-enhanced reporters. EJC-enhanced reporter expression was reduced by 4.8 fold (log2FC: -2.27), which is greater than the 3.3 fold decrease (log2FC: -1.71) for EJC-independent reporters (**Fig. 3B**), consistent with EJC-enhanced targets undergoing more efficient NMD compared to

EJC-independent targets (34-36). These data indicate that NMD efficiency indeed varies based on PTC sequence context.

Next, we investigated if PTC sequence context could explain the wide range of observed NMD activity. Specifically, we were interested in the contribution of -2 and -1 amino acids, the -2 and -1 codons, the stop codon, and the +4 nucleotide. For each of these biological variables, we constructed a linear model to determine the extent to which it explained the variation in NMD activity (i.e. transcript fold change) of the associated sequences. The model allowed us to evaluate the effect size (beta coefficient) and statistical significance (p-value) of individual features within a variable on NMD activity relative to a control feature. For example, in the case of stop codons, we evaluated the effect of TAA, TAG, or TGA on NMD activity relative to UGG (Trp). Negative beta coefficients indicated stronger NMD activity, while positive beta coefficients indicated weaker NMD activity.

As expected, our modeling showed that all stop codons caused significantly stronger NMD activity compared to the Trp control for both EJC-independent and EJC-enhanced reporters (**Fig. 3C**). Notably, the TGA stop codon increased NMD efficiency by ~20% compared to the TAG stop codon for the EJC-independent reporters. However, this difference was not observed for the EJC-enhanced reporters where all stop codons elicit similar levels of NMD activity (**Fig. 3C**). For subsequent analyses, only transcripts expected to undergo NMD (i.e. transcripts with stop codons) were analyzed. When determining the effect of the +4 nt identity on NMD, we found that T, G, and C promoted weaker (~6% decrease) NMD activity compared to A for the EJC-independent reporters. A similar effect was present only for G compared to A for the EJC-enhanced reporters (**Fig. 3D**).

The linear modeling also showed that the identity of the -2 amino acid had a wide range of effects on NMD activity for the EJC-independent reporters, with any stop codon causing the strongest increase (~40%) in NMD activity compared to the control variable Ala (**Fig. 3E**). Codons for the remaining amino acids caused smaller changes in NMD activity, increasing NMD activity up to ~10% (Val) or reducing it up to 20% (Arg) compared to Ala. For the EJC-enhanced reporters, most -2 amino acids did not affect the efficiency of NMD. Interestingly, the presence of any stop codon -2 to a PTC weakened NMD by ~10% compared to the control variable Arg, opposite in trend to the EJC-independent reporters. Lys at the -2 position had the largest effect compared to Arg, weakening EJC-enhanced NMD by ~20% (**Fig. 3E**).

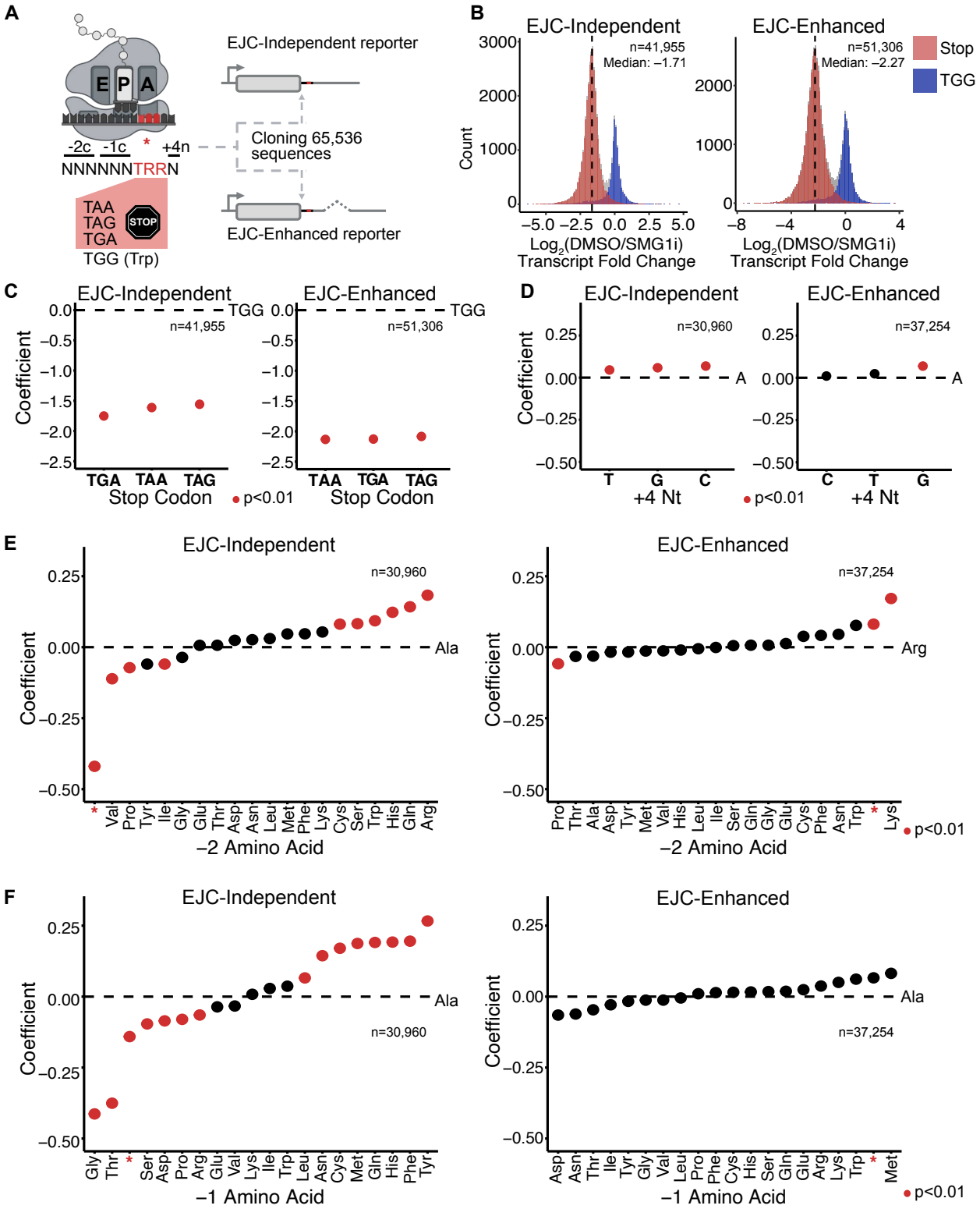


Figure 1. MPRA shows a wide range of PTC context-dependent NMD efficiencies. (A) Design of the MPRA. (B) Histogram of transcript fold change EJC-independent and EJC-enhanced NMD+ reporters comparing DMSO treated cells to SMG1i treated cells. (C-F) Dot

plots of the beta coefficient representing the effect of an individual variable on EJC-independent and EJC-enhanced NMD efficiency. The variables are the (C) stop codon identity, (D) the +4 nt identity, (E) the -2 amino acid, (F) and the -1 amino acid.

Remarkably, when examining the effect of the -1 amino acid on NMD efficiency, we found that Gly-PTC (the same context enriched in the gnomAD database; **Fig 1**) has the strongest effect on EJC-independent NMD activity, reducing it by 40% compared to Ala (**Fig. 3F**). Thr followed closely after and notably, both Gly-PTC and Thr-PTC cause stronger NMD than a stop codon in the -1 position. In contrast, Tyr-PTC had the least efficient NMD, weakening NMD by ~25% compared to Ala-PTC. These effects were not observed for the EJC-enhanced reporters where the -1 amino acid identity did not cause any apparent effects (**Fig. 3F**).

Next, we asked if there is codon-level specificity to the effect of amino acids on EJC-independent NMD activity. We found that the effects of the codons in the -2 and -1 positions were largely grouped by amino acid identity, with only a few that were significantly different (**Supp. Fig. 2C-D**, all codons compared to Lys-AAA). An exception to this trend was Leu at the -1 position whose codon TTA seems to have the opposite effect on NMD compared to the rest of the Leu codons. Reassuringly, all the effects at the amino acid level were replicated at the codon level with the stop codons having the strongest NMD activity at the -2 position and Arg the weakest (**Supp. Fig. 2C**), and Gly having the strongest NMD activity at the -1 position and Tyr the weakest (**Supp. Fig. 2D**). Taken together these data suggest that the -2 and primarily the -1 amino acid identity are key in determining the efficiency of EJC-independent NMD.

Classifier identifies tRNA-peptidyl hydrolysis as the most important feature for EJC-independent NMD substrate activity. To discover other sequence-dependent factors that could influence NMD activity, we expanded the set of biological variables and constructed a random forest classifier trained to discriminate between the high and low NMD activity (**Fig. 4A**). We selected the 3000 sequences with the highest (lowest LFC) and lowest (highest LFC) NMD activity for the EJC-independent and EJC-enhanced reporters.

We introduced additional variables such as the identities of the codons as well as their properties such as usage frequency (GenScript), tRNA binding strength, and optimality (37,38); the identities of the amino acids as well as their physical properties such as charge, solubility, dissociation constants, and hydropathy (ThermoFisher Scientific); the GC and pyrimidine

content of each of the sequences; the prokaryotic peptide release rates based on the identity of the -1 amino acid (18).

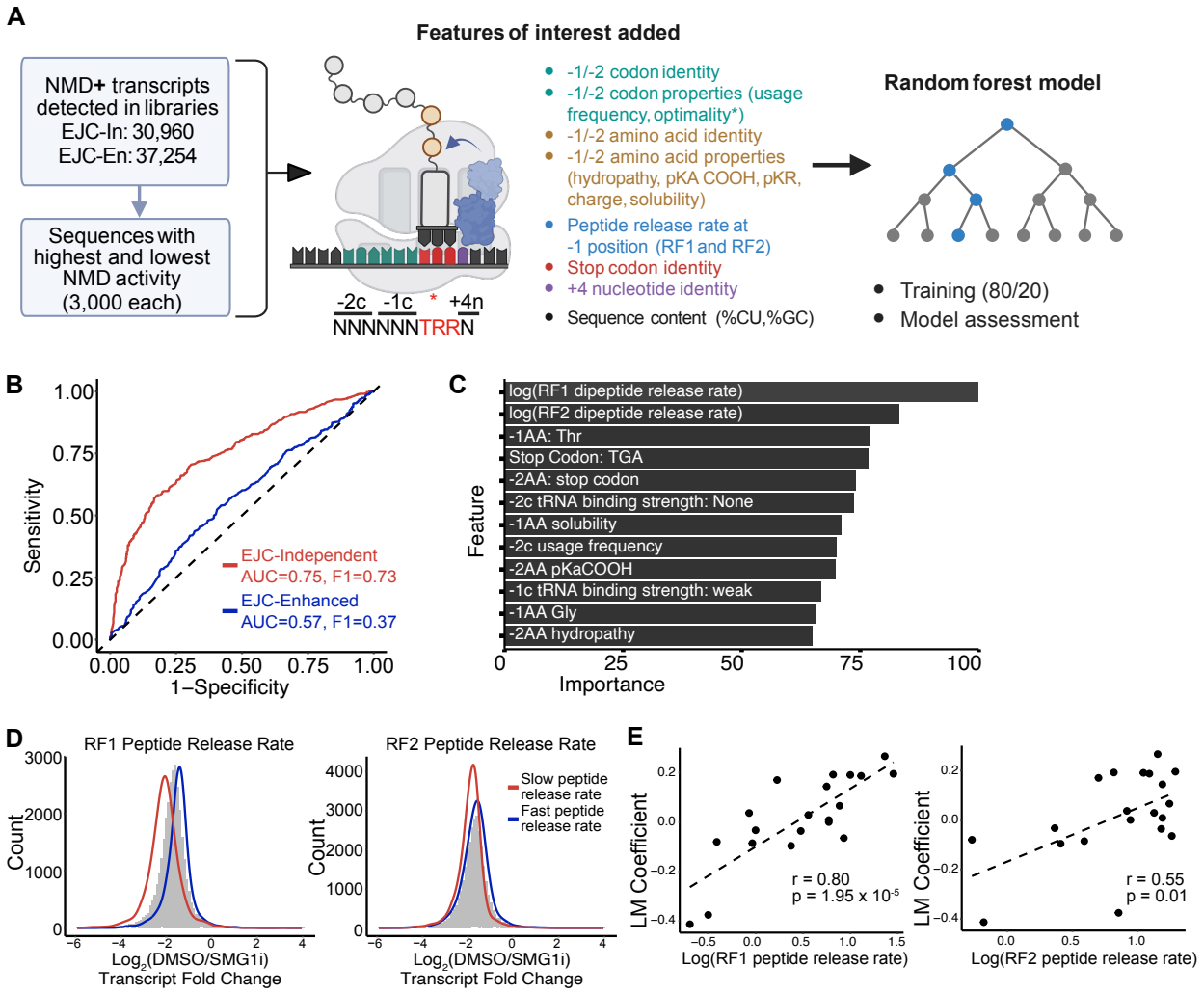


Figure 4. Peptide release rates are (A) Data preparation and variables input into the Random Forest model. (B) ROC curves, AUC values, and F1 scores for the EJC-independent and EJC-enhanced classifiers. (C) Ranked feature importance for the EJC-independent model. (D) Density curves for sequences with slow and fast peptide release rates overlaid on the NMD activity distribution for the EJC-independent NMD+ sequences. Prokaryotic release rates used (18), left: RF1 release rates, right: RF2 release rates. For RF1 slow release rates are defined as log (RF1 release rate) ≤ -0.5, fast release rates is defined as log (RF1 release rate) > 1. For RF2 slow release rates are defined as log (RF2 release rate) ≤ 0, fast release rates is defined as log (RF2 release rate) > 1.25. (E) Scatterplot of the beta coefficients for each -1 amino acid

obtained for the EJC-independent reporters against the RF1 (left) and RF2 (right) peptide release rates.

We trained two separate classifiers for the EJC-independent and EJC-enhanced reporters using the same variables. Classification performance for EJC-independent NMD activity was much better than for EJC-independent NMD activity (**Fig. 4B**), as indicated by the substantially higher area under receiver operating characteristic curve and F1 scores for the EJC-independent (0.75 and 0.73, respectively) compared to the EJC-enhanced (0.57 and 0.37, respectively) classifiers. We concluded that only the classifier for EJC-independent NMD activity was robust and suitable for further interpretation (refer to Tables 1-2 for the confusion matrices and other model metrics). The poor performance of the EJC-enhanced classifier suggests that there are hitherto unknown variables missing in the classifier that may better predict these reporters' NMD activity.

We focused on the features that were most important for the EJC-independent model. Prokaryotic release factor 1 (RF1) and release factor 2 (RF2) peptide release rates emerged as the most important predictor of NMD activity (**Fig 4C**). To examine the effect of peptide release rates on NMD activity, we compared the subpopulation of sequences with the fastest and slowest dipeptide release rates versus all NMD targets (**Fig. 4D**). The two populations were well-separated for RF1 peptide release rates and less robustly for the RF2 peptide release rates (**Fig. 4D**). To examine this relationship more quantitatively we compared the beta coefficients of the effect on NMD activity of the -1 amino acid identity to the peptide release rates of RF1 and RF2 (**Fig 4E**). We observed a strong correlation for the RF1 peptide release rates ($r = 0.8$; $p = 1.95 \times 10^{-5}$) and a weaker but significant correlation for the RF2 peptide release rates ($r = 0.55$; $p = 0.01$) (**Fig 4E**). Taken together these data suggest that peptide release rate may be a major effector of the -1 amino acid-dependent effect on NMD efficiency and present a model where the rate of peptide release could influence the window of opportunity that NMD factors have to detect, bind, and degrade an NMD target. The slower the peptide release rate, the wider the window of opportunity, the more efficient the NMD, and vice versa. Notably, Gly has the slowest peptide release rate and Tyr has the fastest (18), which might explain why Gly causes the most efficient NMD, and Tyr the least efficient.

Luminescence-based NMD reporters recapitulate -1 amino acid-dependent variation in NMD activity in EJC-independent and EJC-enhanced contexts. Results thus far suggest that peptide release rate is a major influencer of NMD, but only in the context of EJC-independent NMD and not EJC-enhanced NMD. This discrepancy could stem from two causes: the highly efficient EJC-enhanced NMD is insensitive to other modifying factors; or that it is a transcript-specific effect limited to the fluorescent reporters. To distinguish between these possibilities, we switched to an independent, luciferase-based reporter system (31). This system is identical to our fluorescent reporter system except that the transfection control is Renilla luciferase and the NMD readout is provided by Firefly luciferase (**Fig. 5A**).

To test whether the -1 amino acid effect was also present within this NMD reporter system we varied the amino acid in the -1 position of either the PTC or the NTC, choosing Gly, Val, Ala, and Tyr, which have a range of peptide release rates from slowest to fastest (**Fig. 5A**). As expected, luciferase activity was ~10-fold lower for the EJC-enhanced reporters compared to EJC-independent (**Fig. 5B**). Remarkably, the relative trend in -1 amino acid-dependent NMD variability was the same for both the EJC-independent and EJC-enhanced reporters. The NMD activity varied dependent on the -1 amino acid, with Gly having ~22% or ~29% more efficient NMD compared to Tyr for the EJC-independent and EJC-enhanced reporters, respectively (**Fig 5B**). Compared to the EJC-independent and EJC-enhanced reporters, the NMD**–** reporters did not vary with the change in the -1 amino acid (**Fig 5B**). Taken together, these data demonstrate that the -1 amino acid effect can extend to both forms of NMD.

Key NMD and translation factors do not influence the -1 amino acid dependent variation in NMD activity. Next, we wanted to ask whether any trans factors played a role in the -1 amino acid-dependent effect on NMD efficiency. To this end, we utilized shRNAs to knock down various NMD- and translation-related factors and test the effect of the knockdown on reporter levels. We decided to use the EJC-independent reporters for the knockdown experiments.

First, we investigated the UPF proteins (UPF1, UPF2, and UPF3B), which are key effectors of NMD (39,40); SMG1, which phosphorylates UPF1 to trigger NMD; and SMG6, which is responsible for cleaving the transcript when NMD is initiated (40,41) (**Fig. 5C**). Reporter protein levels increased upon knockdown of all NMD factors except for UPF3B, likely because UPF3B is known to be redundant to UPF3A (42). However, the relative difference in expression of the reporters based on -1 amino acid identity largely remained the same (**Fig. 5C**; knockdown validation **Supp. Fig. 3A**).

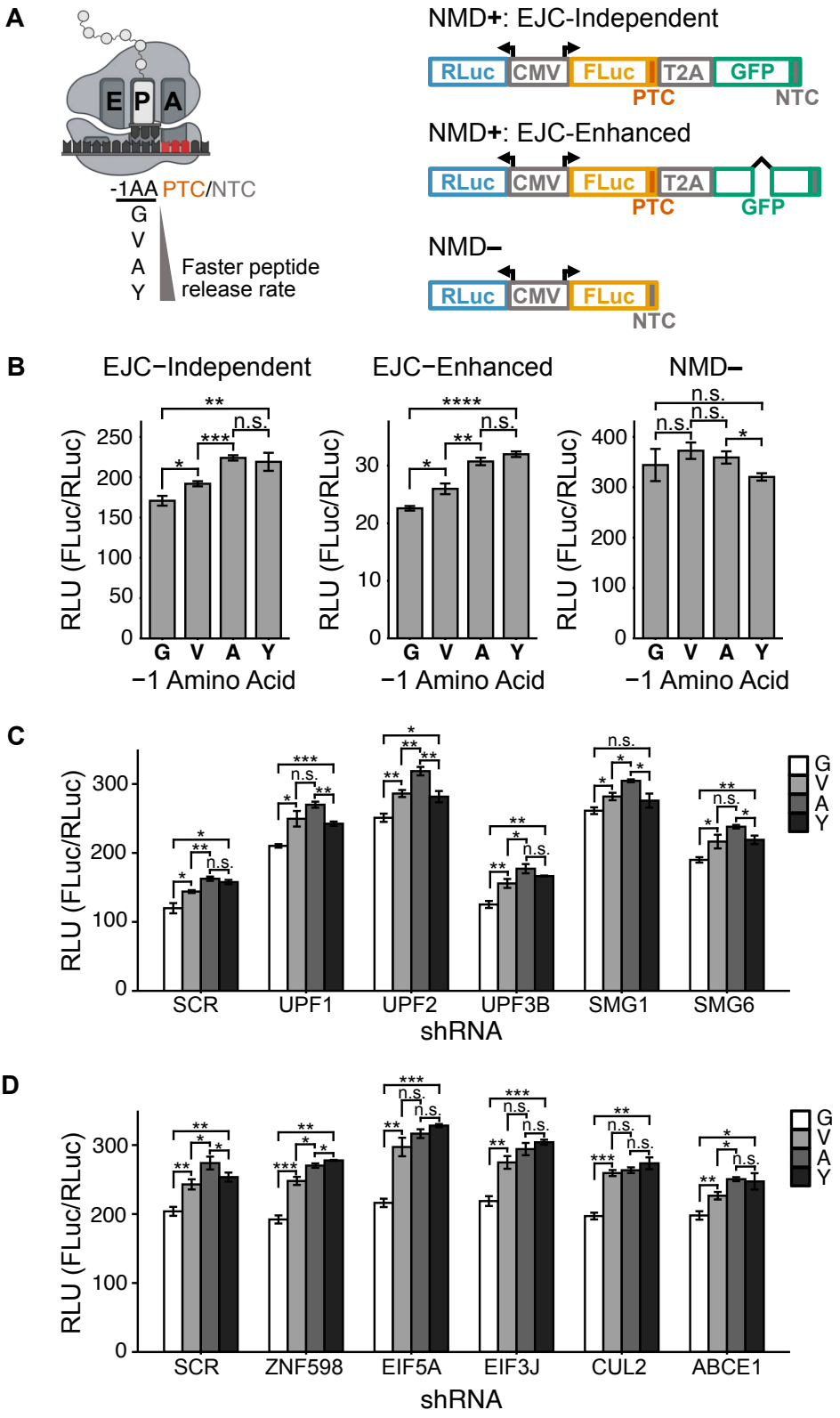


Figure 5. The -1 amino acid effect on NMD is observed for both EJC-independent and EJC-enhanced contexts and is independent of trans factors. (A) Schematic of fluorescence-based NMD reporters used (right) as well as the amino acids tested in the -1 position to the PTC/NTC (left). (B) Luciferase activity for each of the reporters in EJC-independent, EJC-enhanced, and NMD– contexts. (C) shRNA knockdown of key NMD factors and their effect on the EJC-independent reporters. (D) shRNA knockdown of key translation-related factors and their effect on the EJC-independent reporters.

Next, we investigated key translation and protein quality control related factors, ZNF598, EIF5A, EIF3J, CUL2, and ABCE1, using the same approach (**Fig. 5D**). ZNF598 recognizes the interface of collided disomes, thus playing a role in ribosome-associated quality control substrate recognition (43). It is possible that some sequences would promote ribosome collision at the end of a transcript, which would then influence NMD. EIF5A is an elongation factor that aids the ribosome in translating Gly/Pro sequences. EIF5A accumulates at the ribosome during translation termination and is hypothesized to bind at the E-site when the deacylated tRNA leaves (44-46). EIF5A binding to the termination complex may be influenced by the transcript sequence within the ribosome. EIF3J facilitates binding of the release factors to the ribosome, and thus stimulates peptide release (47). CUL2 forms a key part of the protein complex that recognizes C-terminal Gly degrons (28,29). Finally, ABCE1 is a ribosome recycling factor that acts during normal and premature termination (**Fig. 5D**) (48-50). Similar to the NMD factors, none of these conditions eliminated the differences in reporter level expression based on -1 amino acid identity (**Fig. 5D**; knockdown validation **Supp. Fig. 3B**). Together, these data point to peptide release rate as the main driver for the observed variability in NMD activity, and does not implicate any of the trans factors tested.

Eukaryotic translation termination efficiency varies with -1 amino acid identity *in vitro*. The peptide release rate values used in the Random Forest classifier (Fig. 4) came from a prokaryotic system and may not reflect the peptidyl tRNA hydrolysis rate in mammalian cells (18). To determine if the rate of peptide release varied based on the identity of the -1 amino acid in a eukaryotic system, we used the *in vitro* Termi-Luc assay (**Fig. 6A**) in the excess of mutated eukaryotic release factor 1 (eRF1(AGQ)). This protein is able to recognize stop codons, but unable to induce peptidyl-tRNA hydrolysis (51). eRF1(AGQ) binding allows to freeze translation at the termination stage and obtain pretermination ribosomal complexes (preTCs). PreTCs are

then purified from eRF1(AGQ) by sucrose gradient centrifugation, followed by the addition of release factors (working concentrations determined in **Supp. Fig. 4**). The real-time release kinetics of the synthesized NLuc are then measured to determine the translation termination rate of the transcript. We used this system to test the termination rates when Gly, Val, Ala, or Tyr codons are present at the -1 position to the stop codon. If peptide release rate influences the time of residence of the terminating ribosome, we would expect to see a difference in termination rate when different amino acids are present upstream of the stop codon.

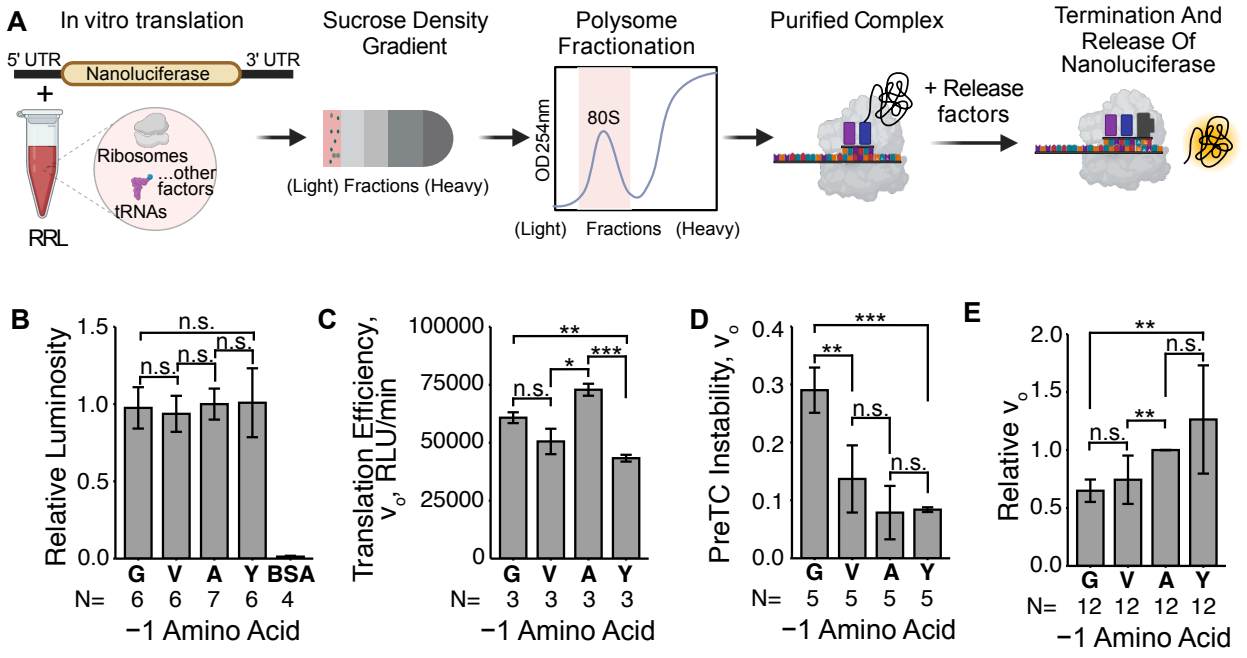


Figure 6. Peptide release rate in an *in vitro* eukaryotic system is slowest when glycine is upstream of the stop codon. (A) Schematic of the Termi-Luc peptide release assay. This assay was performed for nanoluciferase reporters with different amino acids -1 to the stop codon. (B) Tested *in vitro* luminescence activity of recombinant NLuc variants. (C) Translation efficiency of each reporter in the *in vitro* system. (D) Relative rate of peptide release when no release factors are added to the termination reaction i.e., preTC instability. (E) Relative peptide release rates for each reporter.

First, control experiments were performed to ensure that substitution of an amino acid residue did not affect the activity, translation efficiency, or preTC stability of NLuc. To determine whether NLuc activity was impacted by the -1 amino acid, we purified all tested peptide variants as recombinant proteins and determined their maximum luminescence. All NLuc peptide variants

had the same activity (**Fig. 6B**). To determine whether translation efficiency was impacted by changing the -1 amino acid, we calculated the maximal derivative (slope) of the linear portion of the luminescence production curve. The -1 amino acid identity affected translation efficiency with Tyr having the lowest and Ala having the highest efficiency (**Fig. 6C**). To determine preTC instability, the rate of peptide release of each peptide variant was calculated without the addition of any release factors. These data showed that Gly had the most unstable preTC, while Tyr had the most stable preTC (**Fig. 6D**). Finally, we performed the Termi-Luc assay on each peptide variant starting with equal amounts of preTC complexes. We found that Gly had the slowest release rate, and Tyr had the fastest release rate, a difference which was statistically significant and correlated with high and low NMD efficiency, respectively (**Fig. 6E**). Further, the pattern of peptide release rate was opposite to the pattern of preTC instability and did not correlate with translation efficiency (**Fig. 6C-E**). Taken together, these data show that peptide release rates indeed vary based on -1 amino acid identity in a eukaryotic system, with glycine showing the slowest termination rate and tyrosine the fastest, among the amino acids tested.

Gly-PTC variants are tolerant to loss of function. If slow termination kinetics enhances NMD, we would expect Gly-PTC enrichment to be more pronounced in genes that are not essential, relative to essential genes. To test this hypothesis, we analyzed the relative frequency of PTVs in healthy individuals binned in terms of the LoF intolerance score of their host genes (**Fig. 7A-B**). The lower the LoF intolerance score, the more tolerant to loss of function a gene is and the less essential We find that PTVs with Gly-PTC contexts increase in enrichment with decreasing LoF intolerance score, suggesting that these PTVs are more tolerant to loss of function (**Fig. 7B**). Further, if slow termination is a mechanism for LoF tolerance in healthy individuals, we would not expect Gly-PTC contexts to be enriched in disease-associated PTVs. To test this, we examined the amino acids enriched immediately upstream of a PTC in PTVs from the ClinVar database, which contains human genetic variants associated with disease (52). As expected, Gly showed no enrichment -1 to a PTC in disease associated PTVs (**Fig. 7C**). Taken together, our data supports a model where slower peptide release rate offers a longer kinetic window for NMD factors to assemble and degrade a transcript, thus causing highly efficient NMD. Thus, Gly-PTC context may have coevolved with NMD to efficiently eliminate transcripts encoding truncated proteins

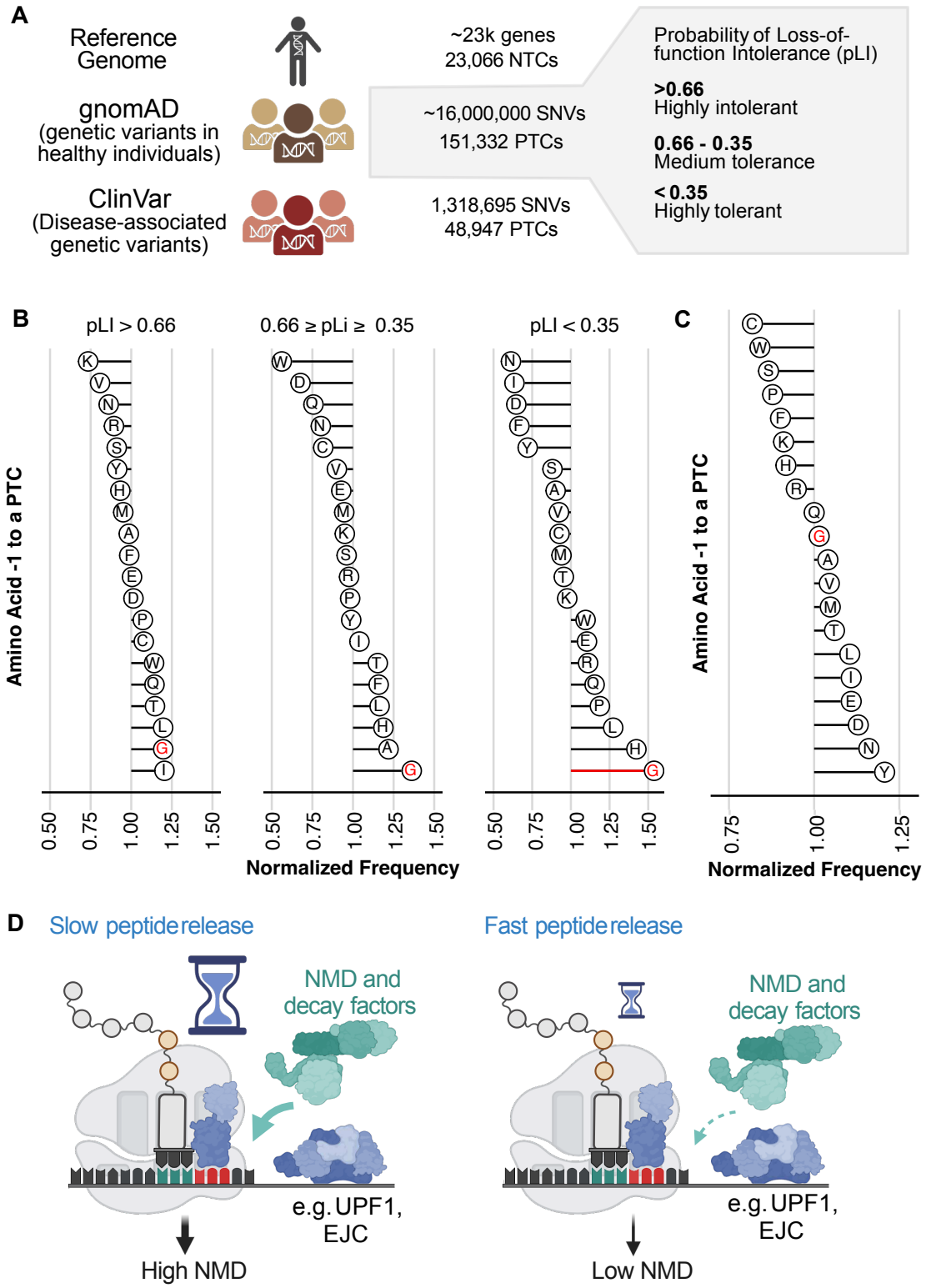


Figure 7. PTVs with Gly -1 to a PTC are tolerant to LoF and not enriched in genetic diseases. (A) Schematic for the analysis of LoF scores for rare PTVs in healthy individuals as well as PTVs in disease contexts. (B) The -1 amino acid enrichment of PTVs binned into 3

422 categories of LoF scores. The lower the score the more tolerant the PTVs are to loss of
423 function. (C) The -1 amino acid enrichment of PTVs in disease contexts. (D) Window of
424 opportunity model for how NMD efficiency is influenced by peptide release rates. A slower
425 peptide release rate gives NMD factors a larger window to bind an act on a target, and vice
426 versa.

DISCUSSION

NMD is often considered a binary outcome that relies on the presence of triggering factors downstream of the PTC. It is increasingly clear that NMD is tunable, with multiple factors influencing the degree to which a target undergoes NMD. However, the factors influencing this variability are incompletely understood. While translation termination is pre-requisite for NMD, it is not thought to quantitatively influence NMD efficiency. Our data support a window of opportunity model where the time of residence of the terminating ribosome, which serves as the platform for the assembly of NMD machinery, determines the effectiveness of NMD (**Fig. 7D**). Consistent with this model, slower peptidyl-tRNA hydrolysis rate enhances NMD efficiency, while faster rate leads to less efficient NMD. Together with the RBP cues that trigger NMD, termination kinetics could be a key modifier of the probability with which a prematurely terminating ribosome leads to NMD.

Several lines of evidence point to the significance of the Gly-PTC context. Gly codon is most enriched before a PTC and least enriched before an NTC, causes the most efficient NMD, and has the slowest associated peptide release rate. Further, PTVs with Gly-PTC contexts are enriched among variants that are tolerant to loss of function in healthy individuals, while they are not enriched among disease-causing variants. Gly is also a C-terminal degron, which suggests that Gly-end proteins, truncated or otherwise, are more efficiently cleared (28,29). Thus, the preponderance of glycine codons at the ends of truncated proteins may be the result of co-evolution of PTVs alongside processes that efficiently eliminate Gly-end proteins at both the RNA and protein levels.

Other amino acids that have similar patterns to Gly are also consistent with our model. For example, Thr and Pro are also enriched before a PTC in healthy individuals and depleted before an NTC. They both cause efficient NMD (less efficient than Gly but more efficient than most other amino acids) and have slow peptide release rates. On the other end of the spectrum, Tyr which had the lowest NMD efficiency in our experiments, was enriched before an NTC but not a PTC. It also has the fastest peptide release rate. PTVs with Tyr at the -1 of a PTC are depleted among variants that are tolerant to loss of function and are enriched in disease contexts.

An interesting observation from the linear modeling of the MPRA data is that the stop codon TGA caused 20% more efficient EJC-independent NMD compared to TAG. This stop codon also emerged as an important predictor of EJC-independent NMD in the Random Forest classifier. This observation is counterintuitive since TGA is known to be the stop codon with the lowest fidelity i.e. it is most likely to be readthrough, which usually stabilizes an NMD target (53).

Our results suggest there are kinetic differences in termination at different stop codons. Additional factors, such as RBPs or other sequence elements, may influence whether the stop is readthrough or whether it causes efficient NMD.

EJC-independent NMD is weaker than EJC-enhanced NMD (31,34-36) and have different cues. For EJC-independent NMD, a long 3′ UTR downstream of the PTC remains bound by RBPs like UPF1, which would ordinarily be dislodged by the ribosome during translation. Some models propose that the length of the 3′ UTR and the number of RBPs bound to it influences the extent of NMD (6,7,54). For EJC-enhanced NMD, the presence of one EJC downstream of the prematurely terminating ribosome is sufficient to trigger efficient NMD (3,55). Thus, the window of opportunity during termination may matter more for a weaker, EJC-independent target, than a stronger EJC-enhanced target. Indeed, linear modeling and the Random Forest classifier of the MPRA data suggested that PTC-sequence context mattered more for EJC-independent NMD compared to EJC-enhanced NMD. In contrast, individual luciferase reporters demonstrated that the -1 amino acid effect applied to both EJC-independent and EJC-enhanced forms of NMD. We interpret these data to mean that the PTC context influences both modes of NMD.

In summary, the effect of various amino acids upstream of the PTC on NMD efficiency and their pattern of enrichment in healthy and diseased populations suggest that NMD has played a key role in shaping the evolution of protein-truncating variants. These findings can be used to better model the phenotypic outcomes of PTVs and thus, develop more effective strategies to counteract PTV-associated diseases.

ACKNOWLEDGEMENTS

We thank Matthew Taliaferro for the gift of HEK293T cells with the loxP cassette. We thank Timothy Stasevich, Tatsuya Morisaki, and Laura White for exploratory studies on glycine's impact on NMD. We thank Srinivas Ramachandran, Olivia Rissland, and all members of the Jagannathan laboratory for insightful manuscript feedback. This work was supported by the RNA Bioscience Initiative (R. F.), AHA Award #831183 (D.K.), 5T32GM136444-03 (M.L.), Polish National Agency for Academic Exchange Bekker Program PPN/BEK/2019/1/00173 (M.P.S.), R35 GM119550 (J.H.), R35GM147025 (N.M.), University of Colorado School of Medicine Translational Research Scholars Program (S.J.), and the National Institutes of Health grant R35GM133433 (S.J.), The study of the effect of the stop codon 5' context on the efficiency of translation termination was supported by the Russian Science Foundation grant 22-14-00279 (E.A.).

AUTHOR CONTRIBUTIONS

Conceptualization (D.K., R.F., S.J.); Data curation (D.K., R.F., A.E.C., S.J.); Formal analysis (D.K., R.F., S.J.); Funding acquisition (D.K., E.A., S.J.); Investigation (D.K., R.F., N.B., A.S., M.L., A.E.C., S.J.); Project administration (S.J.); Software (D.K., R.F., M.P.S., M.A.C., N.M., S.J.); Supervision (J.H., N.M., E.A., S.J.); Validation (D.K., A.E.C.); Visualization (D.K., R.F., A.E.C., S.J.); Writing – original draft (D.K., S.J.); Writing – review and editing (D.K., R.F., A.E.C., M.A.C, M.P.S., J.H., N.M., E.A., J S.J.).

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sujatha Jagannathan (sujatha.jagannathan@cuanschutz.edu).

Materials availability

All unique reagents generated in this study are available from the lead contact with a completed Material Transfer Agreement.

Data and code availability

- All sequencing data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the Key Resources Table.
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the Key Resources Table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and culture conditions

HEK293T cells (female) were obtained from ATCC (CRL-3216; RRID:CVCL_0063). HEK293T Lox2272/LoxP cells were obtained from Taliaferro laboratory. All cell lines were determined to be free of mycoplasma by PCR screening. HEK293T and HEK293T cells with Lox2272/LoxP cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific) supplemented with 10% EqualFETAL (Atlas Biologicals).

METHOD DETAILS

gnomAD and ClinVar analysis

Rare (frequency < 0.01) SNP variants from gnomAD v2.1.1 exome database that pass random forest classification quality filtering were analyzed, filtering for “Stop-gained” consequence via bcftools. Using “cDNA_position” field pre-annotated by VEP, the sequence and codons surrounding variants from Gencode v19 transcript FASTA file was then extracted using bedtools getfasta. For other contexts such as normal stop codons, custom code (found in: https://github.com/jagannathan-lab/2023-kolakada_et_al) was used to extract sequence and codon information using Biostrings and BSgenome in R. Codon enrichment is calculated as observed frequency of a particular codon/amino acid normalized against normal occurrence frequency of the codon/amino acid in the last 10 positions of all normal coding genes, similar to Koren et al (28). Gene-level loss of function scores were retrieved from gnomAD LoF constraint scores. Clinvar data from 2022/07/30 was processed similarly to gnomAD, but with sequence extraction done in R.

Cloning

The NMD reporter plasmids have been previously described (31). For individual fluorescent and luminescent reporters containing Gly, Val, Ala, and Tyr amino acids before the PTC, oligos containing these sequences (Integrated DNA Technologies) and matching the EcoRI (NEB, R3101S) and XhoI (NEB, R0146S) restriction sites were synthesized. These oligos were annealed and ligated using the Quick Ligation Kit (NEB, M2200L) into EJC-independent and EJC-enhanced fluorescent and luminescent backbones digested with EcoRI and XhoI. To make the NMD– reporters for each sequence, each EJC-independent reporter was digested with EcoRI and MfeI (NEB, R3589S), the sticky ends were filled in using Klenow Fragment (NEB, M0212S) and the blunt ends were ligated together using the Quick Ligation Kit (NEB), removing the GFP 3′ UTR. The oligos synthesized for each sequence tested are as follows.

Oligo	Sequence (5’ – 3’)
G1_Top	AATTCAGAACCACCAGAACCACCTTAGCCAGAACCACCAGAACCACCC
G2_Top	AATTCAGAACCACCAGAACCACCTTAACCAGAACCACCAGAACCACCC
A_Top	AATTCAGAACCACCAGAACCACCTTAGGCAGAACCACCAGAACCACCC
V_Top	AATTCAGAACCACCAGAACCACCTTACACAGAACCACCAGAACCACCC
Y_Top	AATTCAGAACCACCAGAACCACCTTAGTAAGAACCACCAGAACCACCC
G1_Bottom	TCGAGGGTGGTTCTGGTGGTTCTGGCTAAGGTGGTTCTGGTGGTTCTG
G2_Bottom	TCGAGGGTGGTTCTGGTGGTTCTGGTTAAGGTGGTTCTGGTGGTTCTG

A_Bottom	TCGAGGGTGGTTCTGGTGGTTCTGCCTAAGGTGGTTCTGGTGGTTCTG
V_Bottom	TCGAGGGTGGTTCTGGTGGTTCTGTGTAAGGTGGTTCTGGTGGTTCTG
Y_Bottom	TCGAGGGTGGTTCTGGTGGTTCTTACTAAGGTGGTTCTGGTGGTTCTG

MPRA plasmid library

An oligo library (Eurofins) of sequence CTA GCA AAC TGG GGC ACA GCC TCG AGG GTG GTT CTG GTG GTN NNN NNT RRN GTG GTT CTG GTG GTT CTG AAT TCG ACT ACA AGG ACC ACG ACG G was synthesized, where N refers to equal proportions of the A, C, G, and T nucleotides, and R refers to equal portions of the A and G nucleotides. This library was resuspended in water to 100 μM. The oligo libraries were then filled in (1 amplification cycle) using the Q5 High-Fidelity DNA Polymerase (NEB, M0492S) according to manufacturer's instructions, using 100 μmols of oligos, and 100 μmols of reverse primer (sequence: CACCGTCGTGGTCCTTGTAGTC), in 2x PCR reactions of 25 μL each. After amplification, the PCR reaction was digested with Exonuclease I, at 37°C for 3 h to digest any remaining single stranded DNA. The DNA was purified using AMPure XP beads (Beckman, A63882), as per manufacturer's instructions.

The EJC-independent and EJC-enhanced fluorescent reporter backbones were linearized using EcoRI (NEB) and XhoI (NEB) at room temperature overnight. Digested plasmid DNA was gel purified using the NucleoSpin Gel and PCR Clean-up kit (Takara Bio, 740609.250). The oligo library was then cloned into the digested backbones using the Gibson Assembly Master Mix (NEB, E2611L) using an insert:vector molar ratio of 7:1 for the EJC-independent backbone and 3:1 for the EJC-enhanced backbone. The reactions were incubated at 50°C for 1 h. The reactions were then purified using AMPure XP beads to remove excess salts and a total of 200 ng of plasmid DNA was transformed into MegaX DH10B T1R Electrocompetent Cells (ThermoFisher, C640003), using a Biorad GenePulser electroporator. The transformed cells were recovered in the recovery medium provided with the electrocompetent cells at 37°C for 1 h, then plated on 33 pre-warmed 15 cm Luria broth (LB) agar-Carbenicillin (RPI, C46000-5.0) plates and incubated at 37°C overnight. The next day colonies were collected by scraping plates, spun-down, and then midi-prepped (NucleoBond Xtra Midi EF kit, 740420.50) to extract the plasmid DNA.

Plasmid transfections

Transfections for fluorescent and luminescent reporters were performed differently since fluorescent proteins have longer maturation rates than luciferase proteins (56). For the fluorescent reporters, cells were split 24 h after transfection equally into two new 6-wells. Forty-eight hours after transfection, one set of cells was harvested for RNA and the other set for flow cytometry. For the luminescent reporters, transfections were performed in 24-well plates at a 40% confluency with 500 ng of plasmid DNA using Lipofectamine 2000 (ThermoFisher Scientific, 11668027), as per manufacturer's instructions. All cells were harvested for luciferase assays.

Fluorescence imaging

Fluorescent reporters were imaged before harvest for RFP, GFP, and BFP using the Cytation Multimode Imaging plate reader (Agilent, CYT5MFW-SN). Images were processed and quantified using the Gen5 software (Agilent).

Generating NMD+ reporter cell lines for the MPRA

To create the EJC-independent and EJC-enhanced reporter cell lines HEK293T cells containing a Lox2272 and LoxP enclosing a blasticidin cassette (courtesy Taliaferro Lab, CU Anschutz), were co-transfected with 97.5% NMD reporter plasmid library and 2.5% Cre-plasmid (Addgene plasmid #27493). These transfections were performed in 60-70% confluent 10 cm plates using 11,800 ng of plasmid library and 295 ng of Cre plasmid per plate for 20 plates. Twenty-four hours after transfection, each plate of cells was split equally into two new plates and puromycin (2 μg/mL) was added to the media to begin selection. Puromycin selection lasted between 1-2 weeks, during which the media was replaced with fresh media supplemented with puromycin every 1-2 days and cells were consolidated into fewer plates when their confluency got too low. Once selection was complete, cells were frozen down in medium containing 10% DMSO. These cells were thawed at a later time for the MPRA.

Targeted sequencing of MPRA libraries

To obtain MPRA sequencing libraries, frozen EJC-independent and EJC-enhanced reporter cell lines were thawed. These cells were split into 6, 15 cm plates, 3 of which were treated with DMSO and the other 3 treated with 0.5 μM SMG1i (technical replicates). These plates were harvested for RNA in TRIzol Reagent (ThermoFisher Scientific, 15596018) after 24 hrs. TRIzol extractions were performed for RNA as per manufacturer's instructions. A DNase digest was performed on 10 μg of RNA from each sample using the Turbo DNA-free kit (ThermoFisher Scientific, AM1907), including reaction inactivation. Following this, 2 μg of

digested RNA was used to make cDNA using Superscript II Reverse Transcriptase in duplicate 20 µL reactions (ThermoFisher Scientific, 18064014).

For library preparation, each cDNA sample was split into 20 PCR reactions (2 µL cDNA/PCR) and amplified using reporter-specific forward and reverse primers with UMIs (forward primer: ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAGACTTCCTCTGCCCTC; reverse primer: AGACGTGTGCTCTTCCGATCTNNNNNNNNtggggcacagcctcga, where the Ns refer to the nucleotides for the UMIs). PCR reactions were performed using Kapa HotStart PCR Kit (Roche, KAPA KK2502) using an annealing temperature of 69°C and 32x cycles. The 20 PCR reactions were pooled and 100 µL of all reactions were purified using AMPure XP beads. A second PCR reaction was performed on 25 ng of the purified product with primers containing Illumina adaptors. Once again, the Kapa HotStart PCR Kit was used, this time with an annealing temperature of 65°C and 8x cycles. Reactions were purified using AMPure XP beads and samples were run on an agarose gel to verify the library size. Libraries were sequenced using the Illumina sequencing platform NovaSeq 6000 (paired-end 2x150 cycles). A total of 50 million and 30 million reads were requested for the EJC-independent and EJC-enhanced libraries, respectively.

MPRA data analysis

To process the MPRA sequencing libraries, the UMI-tools repository was used to add the UMIs to the name of each read in the FASTA files (57). Following this, Cutadapt was used to trim the 5′ and 3′ends of the reads in a sequence dependent manner, up to the 10 nt context of interest (58); ~25% of reads were lost in this process. SeqPrep was used to merge sequencing reads 1 and 2, only accepting reads that perfectly align (https://github.com/jstjohn/SeqPrep); ~5% of reads were lost in this process. Bowtie was used to map the reads to a reference file of all sequence contexts (59); 1% of reads were lost in this process. Samtools was used to create a BAM file of mapped reads, followed by deduplicating the BAM files based on UMIs and converting the BAM files to BED files (60). Around 10% of reads were lost upon deduplication of UMIs. A custom python script was used to remove reads on the negative strand and to get a bed file of reads in the correct orientation.

The NMD activity of an individual reporter transcript was the log2-fold change in between DMSO-treated cells to SMG1i-treated cells calculated with DESeq2 (33). The reporter transcripts that do not undergo NMD, i.e. Trp codon rather than a stop codon at the PTC position, were used similar to housekeeping normalization controls within DESeq2. Both linear

models and random forest classifiers were performed in R. The code and packages used for these analyses can be found in this repository: https://github.com/jagannathan-lab/2023-kolakada_et_al.

Luciferase assays

For cells harvested from 24-well transfections, the media was aspirated, 400 µL of 1x PLB from the Dual Luciferase Reporter Assay System (Promega, E1980) was added to the cellular monolayer per well, and the plate was frozen for subsequent luciferase assays. The frozen cells were brought to room temperature, transferred to tubes, and lysed thoroughly via vortexing for 30 s per sample. 20µL of this dilution was pipetted into a 96-well plate, in triplicate, per sample. The LAR II and Stop & Glo reagents were prepared as per the manufacturer's instructions, using 50 µL per sample, each. Luciferase assay was performed using the Glomax Navigator (Promega, GM2010) as per the Dual Luciferase Assay Protocol.

Lentiviral transductions

Knockdowns were conducted using shRNAs. Plasmids containing shRNAs were first packaged into lentivirus. To do this, HEK293T cells were transfected with 4 µg of the relevant shRNA vector (Functional Genomics Core, CU Anschutz), 3 µg of psPAX2 (Addgene plasmid #12260), and 1 µg of pMD2.G (Addgene plasmid #12259), in 80% confluent 6 cm plates. The shRNAs used were for scrambled (SCH016), UPF1 (TRCN0000022254), UPF2 (TRCN0000151381), UPF3B (TRCN0000152769), SMG1 (TRCN0000194827**)**, SMG6 (TRCN0000040014), ZNF598 (TRCN0000073159), EIF5A (TRCN0000062552), EIF3J (TRCN0000062013), CUL2 (TRCN0000006523), and ABCE1 (TRCN0000158417). Twenty-four hours after transfection the media was replaced with 3 mL of fresh media. Forty-eight hours after transfection media containing lentivirus was harvested and spun down at 1200 rpm for 5 minutes. Aliquots of the supernatant were frozen at -80°C for later use.

For the knockdown experiments with the luciferase reporters, cells were seeded at a 40% confluency in 6-wells and infected with 50 µL of virus. Twenty-four hours post transduction, the cells were split into 24-wells at a 20% confluency. Forty-eight hours post transduction, the luciferase reporters were transfected into cells as described above. Seventy-two hours post-transduction the cells were harvested for luciferase assays. For knockdown validation experiments, the same procedure was followed without the transfection of luciferase reporters.

RNA extraction and RT-qPCR

For RT-qPCR used to validate knockdowns of SMG1, ZNF598, and EIF3J, RNA was extracted from cells using TRIzol Reagent (ThermoFisher Scientific). DNase digestion was performed with DNaseI (ThermoFisher Scientific, 18068-015) on 1 µg of extracted RNA. This was followed by cDNA synthesis using the SuperScript III First-Strand Synthesis kit, using random hexamers. A no-RT sample was included as a control to make sure there was no genomic DNA contamination. The cDNA was then diluted 1:4 and 2 µL were used per 10 µL qPCR reaction. qPCR was performed using iTaq Universal SYBR Green Supermix and primers specific for SMG1, ZNF598, EIF3J, and the housekeeping gene RPL27, were used at a final concentration of 0.25 µM. The following primers were used for each gene:

Oligo	**Sequence (5' – 3')**
RPL27_F	GCAAGAAGAAGATCGCCAAG
RPL27_R	TCCAAGGGGATATCCACAGA
SMG1_F	TGGGAAAGACCACCACTGCACA
SMG1_R	TGCATGTGTTGACTGGCCTGCT
ZNF598_F	TCGTTGGTGGCGAAGACTAC
ZNF598_R	TCGGTCCTCTTCTTCCCTTT
EIF3J_F	GCAGATAAACTGCGGCTAAAGA
EIF3J_R	TCTCTTGAAGATGGGTTCATAGCA

Reactions were set up in triplicate per sample and plated in 384-well plates. The plates were then run on the OPUS Bio-Rad qPCR machine (Bio-Rad, 12011319) using the 2-step Amplification and melting curve protocol. For the knockdown validations, the Livak method (61) was used to quantify differences in RNA levels: SMG1, ZNF598, EIF3J, were normalized to RPL27. Mean RNA levels of 3 different transfections or 3 technical replicates of transductions were plotted, respectively. Error was calculated using standard error of the mean (fluorescent/luminescent reporters) or standard deviation (knockdowns).

Protein isolation and western blotting

Cells were harvested in RIPA buffer supplemented with complete protease inhibitor (Roche, 11836170001). Protein was run on NuPAGE Bis-Tris precast polyacrylamide gels (Thermo Fisher Scientific, NP0323BOX) alongside PageRuler Plus Prestained Protein Ladder (Thermo Fisher Scientific, 26619) and transferred to Odyssey nitrocellulose membrane (LI-COR Biosciences, 926-31092). Membranes were blocked in Intercept (PBS) Blocking Buffer (LI-COR Biosciences, 927-70001) before overnight incubation at 4°C with primary antibodies diluted in Blocking Buffer containing 0.2% Tween 20. Membranes were incubated with IRDye-conjugated

secondary antibodies (LI-COR Biosciences) for 1 h and fluorescent signal visualized using a Sapphire Biomolecular Imager (Azure Biosystems) and Sapphire Capture software (Azure Biosystems). When appropriate, membranes were stripped with Restore Western Blot Stripping Buffer (ThermoFisher Scientific) before being re-probed. Primary antibodies include anti-UPF1 (Abcam, ab109363), anti-UPF2 (Cell Signaling, 11875S), anti-UPF3B (Abcam, ab134566), anti-SMG6 (ABclonal, A10141), anti-CUL2 (Bethyl Laboratories, A302-476A), anti-ABCE1 (Abcam, ab185548), anti-EIF5A (Abcam, ab32407), and anti-GAPDH (Abcam, ab9484). Secondary antibodies used include IRDye 650 Goat anti-Mouse IgG Secondary Antibody (LI-COR Biosciences, 926-65010) and IRDye 800CW Goat anti-Rabbit IgG Secondary Antibody (LI-COR Biosciences, 926-32211).

Luminescent reporters for *in vitro* assays

A pNL-globine vector derived from pNL1.1 vector (Promega), with a β-globin 5′UTR addition before nanoluciferase (NLuc) coding sequence was used (Shuvalov et al., 2021). Additional constructs were created based on pNL-globin, encoding NLuc with various substitutions of the codon before the stop codon using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, cat. 200518-5) was used. Primers were selected in the manufacturer's recommended web-based QuikChange Primer Design Program (www.agilent.com/genomics/qcpd). For obtaining mRNA, the fragment of the plasmid were amplified using RV3L (CTAGCAAAATAGGCTGTCCCCAG) and FLA50 (TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTTGTTTATTGCAGC TTATAATGG) primers, as described in Shuvaolv et al., 2021 (Shuvalov et al., 2021). Templates were run-off transcribed with the T7 RiboMAX ™ Large Scale RNA Production System (Promega, P1320) kit according to the manufacturer's protocol. The mRNA was then purified sequentially by isolation in acidic phenol, precipitation with 3 M LiCl followed by 80% ethanol wash.

Expression and purification of eRF1 and eRF3

Recombinant eRF1 was expressed from the plasmid pET-SUMO-eRF1. To obtain pET-SUMO-eRF1, the eRF1 coding sequence was amplified from the plasmid pET23b-eRF1 (62) using primers petSUMO_eRF1_F (GAGAACAGATTGGTGGTATGGCGGACGACCCCAG) and petSUMO_eRF1_R (CCGAATAAATACCTAAGCTCTAGTAGTCATCAAGGTCAAAAAATTCATCGTCTCCTCC). eRF1 was expressed in E. coli BL21(DE3) cells. The lysate was prepared by ultrasonification of the pelleted cells in a buffer composed of 20 mM Tris-HCl pH 7.5, 500 mM KCl, 10% glycerol,

744 0.1% Triton X-100, 0.5 mM PMSF, and 1 mM DTT. Following this, His-SUMO tag was cleaved
745 by His-tagged Ulp1 protease. Untagged eRF1 was further purified by anion-exchange
746 chromatography (HiTrap Q HP, Cytiva). Fractions enriched with eRF1 were collected, dialyzed
747 in storage buffer, frizzed in liquid nitrogen, and stored at −70 ºC. Recombinant eRF3A cloned
748 into baculovirus vector EMBacY from a MultiBac expression system was expressed in the insect
749 cell line Sf21. Following this, recombinant proteins were purified using Ni-NTA agarose and ion-
750 exchange chromatography, as described previously (25).

751 **Expression, purification, and determination of luminescence or recombinant**
752 **nanoluciferase**

753 To obtain recombinant NLuc with varying amino acids in the position -1 to the stop
754 codon, corresponding constructs of pNL-globine were created using the petSUMO vector
755 (Invitrogen, cat. K30001). Each vector was assembled via Gibson Assembly (NEB, cat. E2611L)
756 using PCR products of the petSUMO backbone (forward primer:
757 AGCTTAGGTATTTATTCGGCGCAAAGTG; reverse primer:
758 ACCACCAATCTGTTCTCTGTGAGC) and pNL-globine inserts (forward primer:
759 GAGAACAGATTGGTGGTATGGTCTTCACACTCGAAGATTTCGTTGG; reverse primer:
760 CCGAATAAATACCTAAGCTTACGCCAGAATGCGTTCGCA) amplified using Q5 High-Fidelity
761 DNA Polymerase (NEB, cat. M0491L). The resulting plasmids were expressed in *E. coli*
762 BL21(DE3) and His-SUMO-Nluc was purified using Ni-affinity chromatography. Recombinant
763 6xHis-ULP1 protease was added to purify 6xHisSUMO-Nluc and incubated for 2 hours. The
764 6xHisSUMO fragment and 6xHis-ULP1 were then removed with Ni-NTA agarose to obtain
765 purified Nluc without tags. 0.1 femtomole of recombinant Nluc was incubated in storage buffer,
766 with the addition of BSA up to 500 µg/mL and 0.5 % NanoGlo (Promega). Luminescence was
767 measured at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) in a 40 min
768 time period. The luminescence of NLuc was calculated as a maximum of relative luminescence
769 units (RLU_{max}).

770 **Purification of the preTC-NLuc and Termi-Luc Assay**

771 For the Termi-Luc assay, preTCs translating NLuc (preTC-NLuc) were purified using
772 previously published methods (63,64). 100% RRL lysate was preincubated in a mixture
773 containing 1 mM $CaCl_2$ and 3 U/µL Micrococcal nuclease (Fermentas) at 30°C for 10 min,
774 followed by the addition of EGTA to a final concentration of 4 mM. The lysate was then diluted
775 to 70% (v/v) and supplemented with 20 mM HEPES-KOH (pH 7.5), 80 mM KOAc, 0.5 mM
776 $Mg(OAc)_2$, 0.3 mM ATP, 0.2 GTP, 0.04 mM of each of 20 amino acids (Promega), 0.5 mM

spermidine, 0.45 μM aminoacylated total rabbit tRNA, 10 mM creatine phosphate, 0.003 U/μL creatine kinase (Sigma), 2 mM DTT, and 0.2 U/μL Ribolock (ThermoFisher) (70% RRL mix).

For preTC-NLuc assembly, 220 μL of 70% RRL mix was preincubated in the presence of 1.7 μM ERF1 G183A mutant (65) at 30° C for 10 min, followed by the addition of 10.5 pmol of NLuc mRNA. The mixture was incubated at 30°C for 40 min. The KOAc concentration was then adjusted to 300 mM and the mixture was layered on 5 ml of a 10–35% linear sucrose gradient in a buffer containing 50 mM HEPES-KOH, pH 7.5, 7.5 mM $Mg(OAc)_2$, 300 mM KOAc, 2 mM DTT. The gradient was centrifuged in a SW55-Ti (Beckman Coulter) rotor at 55 000 rpm (367 598 g_{max}) for 1 h. The gradient was fractionated in 15 fractions of 150 μL from bottom to top and the remaining sucrose was collected separately. Fractions of the first peak, enriched with the preTC-Nluc were analyzed by the Termi-Luc assay, merged, flash-frozen in liquid nitrogen, and stored at −70°C.

A peptide release assay with the preTC-NLuc (Termi-Luc) was performed as previously described with some modifications (63). Peptide release was performed in a solution containing 1.5 pM preTC-Nluc, 45 mM HEPES-KOH pH 7.5, 1.4 mM Mg_2OAc, 56 mM KOAc pH 7.0, 1 mM DTT, 177 μM spermidine, 1.5 % (w/w) sucrose, 0.8 mM $MgCl_2$, 0.2 mM GTP supplemented with equimolar $MgCl_2$, and 0.5 % NanoGlo (Promega), in the absence or the presence of release factors eRF1 and eRF3a at various concentrations. Luminescence was measured at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) for 12 min. The translation efficiency was calculated as the maximal derivative of the growing linear section of the luminescence curve (v_0, RLU/min). For comparison of different preTCs, a single working concentration of 3 nM of eRF1-eRF3A was chosen in the linear section. The instability of the preTC-NLuc was evaluated in Termi-Luc assay in the absence of release factors.

The amount of preTC-Nluc was calculated as a maximum of relative luminescence (RLU_{max}). To determine concentration of preTC-Nluc, 1.25 μL of preTC sample were incubated in the presence of excess of eRF1 and eRF3a (100 nM). Luminescence was measured at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) for 40 min. PreTC concentration was considered in the translation termination rate experiments.

Translation of NLuc in RRL lysates

For translation of NLuc, 19 μL of 70% RRL lysate prepared as described in Termi-Luc section, was mixed with 0.19 μL of NanoGlo and 0.25 pmol of NLuc mRNA. Luminescence was measured at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) for 60 min.

The translation efficiency was calculated as a maximal derivative of the growing linear section of the luminescence curve (v_0, RLU/min).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis, statistical tests, and visualization

All experiments were performed with a sample size of at least n=3. For transient transfections, each replicate represents an independent transfection. For stably integrated cells, the same cell line was plated in 3 different wells, from which RNA and protein were harvested. Statistical significance between various populations was calculated using a student's t-test in R and p-values were two-sided (*p<0.05, **p<0.005, ***p<0.0005). Statistical details of specific experiments can be found in the Results, Methods, and/or Figure Legends. Plots were generated using R plotting functions and/or the ggplot2 package. The exact code and packages used for these analyses can be found in this repository: https://github.com/jagannathan-lab/2023-kolakada_et_al.

SUPPLEMENTAL INFORMATION

SUPPLEMENTARY FIGURES

Figure S1

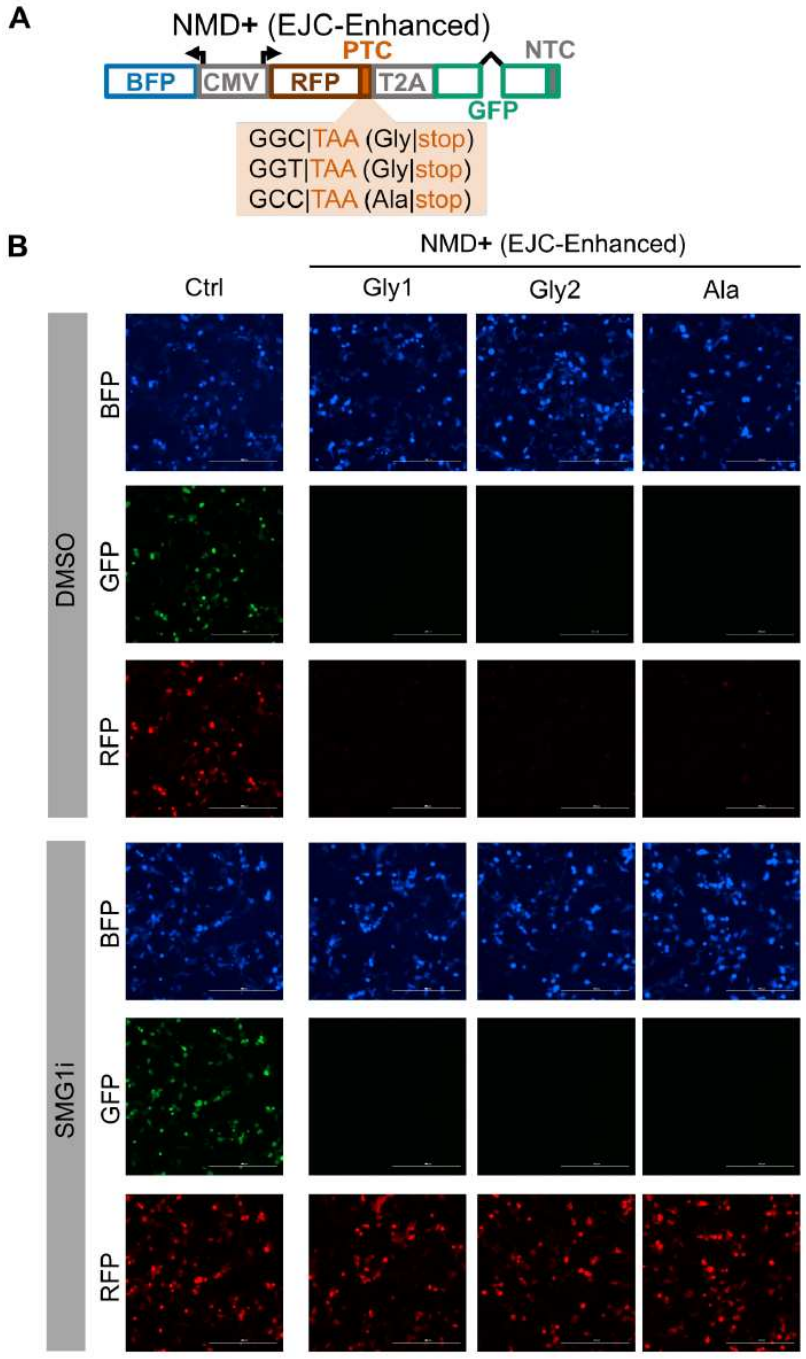
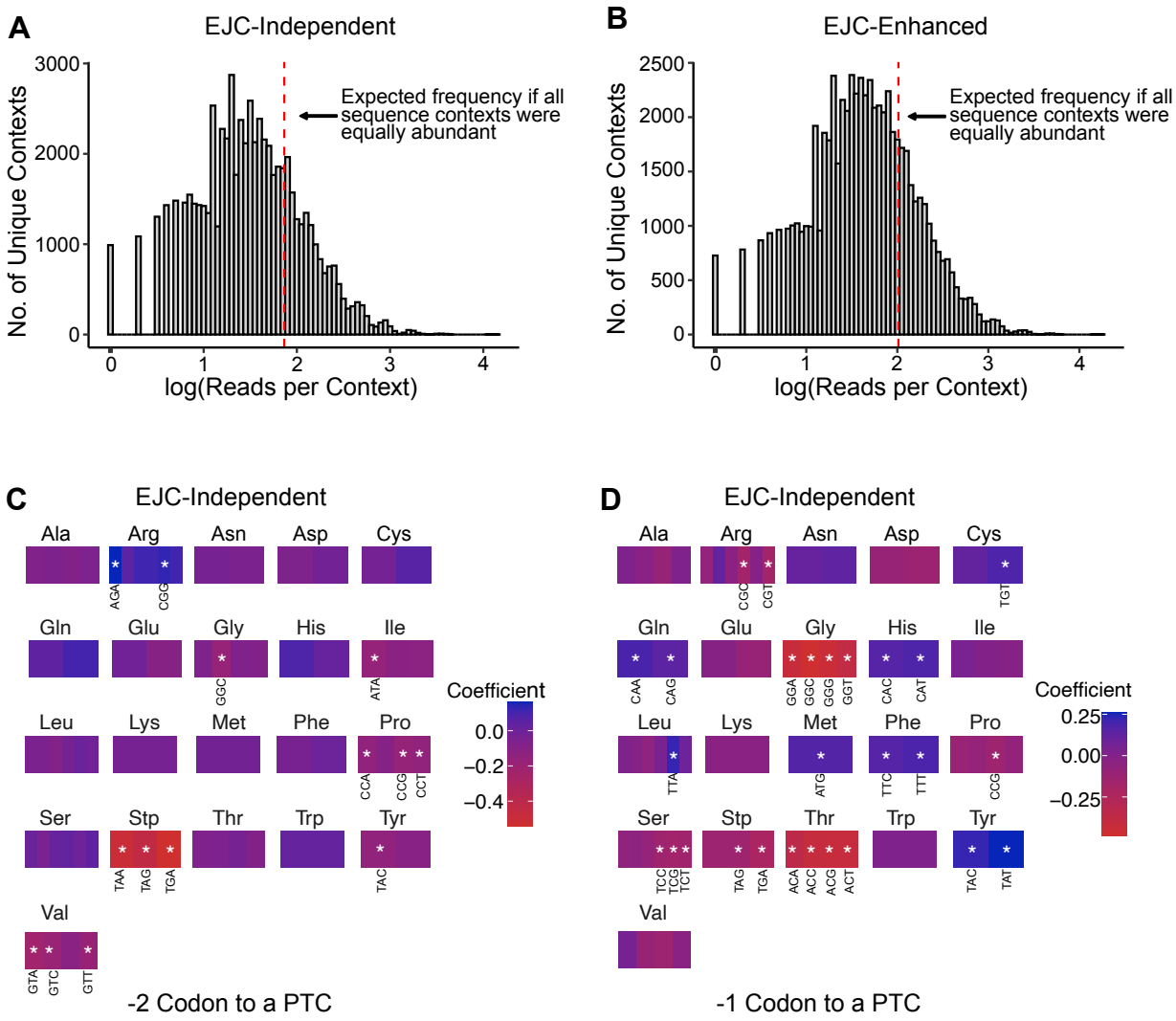


Figure S1. EJC-enhanced reporters do not express RFP in any context. (A) Schematic of EJC-enhanced NMD reporter system used. (B) Fluorescence imaging of BFP, GFP, and RFP

830 from cells transfected with EJC-enhanced NMD reporters containing Gly-PTC or Ala-PTC
831 reporters, treated with a DMSO as a control or SMG1i. Ctrl refers to a construct that robustly
832 expresses all three fluorescent proteins. Gly1 and Gly2 are the GGC and GGT codons for Gly.
833 The codon used for Ala was GCC.

834 **Figure S2.**



835

836 **Figure S2.** MPRA library representation and codon linear modeling. (**A**) Plasmid library
837 representation for the EJC-independent NMD**+** reporter library. (**B**) The same as (A) but for the
838 EJC-enhanced NMD**+** reporter library. (**C**) Heat map depicting the linear modeling coefficients for
839 the EJC-independent library for each codon -2 to a PTC. (**D**) Same as C but for the -1 codon.

840

841

Figure S3.

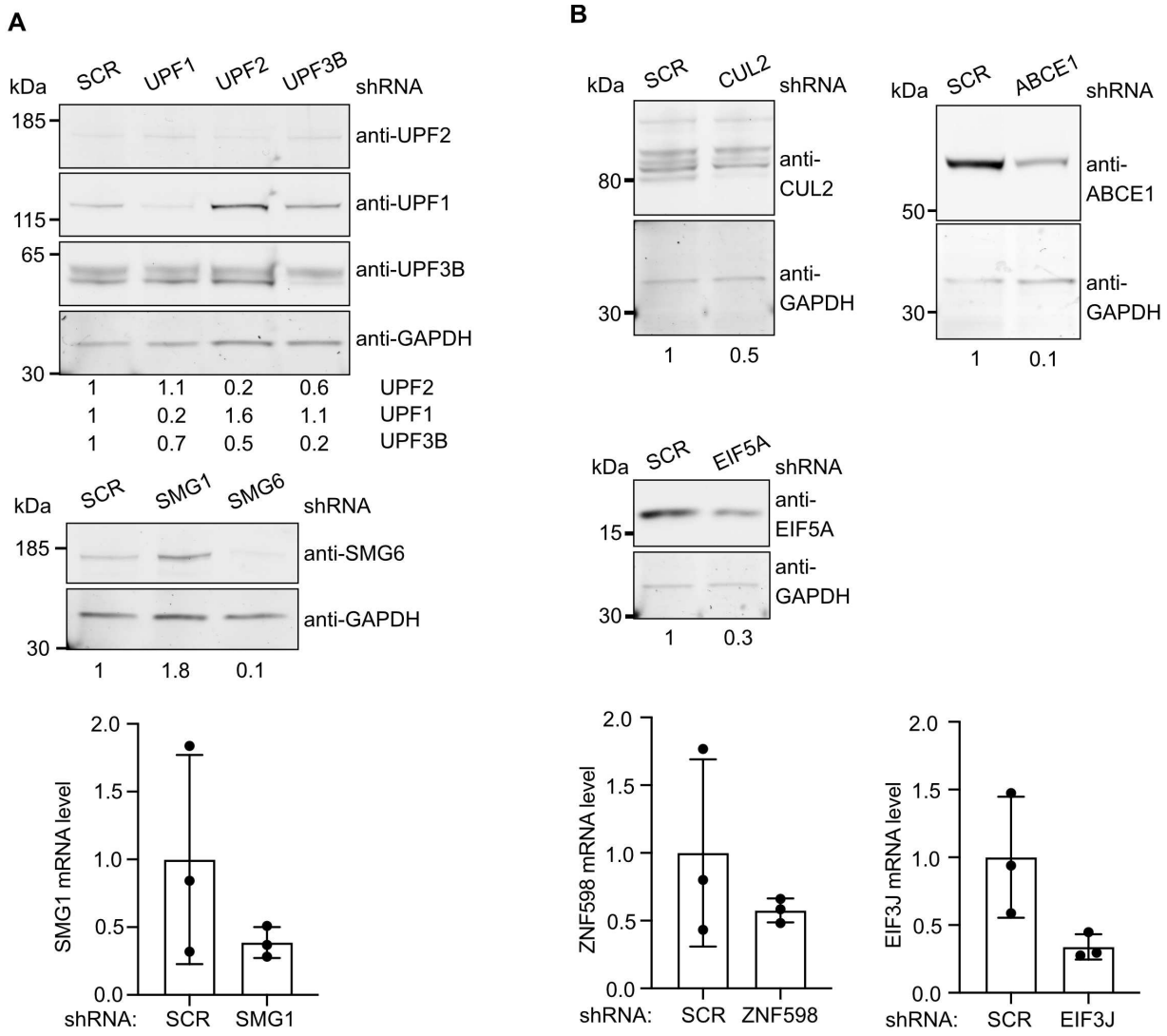
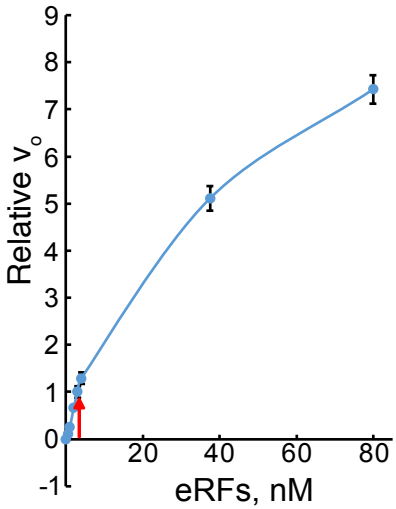


Figure S3. Knockdown validations for all shRNAs used. (**A**) Top: western blot validating UPF1, UPF2, and UPF3B knockdowns; middle: western blot validating SMG6 knockdown; bottom: qPCR validation of SMG1 knockdown. (**B**) Top left: western blot for CUL2 knockdown validation; top right: western blot for ABCE1 knockdown validation; middle left: western blot for EIF5A knockdown validation; bottom left: qPCR validation of ZNF598 knockdown; bottom right: qPCR validation for EIF3J knockdown. GAPDH was the loading control for all western blots. RPL27 was the housekeeping gene for all qPCRs. qPCRs were performed in triplicate.

853 **Figure S4.**

854



855

856 **Figure S4.** Determining the optimum concentration of release factors to use for the Termi-Luc

857 assay.

858
859

860 **Table 1:** Confusion matrices for the EJC-independent and EJC-enhanced classifier

Prediction	Reference	n
EJC-Independent		
NMD Escape	NMD Escape	504
Strong NMD	NMD Escape	96
NMD Escape	Strong NMD	271
Strong NMD	Strong NMD	329
EJC-Enhanced		
NMD Escape	NMD Escape	166
Strong NMD	NMD Escape	434
NMD Escape	Strong NMD	131
Strong NMD	Strong NMD	469

861

862 **Table 2:** Model metrics for the EJC-independent and EJC-enhanced classifier

EJC-Independent	
Sensitivity	0.84
Specificity	0.55
Pos Pred Value	0.65
Neg Pred Value	0.77
Precision	0.65
Recall	0.84
F1	0.73
Prevalence	0.50
Detection Rate	0.42
Detection Prevalence	0.65
Balanced Accuracy	0.69
EJC-Enhanced	
Sensitivity	0.28
Specificity	0.78
Pos Pred Value	0.56
Neg Pred Value	0.52
Precision	0.56
Recall	0.28
F1	0.37
Prevalence	0.50
Detection Rate	0.14
Detection Prevalence	0.25
Balanced Accuracy	0.53

SUPPLEMENTARY DATA

Supplementary Table 1. DEseq2 results for the EJC-independent library of sequences representing the fold change of DMSO over SMG1i treatment of cells.

Supplementary Table 2. DEseq2 results for the EJC-enhanced library of sequences representing the fold change of DMSO over SMG1i treatment of cells.

REFERENCES

1. Miller, J.N. and Pearce, D.A. (2014) Nonsense-mediated decay in genetic disease: friend or foe? *Mutat Res Rev Mutat Res*, **762**, 52-64.

2. Mort, M., Ivanov, D., Cooper, D.N. and Chuzhanova, N.A. (2008) A meta-analysis of nonsense mutations causing human genetic disease. *Human mutation*, **29**.

3. Nagy, E. and Maquat, L.E. (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci*, **23**, 198-199.

4. Gerbracht, J.V., Boehm, V., Britto-Borges, T., Kallabis, S., Wiederstein, J.L., Ciriello, S., Aschemeier, D.U., Kruger, M., Frese, C.K., Altmuller, J. *et al.* (2020) CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex. *Nucleic Acids Res*, **48**, 8626-8644.

5. Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science*, **293**, 1836-1839.

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

7. Kurosaki, T. and Maquat, L.E. (2013) Rules that govern UPF1 binding to mRNA 3' UTRs. *Proc Natl Acad Sci U S A*, **110**, 3357-3362.

8. Sato, H. and Singer, R.H. (2021) Cellular variability of nonsense-mediated mRNA decay. *Nat Commun*, **12**, 7203.

9. Neu-Yilik, G., Amthor, B., Gehring, N.H., Bahri, S., Paidassi, H., Hentze, M.W. and Kulozik, A.E. (2011) Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. *RNA*, **17**, 843-854.

10. Zhang, J. and Maquat, L.E. (1997) Evidence that translation reinitiation abrogates nonsense-mediated mRNA decay in mammalian cells. *EMBO J*, **16**, 826-833.

11. Lindeboom, R.G.H., Vermeulen, M., Lehner, B. and Supek, F. (2019) The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet*, **51**, 1645-1651.

12. Jagannathan, S. and Bradley, R.K. (2016) Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Res*, **26**, 1639-1650.

13. Teran, N.A., Nachun, D.C., Eulalio, T., Ferraro, N.M., Smail, C., Rivas, M.A. and Montgomery, S.B. (2021) Nonsense-mediated decay is highly stable across individuals and tissues. *Am J Hum Genet*, **108**, 1401-1408.

14. Loughran, G., Chou, M.Y., Ivanov, I.P., Jungreis, I., Kellis, M., Kiran, A.M., Baranov, P.V. and Atkins, J.F. (2014) Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res*, **42**, 8928-8938.

15. Tork, S., Hatin, I., Rousset, J.P. and Fabret, C. (2004) The major 5' determinant in stop codon read-through involves two adjacent adenines. *Nucleic Acids Res*, **32**, 415-421.

16. Baker, S.L. and Hogg, J.R. (2017) A system for coordinated analysis of translational readthrough and nonsense-mediated mRNA decay | PLOS ONE. *PLoS ONE*, **12(3): e0173980**.

17. McCaughan, K.K., Brown, C.M., Dalphin, M.E., Berry, M.J. and Tate, W.P. (1995) Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proc Natl Acad Sci U S A*, **92**, 5431-5435.

18. Pierson, W.E., Hoffer, E.D., Keedy, H.E., Simms, C.L., Dunham, C.M. and Zaher, H.S. (2016) Uniformity of Peptide Release Is Maintained by Methylation of Release Factors. *Cell Rep*, **17**, 11-18.

19. Mangkalaphiban, K., He, F., Ganesan, R., Wu, C., Baker, R. and Jacobson, A. (2021) Transcriptome-wide investigation of stop codon readthrough in Saccharomyces cerevisiae. *PLoS Genet*, **17**, e1009538.

20. Bonetti, B., Fu, L., Moon, J. and Bedwell, D.M. (1995) The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. *J Mol Biol*, **251**, 334-345.

21. Cridge, A.G., Crowe-McAuliffe, C., Mathew, S.F. and Tate, W.P. (2018) Eukaryotic translational termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA channel. *Nucleic Acids Res*, **46**, 1927-1944.

22. Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S. and Jacobson, A. (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature*, **432**, 112-118.

23. Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V. and Izaurralde, E. (2007) A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J*, **26**, 1591-1601.

24. Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W. and Kulozik, A.E. (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J*, **27**, 736-747.

25. Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A., Schaffitzel, C. and Alkalaeva, E. (2016) PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Res*, **44**, 7766-7776.

26. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B. *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science*, **335**, 823-828.

27. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434-443.

28. Koren, I., Timms, R.T., Kula, T., Xu, Q., Li, M.Z. and Elledge, S.J. (2018) The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell*, **173**, 1622-1635 e1614.

29. Lin, H.C., Yeh, C.W., Chen, Y.F., Lee, T.T., Hsieh, P.Y., Rusnac, D.V., Lin, S.Y., Elledge, S.J., Zheng, N. and Yen, H.S. (2018) C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases. *Mol Cell*, **70**, 602-613 e603.

30. Loughran, G., Li, X., O'Loughlin, S., Atkins, J.F. and Baranov, P.V. (2023) Monitoring translation in all reading frames downstream of weak stop codons provides mechanistic insights into the impact of nucleotide and cellular contexts. *Nucleic Acids Res*, **51**, 304-314.

31. Kolakada, D., Campbell, A.E., Galvis, L.B., Li, Z., Lore, M. and Jagannathan, S. (2023) A system of reporters for comparative investigation of EJC-independent and EJC-enhanced nonsense-mediated mRNA decay. *bioRxiv*, 2023.2011.2014.567061.

32. Khandelia, P., Yap, K. and Makeyev, E.V. (2011) Streamlined platform for short hairpin RNA interference and transgenesis in cultured mammalian cells. *Proc Natl Acad Sci U S A*, **108**, 12799-12804.

33. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.

34. Buhler, M., Steiner, S., Mohn, F., Paillusson, A. and Muhlemann, O. (2006) EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat Struct Mol Biol*, **13**, 462-464.

35. Chu, V., Feng, Q., Lim, Y. and Shao, S. (2021) Selective destabilization of polypeptides synthesized from NMD-targeted transcripts. *Mol Biol Cell*, **32**, ar38.

973 36. Singh, G., Rebbapragada, I. and Lykke-Andersen, J. (2008) A competition between
974 stimulators and antagonists of Upf complex recruitment governs human nonsense-
975 mediated mRNA decay. *PLoS Biol*, **6**, e111.

976 37. Grosjean, H. and Westhof, E. (2016) An integrated, structure- and energy-based view of
977 the genetic code. *Nucleic Acids Res*, **44**, 8020-8040.

978 38. Bazzini, A.A., Del Viso, F., Moreno-Mateos, M.A., Johnstone, T.G., Vejnar, C.E., Qin, Y.,
979 Yao, J., Khokha, M.K. and Giraldez, A.J. (2016) Codon identity regulates mRNA stability
980 and translation efficiency during the maternal-to-zygotic transition. *EMBO J*, **35**, 2087-
981 2103.

982 39. Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M.,
983 Dreyfuss, G. and Ohno, S. (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex
984 (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-
985 mediated mRNA decay. *Genes Dev*, **20**, 355-367.

986 40. Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu,
987 R., Morita, T., Iwamatsu, A., Hachiya, T. *et al.* (2009) SMG-8 and SMG-9, two novel
988 subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex
989 during nonsense-mediated mRNA decay. *Genes Dev*, **23**, 1091-1105.

990 41. Arias-Palomo, E., Yamashita, A., Fernandez, I.S., Nunez-Ramirez, R., Bamba, Y., Izumi,
991 N., Ohno, S. and Llorca, O. (2011) The nonsense-mediated mRNA decay SMG-1 kinase
992 is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev*,
993 **25**, 153-164.

994 42. Wallmeroth, D., Lackmann, J.W., Kueckelmann, S., Altmuller, J., Dieterich, C., Boehm,
995 V. and Gehring, N.H. (2022) Human UPF3A and UPF3B enable fault-tolerant activation
996 of nonsense-mediated mRNA decay. *EMBO J*, **41**, e109191.

997 43. Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V. and Hegde,
998 R.S. (2018) ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Mol Cell*, **72**,
999 469-481 e467.

1000 44. Schrader, R., Young, C., Kozian, D., Hoffmann, R. and Lottspeich, F. (2006)
1001 Temperature-sensitive eIF5A mutant accumulates transcripts targeted to the nonsense-
1002 mediated decay pathway. *J Biol Chem*, **281**, 35336-35346.

1003 45. Pelechano, V. and Alepuz, P. (2017) eIF5A facilitates translation termination globally
1004 and promotes the elongation of many non polyproline-specific tripeptide sequences.
1005 *Nucleic Acids Res*, **45**, 7326-7338.

1006 46. Schuller, A.P., Wu, C.C., Dever, T.E., Buskirk, A.R. and Green, R. (2017) eIF5A
1007 Functions Globally in Translation Elongation and Termination. *Mol Cell*, **66**, 194-205
1008 e195.

1009 47. Egorova, T., Biziaev, N., Shuvalov, A., Sokolova, E., Mukba, S., Evmenov, K., Zotova,
1010 M., Kushchenko, A., Shuvalova, E. and Alkalaeva, E. (2021) eIF3j facilitates loading of
1011 release factors into the ribosome. *Nucleic Acids Res*, **49**, 11181-11196.

1012 48. Jackson, R.J., Hellen, C.U. and Pestova, T.V. (2012) Termination and post-termination
1013 events in eukaryotic translation. *Adv Protein Chem Struct Biol*, **86**, 45-93.

1014 49. Hellen, C.U.T. (2018) Translation Termination and Ribosome Recycling in Eukaryotes.
1015 *Cold Spring Harb Perspect Biol*, **10**.

1016 50. Zhu, X., Zhang, H. and Mendell, J.T. (2020) Ribosome Recycling by ABCE1 Links
1017 Lysosomal Function and Iron Homeostasis to 3' UTR-Directed Regulation and
1018 Nonsense-Mediated Decay. *Cell Rep*, **32**, 107895.

1019 51. Frolova, L.Y., Tsivkovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov,
1020 V.M., Tatkov, S.I. and Kisselev, L.L. (1999) Mutations in the highly conserved GGQ motif
1021 of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-
1022 tRNA hydrolysis. *RNA*, **5**, 1014-1020.

1023 52. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B.,
1024 Hart, J., Hoffman, D., Jang, W. *et al.* (2018) ClinVar: improving access to variant
1025 interpretations and supporting evidence. *Nucleic Acids Res*, **46**, D1062-D1067.

53. Tate, W.P. and Mannering, S.A. (1996) Three, four or more: the translational stop signal at length. *Mol Microbiol*, **21**, 213-219.

54. Imamachi, N., Salam, K.A., Suzuki, Y. and Akimitsu, N. (2017) A GC-rich sequence feature in the 3' UTR directs UPF1-dependent mRNA decay in mammalian cells. *Genome Res*, **27**, 407-418.

55. Lejeune, F. and Maquat, L.E. (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol*, **17**, 309-315.

56. Mazo-Vargas, A., Park, H., Aydin, M. and Buchler, N.E. (2014) Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy. *Mol Biol Cell*, **25**, 3699-3708.

57. Smith, T., Heger, A. and Sudbery, I. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*, **27**, 491-499.

58. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*, **17**, 3.

59. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.

60. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

61. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402-408.

62. Frolova, L., Seit-Nebi, A. and Kisselev, L. (2002) Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA*, **8**, 129-136.

1051 63. Susorov, D., Egri, S. and Korostelev, A.A. (2020) Termi-Luc: a versatile assay to monitor
1052 full-protein release from ribosomes. *RNA*, **26**, 2044-2050.

1053 64. Shuvalov, A., Shuvalova, E., Biziaev, N., Sokolova, E., Evmenov, K., Pustogarov, N.,
1054 Arnautova, A., Matrosova, V., Egorova, T. and Alkalaeva, E. (2021) Nsp1 of SARS-CoV-
1055 2 stimulates host translation termination. *RNA Biol*, **18**, 804-817.

1056 65. Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L. and Pestova, T.V. (2006) In
1057 vitro reconstitution of eukaryotic translation reveals cooperativity between release factors
1058 eRF1 and eRF3. *Cell*, **125**, 1125-1136.

1059