Systematic analysis of nonsense variants uncovers peptide release rate as a novel modifier of nonsense-mediated mRNA decay efficiency

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[1,5], Jay R. Hesselberth[1,3], Neelanjan Mukherjee[1,3], Elena Alkalaeva[4], and Sujatha Jagannathan[1,3,6]*

[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]Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland

[6]Lead contact

*Corresponding Author: sujatha.jagannathan@cuanschutz.edu

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

Keywords: NMD, PTC, translation, termination, splicing

ABSTRACT

Nonsense variants underlie many genetic diseases. The phenotypic impact of nonsense variants is determined by Nonsense-mediated mRNA decay (NMD), which degrades transcripts with premature termination codons (PTCs). NMD activity varies across transcripts and cellular contexts via poorly understood mechanisms. Here, by leveraging human genetic datasets, we uncover that the amino acid preceding the PTC dramatically affects NMD activity in human cells. We find that glycine codons in particular support high levels of NMD and are enriched before PTCs but depleted before normal termination codons (NTCs). Gly-PTC enrichment is most pronounced in human genes that tolerate loss-of-function variants. This suggests a strong biological impact for Gly-PTC in ensuring robust elimination of potentially toxic truncated proteins from non-essential genes. Biochemical assays revealed that the peptide release rate during translation termination is highly dependent on the identity of the amino acid preceding the stop codon. This release rate is the most critical feature determining NMD activity across our massively parallel reporter assays. Together, we conclude that NMD activity is significantly modulated by the "window of opportunity" offered by translation termination kinetics. Integrating the window of opportunity model with the existing framework of NMD would enable more accurate nonsense variant interpretation in the clinic.

INTRODUCTION

Mutations that introduce a premature stop codon account for over 30% of inherited human diseases due to the action of nonsense-mediated RNA decay (NMD) (Miller & Pearce, 2014). NMD eliminates transcripts that contain premature termination codons (PTCs) that would otherwise generate truncated protein products. Much of our understanding of NMD comes from studies of naturally occurring nonsense mutations and genotype-phenotype correlations in genes involved in Mendelian disorders such as β-globin (Neu-Yilik et al., 2011) (*HBB*), Dystrophin (Kerr et al., 2001) (*DMD*), and Lamin A/C (van Engelen et al., 2005) (*LMNA*). For example, individuals with a nonsense mutation in early exons of *HBB* develop a recessive form of β-thalassemia due to degradation of the mutant mRNA via NMD (Chang & Kan, 1979), whereas nonsense mutations in the last exon of *HBB* cause a dominant form of the disease due to the mutant mRNA escaping NMD and leading to the production of a C-terminally truncated beta-globin protein (Kazazian et al., 1992). Such studies have helped refine the rules of NMD, including the “55-nt rule”, which describes the NMD insensitivity of nonsense mutations in the last and penultimate exons up to 55 nucleotides upstream of the final exon junction complex (Popp & Maquat, 2016).

Several exceptions to these rules have also been observed: nonsense mutations found close to the start codon often escape NMD due to downstream reinitiation of translation (Inacio et al., 2004; Pereira et al., 2015; Zhang & Maquat, 1997); stop codon readthrough has been reported to contribute to the escape of some nonsense variants from NMD (Hogg & Goff, 2010; Keeling et al., 2004); and NMD efficiency has been found to vary across cell types (Linde et al., 2007), tissues (Bateman et al., 2003; Zetoune et al., 2008), and individuals (Kerr et al., 2001; Nguyen et al., 2014), offering potential explanations for tissue tropism and varying penetrance across the population of disease-causing nonsense variants. A systematic study of PTCs in the Cancer Genome Atlas database found that the 55-nt rule only predicted the fate of ~50% of PTCs; while additional factors such as exon length could account for another ~20% of NMD variability, ~30% of NMD variability is still unexplained (Lindeboom et al., 2016; Lindeboom et al., 2019). Another study investigating nonsense variants that are highly frequent in the healthy human population found that ~50% of PTC-containing transcripts escape NMD via mechanisms including alternative splicing, stop codon readthrough, and alternative translation initiation (Jagannathan & Bradley, 2016). If every premature termination event that satisfies the 50-55 nt rule does not lead to NMD, what other factors influence the probability and activity of NMD?

As premature translation termination is the key signal for NMD, termination efficiency may impact 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 (Baker & Hogg, 2017; Loughran et al., 2014; McCaughan et al., 1995; Pierson et al., 2016; Tork et al., 2004). For example, specific elements within the stop codon sequence context and certain viral RNA structures can reduce termination fidelity (Baker & Hogg, 2017; Bonetti et al., 1995; Cridge et al., 2018; Loughran et al., 2014; Mangkalaphiban et al., 2021; Tork et al., 2004). 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 (McCaughan et al., 1995; Pierson et al., 2016). However, whether and how local sequence context and translation termination fidelity or kinetics influence the probability that a given transcript will undergo NMD remains to be determined.

In this study, we hypothesized that the sequence contexts surrounding PTCs may hold the key to understanding NMD variability among transcripts. Indeed, we found that glycine codons are highly enriched immediately upstream of PTCs among healthy individuals and that this sequence context potently enhances NMD activity. To systematically assess the impact of PTC sequence context on NMD activity, we performed a massively parallel reporter assay (MPRA) and found a broad range of NMD activities associated with different sequences. Statistical modeling of the MPRA data identified peptide release rate during translation termination as the major predictor of NMD activity, which we verified using biochemical assays. We propose that variable peptide release rates alter NMD activity by modulating the kinetic window available for NMD factors to bind and enact NMD during translation termination. Our data suggest that the Gly-PTC context coevolved with NMD to efficiently eliminate transcripts containing premature stop codons.

RESULTS

Glycine codons are enriched preceding PTCs in non-essential genes. Healthy individuals harbor ~100 putative loss of function genetic variants, most of which are protein-truncating (MacArthur et al., 2012). In previous work, we showed that such PTCs often escape NMD via mechanisms including alternative splicing or stop codon readthrough, allowing them to rise to high frequencies in the healthy population (minor allele frequency, MAF > 5%) (Jagannathan & Bradley, 2016). In this study, we focused on rare variants with MAF ≤ 1% in healthy individuals,

reasoning that such variants are more likely to undergo NMD and yet may have varying

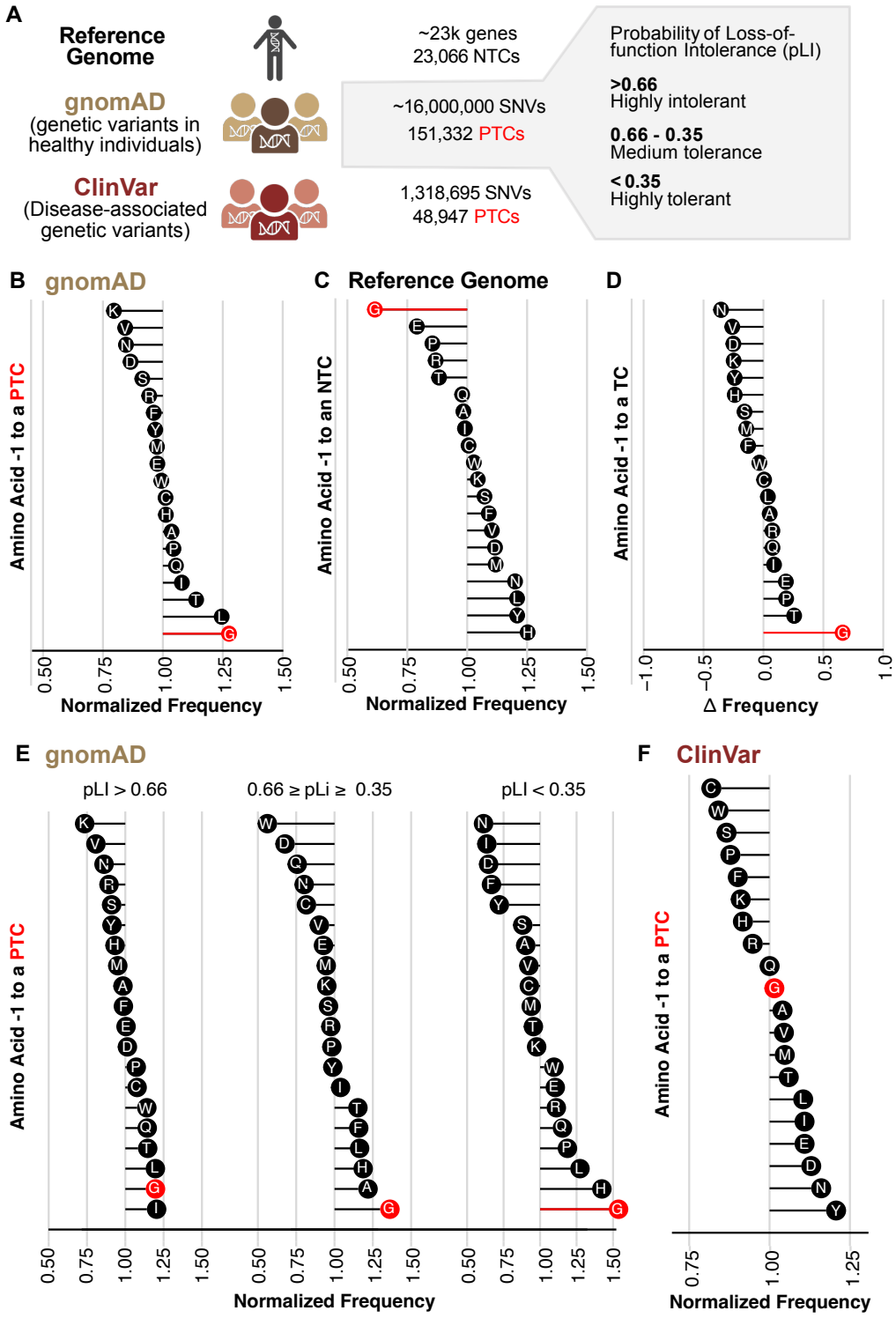


Figure 1. Gly is enriched preceding a PTC in non-essential genes. (**A**) Schematic for analyzing LoF scores for rare PTCs in healthy individuals and PTCs in disease contexts. (**B-C**) Amino acid enrichment -1 to a PTC (B) or an NTC (C). (**D**) Each amino acid's frequency delta (enrichment before a PTC-NTC) at the -1 position. (**E**) The -1 amino acid enrichment of PTCs binned into 3 categories of probability of loss-of-

110 function intolerance (pLI). The lower the pLI, the more tolerant the PTCs are to loss of function. (**F**) The -1
111 amino acid enrichment of PTCs in disease contexts.

112

113 efficiencies. Using the gnomAD database (n=151,332, v2.1.1), which contains genetic variant
114 data for >700,000 healthy humans (Karczewski et al., 2020), we analyzed amino acid enrichment
115 within the 10 amino acids preceding PTCs (**Fig. 1A, Supp. Fig. 1A)**. We compared the enrichment
116 of these amino acids to that of normal termination codon (NTC) contexts in the reference genome
117 (n=23,066, **Fig. 1A, Supp. Fig. 1B**) using the method similar to Koren *et al* (Koren et al., 2018).
118 Gly codons at the -1 position stood out among all examined positions as the most enriched before
119 PTCs and least enriched before the NTC (**Fig. 1B-C,** and **Supp. Fig. 1**). Threonine (Thr) and
120 proline (Pro) were the next most enriched in PTC compared to NTC at the -1 position (**Fig. 1D**).
121 Other amino acids show the opposite trend – enriched before an NTC and depleted before a PTC
122 – including asparagine (Asn), valine (Val), and aspartic acid (Asp). These data suggest that the
123 enrichment of Gly immediately upstream of a PTC (Gly-PTC) has a functional impact that allowed
124 it to rise to high frequency in healthy individuals.

125 To further explore the functional impact of Gly-PTC enrichment in healthy individuals, we
126 analyzed the relative frequency of PTC-inducing variants binned in terms of the probability of LoF
127 intolerance (pLI) score of their host genes (**Fig. 1A** and **1E**). The lower the pLI, the more tolerant
128 to loss of function a gene is and the less essential. We found that Gly-PTC contexts were enriched
129 in genes with lower pLI, suggesting that these PTC-containing variants were more tolerant to loss
130 of function (**Fig. 1E**). We also examined the amino acids enriched immediately upstream of PTCs
131 from the ClinVar database, which contains human genetic variants associated with disease
132 (n=48,847, **Fig. 1A**) (Landrum et al., 2018). Intriguingly, we found no enrichment of Gly at the -1
133 position of disease-associated PTCs (**Fig. 1F**). In summary, the Gly-PTC enrichment in non-
134 essential genes in comparison to essential or disease-relevant genes suggests a non-pathogenic
135 and homeostatic role for this sequence context.

136

137 **Massively parallel reporter assay (MPRA) captures widespread sequence-dependent**
138 **variation in NMD activity.** To systematically examine the relationship between NMD activity and
139 the PTC sequence context, we conducted an MPRA using a set of NMD reporters we previously
140 developed (Kolakada et al., 2024). These reporters contain two types of NMD targets: one
141 triggered by the presence of exon-junction complexes downstream of a PTC ("EJC-enhanced")
142 and the other by the presence of a long 3' UTR ("EJC-independent"). Because these targets rely

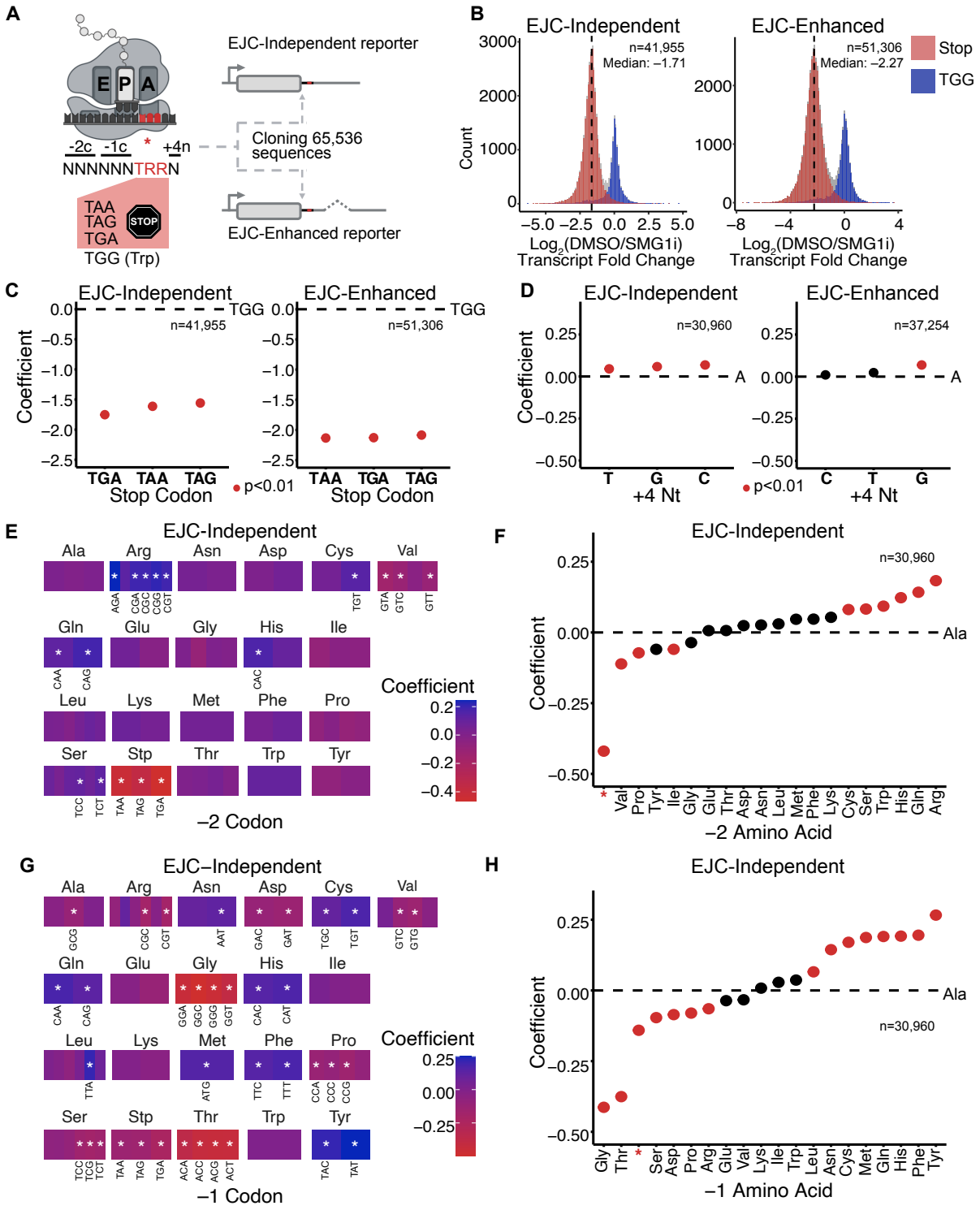
143 on related but distinct cues to trigger NMD, they might respond differently to PTC sequence
144 variation.

145 We varied the 6 nucleotides before the PTC, the PTC itself, and 1 nucleotide after the PTC in
146 an unbiased fashion (N_6TRRN; corresponding to 65,536 unique sequences; **Fig. 2A**) and asked
147 how these sequences impacted NMD activity in both EJC-independent and EJC-enhanced
148 reporters. Because we employed stop codons with "TRR" for array-based oligonucleotide
149 synthesis ("R" being a purine), 25% of our sequence library consists of a tryptophan (Trp) codon
150 (TGG) instead of a PTC – thereby serving as control sequences that do not undergo NMD. All
151 sequences were incorporated into both EJC-independent and EJC-enhanced fluorescent reporter
152 backbones (**Supp. Fig. 2A-B**) and integrated into a single genomic location in HEK293T cells via
153 Cre-Lox mediated recombination (Khandelia et al., 2011). These cells were then treated with
154 either DMSO or SMG1i, harvested for RNA, and subjected to targeted sequencing to quantify the
155 levels of each sequence context in the pool of cells. The NMD activity of an individual reporter
156 sequence was calculated as the $\log_2$ transcript fold change between DMSO-treated cells (NMD
157 active) and SMG1i-treated cells (NMD-inhibited) as measured by DESeq2 (Love et al., 2014),
158 thus a lower (negative) value indicates higher NMD activity, and vice versa. The ~1/4th of control
159 sequences containing the Trp codon rather than a stop codon were used for normalization.

160 The NMD activity of both EJC-independent and EJC-enhanced reporters had a bimodal
161 distribution with clear separation between constructs containing stop codons and the Trp-
162 encoding TGG codon (**Fig. 2B**): sequences containing a stop codon were left-shifted compared
163 to the sequences containing the Trp control, indicating that the former are targets of NMD. In the
164 population of sequences undergoing NMD (i.e. those containing a stop codon), there was a wide
165 range of NMD activity for both the EJC-independent and the EJC-enhanced reporters. EJC-
166 enhanced reporter expression was reduced by an average of 4.8-fold compared to the cells
167 treated with the NMD inhibitor (log2FC: -2.27), while the EJC-independent reporters experienced
168 an average 3.3-fold decrease (log2FC: -1.71) (**Fig. 2B**). These results indicate that EJC-
169 enhanced targets undergo more efficient NMD compared to EJC-independent targets, consistent
170 with prior research (Buhler et al., 2006; Chu et al., 2021; Singh et al., 2008).

171 Next, we investigated if the PTC sequence context could explain the wide range of observed
172 NMD activity. Specifically, we were interested in the contribution of the codons and amino acids
173 at the -1 and -2 positions, the stop codon, and the nucleotide at the +4 position of the PTC. We
174 constructed a linear model for each of these biological variables to determine how much it

175 explained the variation in NMD activity (i.e. transcript fold change) of the associated sequences.



176

177 **Figure 2. MPRA shows a wide range of PTC context-dependent NMD efficiencies. (A)** Design of
178 the MPRA. **(B)** Histogram of transcript fold change EJC-independent and EJC-enhanced NMD+ reporters
179 comparing DMSO treated cells to SMG1i treated cells. **(C-D)** Dot plots of the beta coefficient representing
180 the effect of the stop codon (**C**) and the +4 nt (**D**) on EJC-independent and EJC-enhanced NMD activity.

(E-H) Plots depicting a variable's effect on EJC-independent NMD: (**E**) Heat map for the -2 codon, (**F**) Dot plot for the -2 amino acid, (**G**) Heat map for the -1 codon, and (**H**) Dot plot for the -1 amino acid.

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 TGG (Trp). Negative beta coefficients indicated stronger NMD activity, while positive ones indicated weaker NMD activity. It is important to note that each model (represented by individual plots) is unique, and the beta coefficients cannot be quantitatively compared across the plots.

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. 2C**). Notably, the TGA stop codon had a beta coefficient 0.2 lower than TAG, indicating a 20% increase in NMD activity compared to the TAG stop codon for the EJC-independent reporters (**Supp. Fig. 3A**). The NMD activity of TAA was also higher than TAG (~5%) but not to the extent of TGA for the EJC-Independent reporters. For the EJC-enhanced reporters, both TGA and TAA had higher NMD activity than TGA (~5%) (**Supp. Fig. 3A**). For subsequent analyses, only the 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 ~6% weaker NMD activity compared to A for the EJC-independent reporters. A similar effect was observed for G compared to A for the EJC-enhanced reporters (**Fig. 2D**). Overall, the nucleotide after the PTC only subtly impacted NMD activity.

The identity of the codon at the –2 position had a wide range of effects on NMD activity for the EJC-independent reporters. As the –2 codon position is coded by randomized nucleotides (NNN), it can also include all three stop codons. We found that the effects of the codons were largely grouped by the identity of the encoded amino acid, with only a few that were significantly different (**Fig. 2E**; all codons compared to Ala-GCA). Therefore, we next looked at the codons based on their encoded amino acid identity in a new linear model. We found that a stop codon at the –2 position caused the strongest increase (~40%) in NMD activity compared to the Ala control (**Fig. 2E-F**). More subtle effects were observed across various amino acids, with an increase up to ~10% in NMD activity with Val at the -2 position, or a reduction of up to 20% when Arg was encoded at the -2 position.

213 When examining the effect of the codon at the -1 position on NMD activity, all codons except
214 one Leu codon, were again grouped by the encoded amino acid (**Fig. 2G**; all codons compared
215 to Ala-GCA). Remarkably, we found that Gly-PTC (the same context enriched in the gnomAD
216 database; **Fig. 1**) has the strongest effect on NMD activity, reducing it by 41% compared to Ala
217 (**Fig. 2G-H**). Thr exhibited almost as potent an effect as Gly (~38% enhanced compared to Ala):
218 both Gly-PTC and Thr-PTC resulted in much stronger NMD than a stop codon in the -1 position
219 (~14% enhanced compared to Ala). Indeed, Thr was the second-most enriched amino acid
220 encoded immediately upstream to PTCs in our gnomAD analysis (**Fig. 1D**). In contrast, Tyr-PTC
221 had the least efficient NMD, weakening NMD by ~25% compared to Ala-PTC (**Fig. 2G-H**).

222 For the EJC-enhanced reporters, there were fewer significant changes at both -1 and -2
223 positions. Note that the control feature for the -2 position is Arg or an Arg codon and for the -1
224 position it is Ala or an Ala codon. Control features were picked based on amino acids or codons
225 with beta coefficients in the middle range of the model. One Ala and one Lys codon at the -2
226 position have significant effects compared to Arg-AGA, while one Arg codon shows a significant
227 difference in the -1 position compared to Ala-GCA. That said, the NMD effect size range was
228 much smaller overall compared to the EJC-independent models. Upon aggregation of codons to
229 the encoded amino acids, most amino acids did not affect the efficiency of NMD at the -2 or -1
230 positions, except Lys and Pro at the -2 position, which decreased NMD activity by ~17% or
231 increased it by ~6%, respectively (**Supp Fig. 3A-D**). Interestingly, a stop codon in the -2 position
232 weakened NMD by ~8% compared to the control variable Arg – the opposite effect that we
233 observed with the EJC-independent reporters. Altogether, some codons at the -2 position can
234 marginally influence NMD activity in a manner that appears to depend on whether NMD proceeds
235 via an EJC-independent or EJC-enhanced pathway.

236 At first glance, the discrepancy in results between the EJC-independent and EJC-enhanced
237 reporters is puzzling. However, it is possible that the highly efficient EJC-enhanced NMD is
238 insensitive to other modifying factors or that a narrow range of biological effects may not emerge
239 as statistically significant in a large pool of ~65k sequences. To further explore the basis for this
240 discrepancy, we switched to an independent, luciferase-based reporter system to test a limited
241 number of sequence contexts and their relative impact on EJC-independent and EJC-enhanced
242 NMD activity (Kolakada et al., 2024). In this reporter system, the transfection control is Renilla
243 luciferase and the NMD readout is provided by Firefly luciferase (**Supp. Fig. 3F-G**).

244 As the -1 position of the PTC had the most significant impact on EJC-independent NMD
245 activity in our MPRA, we checked its effect on EJC-independent and EJC-enhanced NMD activity

246 in the luciferase reporter system. We varied the amino acid in the -1 position of either the PTC or
247 the NTC, choosing Gly, Val, Ala, and Tyr, which display a range of NMD efficiencies from highest
248 to lowest (**Supp. Fig. 3F**). As expected, luciferase activity was ~10-fold lower for the EJC-
249 enhanced reporters compared to the EJC-independent reporters (**Supp. Fig. 3G**). Notably, the
250 relative trend in -1 amino acid-dependent NMD variability was the same for both the EJC-
251 independent and EJC-enhanced reporters. The NMD activity varied depending on the -1 amino
252 acid, with Gly having ~22% or ~29% more efficient NMD compared to Tyr for the EJC-independent
253 and EJC-enhanced reporters, respectively (**Supp. Fig. 3G**). Compared to the EJC-independent
254 and EJC-enhanced reporters, the NMD– reporters did not vary with the change in the -1 amino
255 acid (**Supp. Fig. 3G**). Taken together, these data demonstrate that the -1 amino acid effect can
256 extend to both forms of NMD.

257

258 **Key NMD and translation factors do not influence the -1 amino acid-dependent variation in**
259 **NMD activity.** To determine the mechanistic basis of the -1 amino acid-dependent variation in
260 NMD, we next asked whether candidate trans factors played a role in this effect. To this end, we
261 utilized shRNAs to knock down various NMD- and translation-related factors and test the effect
262 of the knockdown on reporter levels using the EJC-independent luminescent reporters.

263 We investigated the UPF proteins (UPF1, UPF2, and UPF3B), which are key effectors of NMD
264 (Kashima et al., 2006; Yamashita et al., 2009); SMG1, which phosphorylates UPF1 to trigger
265 NMD; and SMG6, which is responsible for cleaving the transcript when NMD is initiated (Arias-
266 Palomo et al., 2011; Yamashita et al., 2009) (**Fig. 3A-B**). Reporter protein levels increased upon
267 knockdown of all NMD factors except for UPF3B, likely because UPF3B is known to be redundant
268 to UPF3A (Wallmeroth et al., 2022). However, the relative difference in expression of the reporters
269 based on -1 amino acid identity largely remained the same (**Fig. 3B**; **Supp. Fig. 4A**).

270 Next, using the same approach, we investigated key translation and protein quality control-
271 related factors, CUL2, ZNF598, ABCE1, EIF3J, and EIF5A (**Fig. 3C-D**). CUL2 forms a key part of
272 the protein complex that recognizes C-terminal Gly degrons and could contribute to the decrease
273 observed in Gly-PTC luciferase reporters (Koren et al., 2018; Lin et al., 2018). ZNF598 recognizes
274 the interface of collided disomes, thus playing a role in ribosome-associated quality control
275 substrate recognition (Juszkiewicz et al., 2018). It is possible that some sequences would promote
276 ribosome collision at the end of a transcript, which would then influence NMD. ABCE1 is a
277 ribosome recycling factor that acts during normal and premature termination (**Fig. 3C**) (Hellen,
278 2018; Jackson et al., 2012; Zhu et al., 2020). EIF3J facilitates the binding of the release factors

to the ribosome, and thus stimulates peptide release (Egorova et al., 2021). Finally, 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 (Pelechano & Alepuz, 2017; Schrader et al., 2006; Schuller et al., 2017). EIF5A binding to the termination complex may be influenced by the transcript sequence within the ribosome. Similar to the NMD factors, none of these conditions eliminated the differences in reporter level expression based on -1 amino acid identity (**Fig. 3D**; knockdown validation **Supp. Fig. 4B**).

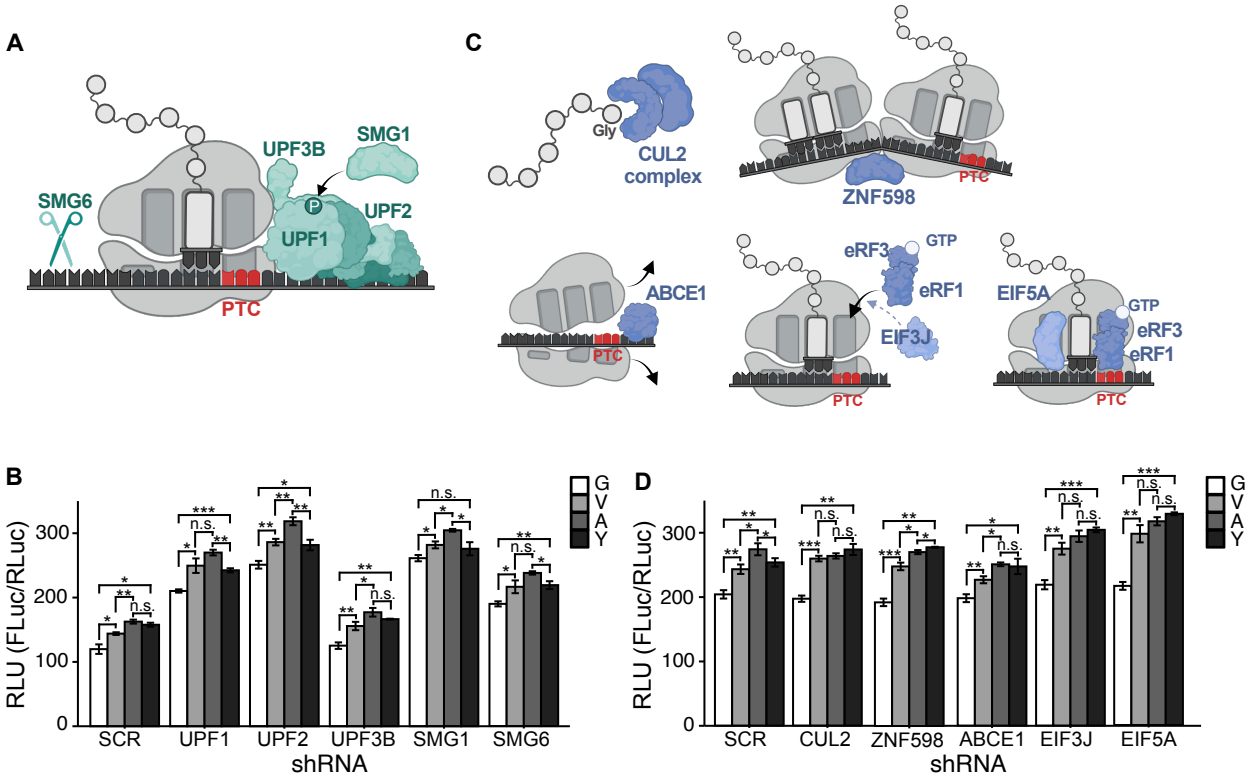


Figure 3. Key NMD and translation factors do not influence the -1 amino acid-dependent variation in NMD activity. (A) Model of NMD factors tested. **(B)** shRNA knockdown of key NMD factors and their effect on the luminescence-based EJC-independent reporters. **(C)** Models of the mechanisms of translation termination, ribosome recycling, and protein degradation factors tested. **(D)** shRNA knockdown of key translation-related factors and their effect on the luminescence-based EJC-independent reporters.

These data show that key NMD, translation, and protein quality control-related trans-factors do not have a role in the -1 amino acid-dependent variation of NMD. Instead, the mechanistic basis of this effect may lie within the intrinsic properties of the sequence.

298 **Peptide release rate is the most important feature that classifies EJC-independent NMD**
299 **substrate activity.** To discover sequence-intrinsic factors that influence NMD, we constructed a
300 random forest classifier trained to discriminate between constructs with high versus low NMD
301 activity (**Fig. 4A**). We selected the 3000 sequences with the highest and lowest NMD activity for
302 the EJC-independent and EJC-enhanced reporters. We then added sequence-intrinsic variables
303 that could influence the classification of high and low NMD activity, including the identities of the
304 codons as well as their properties such as usage frequency (GenScript), tRNA binding strength,
305 and optimality (Bazzini et al., 2016; Grosjean & Westhof, 2016); the identities of the amino acids
306 as well as their physical properties such as charge, solubility, dissociation constants, and
307 hydropathy (ThermoFisher Scientific); and the GC and pyrimidine content of each of the
308 sequences. As termination and NMD are intrinsically linked, we also included the rate of peptide
309 release, which is a measure of the time taken for a release factor to trigger peptidyl-tRNA
310 hydrolysis within the P-site (i.e. the -1 position of a stop codon) and the release of the polypeptide
311 itself. These release rates vary based on the identity of the amino acid present within the P-site
312 during termination. Since release rates are not available for eukaryotic systems, we used the
313 prokaryotic peptide release rates (Pierson et al., 2016). Prokaryotes have two release factors RF1
314 and RF2 that each recognize a subset of the three stop codons, and therefore, we included the
315 rates calculated for both RF1 and RF2 in our classifier.

316 We trained two separate classifiers via train (80%)/ test (20%) cross-validation for the EJC-
317 independent and EJC-enhanced reporters using the same input variables. Classification
318 performance for EJC-independent NMD activity was much better than for EJC-enhanced NMD
319 activity (**Fig. 4B**), as indicated by the substantially higher area under the receiver operating
320 characteristic curve and F1 scores for the EJC-independent (0.75 and 0.73, respectively)
321 compared to the EJC-enhanced (0.57 and 0.37, respectively) classifiers. We concluded that only
322 the classifier for EJC-independent NMD activity was robust and suitable for further interpretation
323 (refer to **Supplementary Tables 1-2** for the confusion matrices and other model metrics). The
324 poor performance of the EJC-enhanced classifier suggests that hitherto unknown variables are
325 missing in the classifier that may better predict these reporters' NMD activity.

326 We focused on the features that were most important for the EJC-independent model.
327 Prokaryotic release factor 1 (RF1) and release factor 2 (RF2) peptide release rates emerged as
328 the most important predictor of NMD activity (**Fig 4C**). Although these rates are measured from a
329 prokaryotic context, the similarity in peptide release mechanisms between prokaryotes and
330 eukaryotes suggests that the relative magnitudes of release rates may also apply to eukaryotic

systems (Hellen, 2018). To visualize 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 of the peptide release rates for the -1 amino acid on NMD activity (**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 factor contributing to the -1 amino acid-dependent impact on NMD activity.

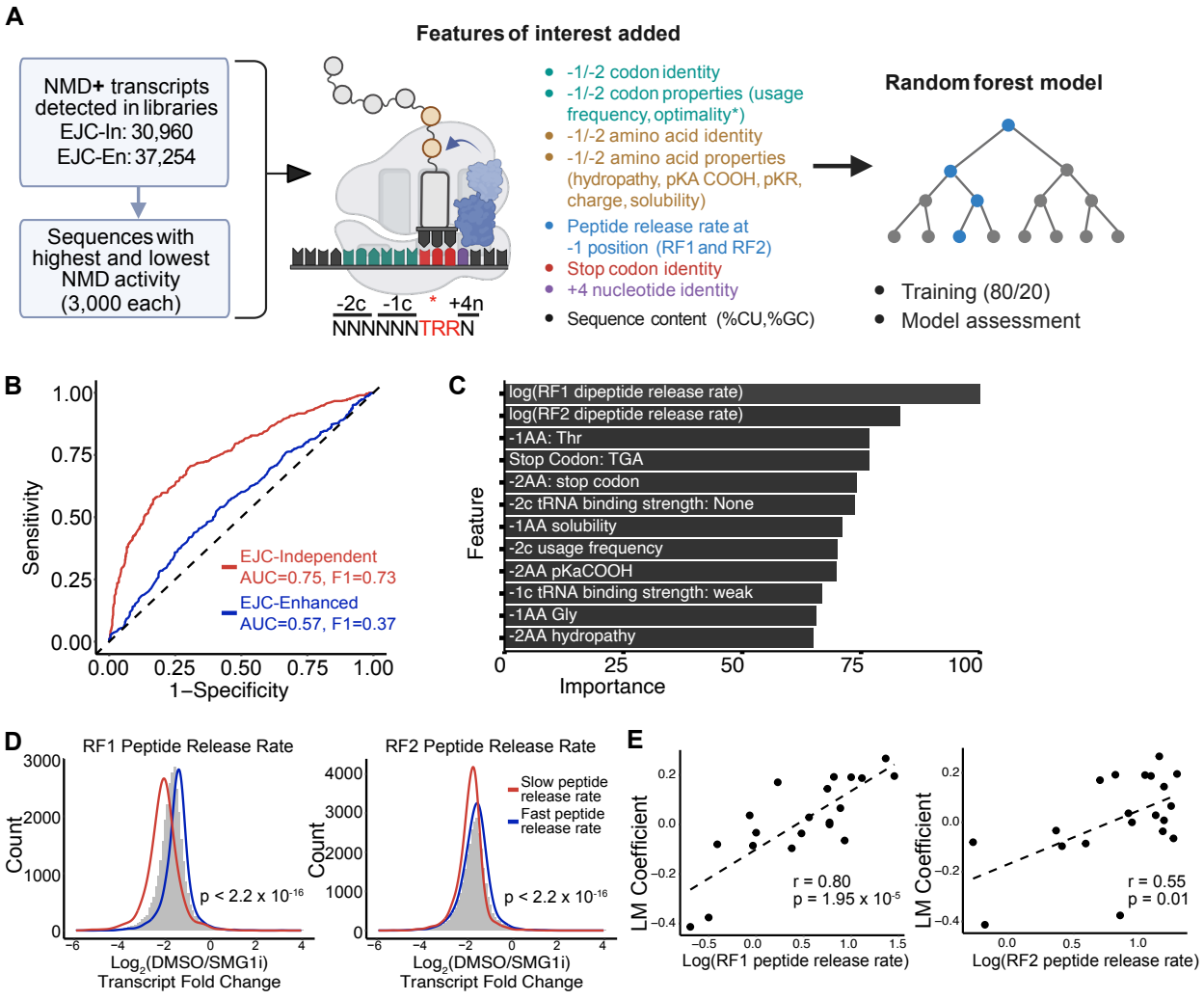


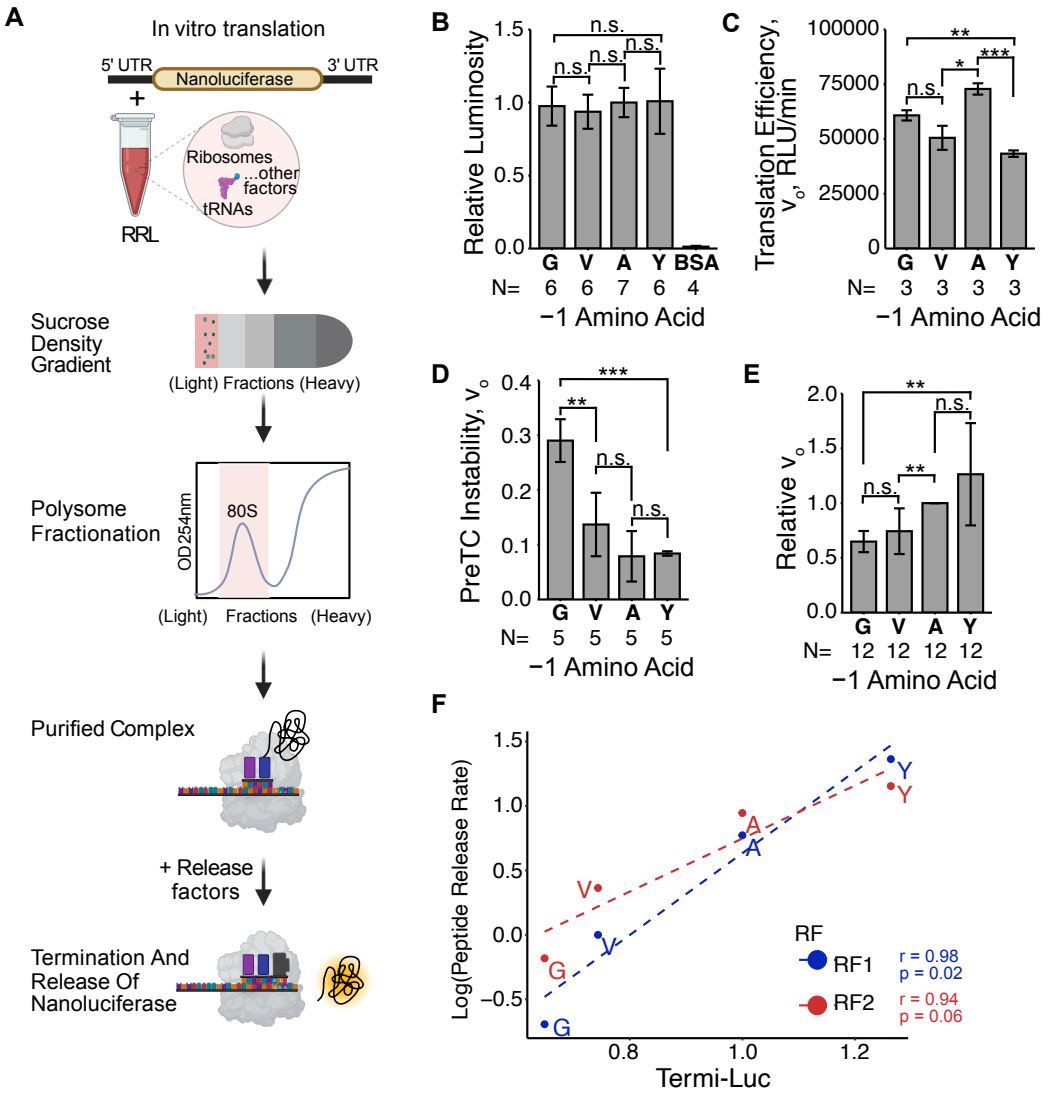
Figure 4. Peptide release rates predict NMD activity (A) Data preparation and variables input into the Random Forest model. **(B)** ROC curves, AUC values, and F1 scores for the EJC-independent and EJC-

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 (Pierson et al., 2016), left: RF1 release rates, right: RF2 release rates. For RF1, slow release rates are defined as log (RF1 release rate) £ –0.5, and fast release rates are defined as log (RF1 release rate) > 1. For RF2, slow release rates are defined as log (RF2 release rate) £ 0, and fast release rates are defined as log (RF2 release rate) > 1.25. p, Wilcoxon test p-value **(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. "r" refers to the Pearson's correlation coefficient and "p" is the corresponding P-value.

Eukaryotic translation termination efficiency varies with the identity of the amino acid upstream of the stop codon *in vitro*. The peptide release rate values used in the Random Forest classifier (**Fig. 4**) came from a prokaryotic system and we wanted to determine whether they reflect the relative peptide release rate in a eukaryotic system (Pierson et al., 2016). 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. 5A**). In this assay, we provide an excess of mutated eukaryotic release factor 1 (eRF1(AGQ)), which can recognize stop codons but unable to induce peptidyl-tRNA hydrolysis (Frolova et al., 1999). eRF1(AGQ) binding thereby traps translation at the termination stage, allowing us to purify pre-termination ribosomal complexes (preTCs) from cells. PreTCs are then exposed to purified, functional release factors (working concentrations determined in **Supp. Fig. 5**), and 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 compare termination rates when Gly, Val, Ala, or Tyr codons (representing the full range of observed NMD efficiencies from high to low) were present at the -1 position to the stop codon.

To check if the substitution of an amino acid residue affected NLuc activity, translation efficiency, or preTC stability of NLuc, control experiments were first performed. To determine whether NLuc activity was impacted by altering the last 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. 5B**). To determine whether translation efficiency was impacted by changing the last amino acid, we calculated the maximal derivative (slope) of the linear portion of the luminescence production curve. The identity of the last amino acid affected translation efficiency with Tyr having the lowest and Ala having the highest efficiency (**Fig. 5C**). To determine preTC stability, the rate of peptide release of each peptide variant was calculated without adding any release factors. These data showed that Gly had the most unstable preTC, while Tyr had the most stable preTC (**Fig. 5D**). Finally, we performed the Termi-Luc assay on

381 each peptide variant starting with equal amounts of preTC complexes to account for the variability
382 in translation efficiency and preTC stability. We found that Gly had the slowest release rate, and
383 Tyr had the fastest release rate. This difference was statistically significant and correlated with
384 high and low NMD activity, respectively (**Fig. 5E**). Finally, the release rates observed via the
385 TermiLuc assay were highly correlated with the values obtained in the prokaryotic system by
386 Pierson *et al.* (**Fig. 5F**; $r_{RF1} = 0.98$; $r_{RF2} = 0.94$) (Pierson et al., 2016). These data show that peptide
387 release rates indeed vary based on the identity of the last amino acid in a eukaryotic system, with
388 glycine showing the slowest termination rate and tyrosine the fastest, among the amino acids
389 tested.



390
391 **Figure 5. Peptide release rate in an *in vitro* eukaryotic system varies with the identity of the amino**
392 **acid upstream of the stop codon. (A)** Schematic of the Termi-Luc peptide release assay. This assay
393 was performed for nanoluciferase reporters with different amino acids -1 to the stop codon. **(B)** Tested *in*

394 *vitro* luminescence activity of recombinant NLuc variants. **(C)** Translation efficiency of each reporter in the
395 *in vitro* system. **(D)** Relative rate of peptide release when no release factors are added to the termination
396 reaction i.e., preTC instability. **(E)** Relative peptide release rates for each reporter **(F)** A comparison of
397 Termi-Luc determined eukaryotic release rates with Pierson et al prokaryotic release rates (Pierson et al.,
398 2016). "r" refers to the Pearson's correlation coefficient and "p" is the corresponding P-value.

399

400 Taken together, our data supports a model where a slower peptide release rate offers a longer
401 kinetic window for NMD factors to assemble and degrade a transcript, thus causing highly efficient
402 NMD (**Fig. 6**).

403
404

405 **DISCUSSION**

406 NMD is often considered a binary outcome that relies on the presence of triggering factors
407 downstream of the PTC. While recent results increasingly suggest that NMD is tunable
408 (Jagannathan & Bradley, 2016; Lindeboom et al., 2019), the factors influencing this variability are
409 incompletely understood. Our data support a window of opportunity model where the time of
410 residence of the terminating ribosome, which serves as the platform for the assembly of NMD
411 machinery, determines the effectiveness of NMD on a given transcript (**Fig. 6**). Consistent with
412 this model, a slower peptide release rate enhances NMD activity, while a faster rate leads to less
413 efficient NMD. Together with the RBP cues that trigger NMD, termination kinetics are a key
414 modifier of the probability with which a prematurely terminating ribosome leads to NMD.

415 The key finding that led us to the window of opportunity model was the strong enrichment
416 of glycine in the position immediately before the PTC in prematurely terminating transcripts. While
417 Gly codons are the most enriched before a PTC, they are the least enriched before an NTC. Even
418 further, PTCs with Gly-PTC contexts are enriched among variants that are tolerant to loss of
419 function in healthy individuals, while they are not enriched among disease-causing variants.
420 These enrichment patterns became logically coherent once we showed that Gly-PTC induces the
421 most efficient NMD among all possible amino acid contexts and does so through its slow peptide
422 release rate from the terminating ribosome. Interestingly, Gly is also a C-terminal degron,
423 meaning that Gly-end proteins, truncated or otherwise, are more efficiently cleared (Koren et al.,
424 2018; Lin et al., 2018). Thus, the preponderance of glycine codons at the ends of truncated
425 proteins may be the result of the co-evolution of PTCs alongside processes that efficiently
426 eliminate Gly-end proteins at both the RNA and protein levels.

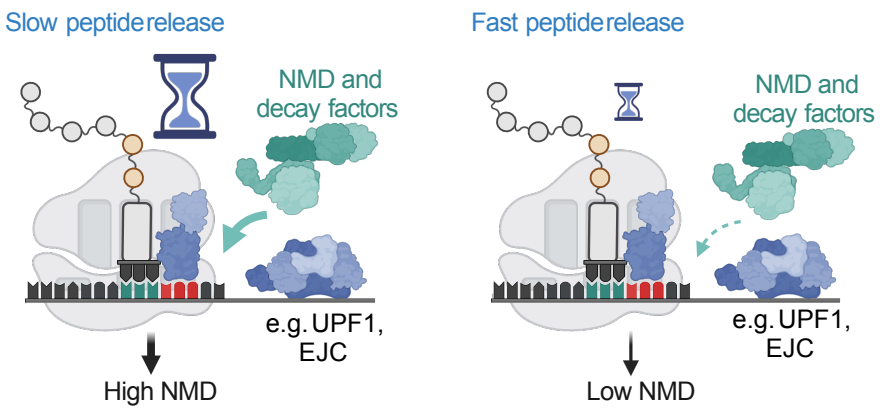


Figure 6. A kinetic model for the variability in NMD activity dependent upon the codon preceding the PTC. A window of opportunity model for how peptide release rates influence NMD activity. A slower peptide release rate gives NMD factors a larger window to bind an act on a target, and vice versa.

Other amino acids having patterns similar 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 support efficient NMD and exhibit slow peptide release rates. On the other end of the spectrum, Tyr had the lowest NMD activity in our experiments and was enriched before NTCs but not PTCs; Tyr also has the fastest peptide release rate. PTCs with Tyr at the -1 of a PTC are depleted among variants tolerant to loss of function and enriched in disease contexts.

Apart from the -1 amino acid position, our massively parallel assessment of NMD activity allowed us to uncover other local sequence preferences for efficient NMD. Surprisingly, we found that the stop codon TGA caused 20% more efficient EJC-independent NMD than TAG. TGA 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 highest probability of readthrough, which usually stabilizes an NMD target (Tate & Mannering, 1996). Our results suggest that there are kinetic differences in termination at different stop codons. Additional factors, such as RBPs or other sequence elements, may influence whether a stop codon is read through or whether it causes efficient NMD.

Our study highlights the power of human genetics to discover novel molecular mechanisms that can explain the real-world variability observed in the expression of genes containing nonsense variants. Our results point to a key role for the kinetics of translation termination in determining NMD activity, ultimately determining the efficiency with which prematurely terminating transcripts are cleared from the cell. Future work will be required to determine whether the influence of translation termination kinetics may also extend to other translation-associated quality

453 control processes. The impact of amino acid identity upstream of the PTC on NMD activity and
454 patterns of amino acid-PTC enrichment in healthy and diseased populations suggest that NMD
455 has played a key role in shaping the evolution of nonsense variants. These findings can be used
456 to better model the phenotypic outcomes of PTCs, and thus develop more effective strategies to
457 counteract PTC-associated diseases.

458

459

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 Monkol Lek and Sander Pajusalu for helpful discussions on gnomAD analysis. 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.

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 (GSE261151) and will be publicly available as of publication.
- All original code is available at the GitHub repository: https://github.com/jagannathan-lab/2024-kolakada_et_al.
- 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).

512 **METHOD DETAILS**

513 **gnomAD and ClinVar analysis**

514 Rare (frequency < 0.01) SNP variants from gnomAD v2.1.1 exome database that pass
515 random forest classification quality filtering were analyzed, filtering for "Stop-gained"
516 consequence via bcftools. Using "cDNA_position" field pre-annotated by VEP, the sequence and
517 codons surrounding variants from Gencode v19 transcript FASTA file was then extracted using
518 bedtools getfasta. For other contexts such as normal stop codons, custom code (found in:
519 https://github.com/jagannathan-lab/2024-kolakada_et_al) was used to extract sequence and
520 codon information using Biostrings and BSgenome in R. Codon enrichment is calculated as
521 observed frequency of a particular codon/amino acid normalized against normal occurrence
522 frequency of the codon/amino acid in the last 10 positions of all normal coding genes, similar to
523 Koren et al (Koren et al., 2018). Gene-level loss of function scores were retrieved from gnomAD
524 LoF constraint scores. Clinvar data from 2022/07/30 was processed similarly to gnomAD, but with
525 sequence extraction done in R.

526 **Cloning**

527 The NMD reporter plasmids have been previously described (Kolakada et al., 2024). For
528 individual fluorescent and luminescent reporters containing Gly, Val, Ala, and Tyr amino acids
529 before the PTC, oligos containing these sequences (Integrated DNA Technologies) and matching
530 the EcoRI (NEB, R3101S) and XhoI (NEB, R0146S) restriction sites were synthesized. These
531 oligos were annealed and ligated using the Quick Ligation Kit (NEB, M2200L) into EJC-
532 independent and EJC-enhanced fluorescent and luminescent backbones digested with EcoRI
533 and XhoI. To make the NMD– reporters for each sequence, each EJC-independent reporter was
534 digested with EcoRI and MfeI (NEB, R3589S), the sticky ends were filled in using Klenow
535 Fragment (NEB, M0212S) and the blunt ends were ligated together using the Quick Ligation Kit
536 (NEB), removing the GFP 3′ UTR. The oligos synthesized for each sequence tested are as
537 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

538

539 **MPRA plasmid library**

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

551 The EJC-independent and EJC-enhanced fluorescent reporter backbones were linearized
552 using EcoRI (NEB) and XhoI (NEB) at room temperature overnight. Digested plasmid DNA was
553 gel purified using the NucleoSpin Gel and PCR Clean-up kit (Takara Bio, 740609.250). The oligo
554 library was then cloned into the digested backbones using the Gibson Assembly Master Mix (NEB,
555 E2611L) using an insert:vector molar ratio of 7:1 for the EJC-independent backbone and 3:1 for
556 the EJC-enhanced backbone. The reactions were incubated at 50°C for 1 h. The reactions were
557 then purified using AMPure XP beads to remove excess salts and a total of 200 ng of plasmid
558 DNA was transformed into MegaX DH10B T1R Electrocompetent Cells (ThermoFisher,
559 C640003), using a Biorad GenePulser electroporator. The transformed cells were recovered in
560 the recovery medium provided with the electrocompetent cells at 37°C for 1 h, then plated on 33
561 pre-warmed 15 cm Luria broth (LB) agar-Carbenicillin (RPI, C46000-5.0) plates and incubated at
562 37°C overnight. The next day colonies were collected by scraping plates, spun-down, and then
563 midi-prepped (NucleoBond Xtra Midi EF kit, 740420.50) to extract the plasmid DNA.

564

565 **Plasmid transfections**

566 Transfections for fluorescent and luminescent reporters were performed differently since
567 fluorescent proteins have longer maturation rates than luciferase proteins (Mazo-Vargas et al.,
568 2014). For the fluorescent reporters, cells were split 24 h after transfection equally into two new
569 6-wells. Forty-eight hours after transfection, one set of cells was harvested for RNA and the other
570 set for flow cytometry. For the luminescent reporters, transfections were performed in 24-well
571 plates at a 40% confluency with 500 ng of plasmid DNA using Lipofectamine 2000 (ThermoFisher
572 Scientific, 11668027), as per manufacturer's instructions. All cells were harvested for luciferase
573 assays.

574 **Generating NMD+ reporter cell lines for the MPRA**

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

586 **Targeted sequencing of MPRA libraries**

587 To obtain MPRA sequencing libraries, frozen EJC-independent and EJC-enhanced
588 reporter cell lines were thawed. These cells were split into 6, 15 cm plates, 3 of which were treated
589 with DMSO and the other 3 treated with 0.5 μM SMG1i (technical replicates). These plates were
590 harvested for RNA in TRIzol Reagent (ThermoFisher Scientific, 15596018) after 24 hrs. TRIzol
591 extractions were performed for RNA as per manufacturer's instructions. A DNase digest was
592 conducted on 10 μg of RNA from each sample using the Turbo DNA-free kit (ThermoFisher
593 Scientific, AM1907), including reaction inactivation. Following this, 2 μg of digested RNA was
594 used to make cDNA using Superscript II Reverse Transcriptase in duplicate 20 μL reactions
595 (ThermoFisher Scientific, 18064014).

596 For library preparation, each cDNA sample was split into 20 PCR reactions (2 μL
597 cDNA/PCR) and amplified using reporter-specific forward and reverse primers with UMIs (forward

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

609 **MPRA data analysis**

610 To process the MPRA sequencing libraries, the UMI-tools repository was used to add the
611 UMIs to the name of each read in the FASTA files (Smith et al., 2017). Following this, Cutadapt
612 was used to trim the 5′ and 3′ends of the reads in a sequence dependent manner, up to the 10 nt
613 context of interest (Martin, 2011); ~25% of reads were lost in this process. SeqPrep was used to
614 merge sequencing reads 1 and 2, only accepting reads that perfectly align
615 (https://github.com/jstjohn/SeqPrep); ~5% of reads were lost in this process. Bowtie was used to
616 map the reads to a reference file of all sequence contexts (Langmead & Salzberg, 2012); 1% of
617 reads were lost in this process. Samtools was used to create a BAM file of mapped reads, followed
618 by deduplicating the BAM files based on UMIs and converting the BAM files to BED files (Li et al.,
619 2009). Around 10% of reads were lost upon deduplication of UMIs. A custom python script was
620 used to remove reads on the negative strand and to get a bed file of reads in the correct
621 orientation.

622 The NMD activity of an individual reporter transcript was the log2-fold change in between
623 DMSO-treated cells to SMG1i-treated cells calculated with DESeq2 (Love et al., 2014). The
624 reporter transcripts that do not undergo NMD, i.e. Trp codon rather than a stop codon at the PTC
625 position, were used similar to housekeeping normalization controls within DESeq2. Both linear
626 models and random forest classifiers were performed in R. Random Forest models were built
627 utilizing the caret R package (doi:10.18637/jss.v028.i05). First, features with variance close to
628 zero were eliminated, after which the linear correlation for all numeric features was calculated.
629 For the features with correlation ≥ 0.85 one feature was randomly selected, and the others
630 discarded from the model. Data were divided into training and validation set using 0.8:0.2 ratio.

631 Hyperparameters tuning was performed for train set using 5x train-test cross validation for the
632 following model parameters: mtry, maxnodes and ntree. Performance of final model was
633 measured using F1 scores and on validation set. The code and packages used for these analyses
634 can be found in this repository: https://github.com/jagannathan-lab/2024-kolakada_et_al.

635 **Luciferase assays**

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

644 **Lentiviral transductions**

645 Knockdowns were conducted using shRNAs. Plasmids containing shRNAs were first
646 packaged into lentivirus. To do this, HEK293T cells were transfected with 4 µg of the relevant
647 shRNA vector (Functional Genomics Core, CU Anschutz), 3 µg of psPAX2 (Addgene plasmid
648 #12260), and 1 µg of pMD2.G (Addgene plasmid #12259), in 80% confluent 6 cm plates. The
649 shRNAs used were for scrambled (SCH016), UPF1 (TRCN0000022254), UPF2
650 (TRCN0000151381), UPF3B (TRCN0000152769), SMG1 (TRCN0000194827**)**, SMG6
651 (TRCN0000040014), ZNF598 (TRCN0000073159), EIF5A (TRCN0000062552), EIF3J
652 (TRCN0000062013), CUL2 (TRCN0000006523), and ABCE1 (TRCN0000158417). Twenty-four
653 hours after transfection the media was replaced with 3 mL of fresh media. Forty-eight hours after
654 transfection media containing lentivirus was harvested and spun down at 1200 rpm for 5 minutes.
655 Aliquots of the supernatant were frozen at -80°C for later use.

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

662 **RNA extraction and RT-qPCR**

663 For RT-qPCR used to validate knockdowns of SMG1, ZNF598, and EIF3J, RNA was
664 extracted from cells using TRIzol Reagent (ThermoFisher Scientific). DNase digestion was
665 performed with DNaseI (ThermoFisher Scientific, 18068-015) on 1 µg of extracted RNA. This was
666 followed by cDNA synthesis using the SuperScript III First-Strand Synthesis kit, using random
667 hexamers. A no-RT sample was included as a control to make sure there was no genomic DNA
668 contamination. The cDNA was then diluted 1:4 and 2 µL were used per 10 µL qPCR reaction.
669 qPCR was performed using iTaq Universal SYBR Green Supermix and primers specific for
670 SMG1, ZNF598, EIF3J, and the housekeeping gene RPL27, were used at a final concentration
671 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

672 Reactions were set up in triplicate per sample and plated in 384-well plates. The plates
673 were then run on the OPUS Bio-Rad qPCR machine (Bio-Rad, 12011319) using the 2-step
674 Amplification and melting curve protocol. For the knockdown validations, the Livak method (Livak
675 & Schmittgen, 2001) was used to quantify differences in RNA levels: SMG1, ZNF598, EIF3J, were
676 normalized to RPL27. Mean RNA levels of 3 different transfections or 3 technical replicates of
677 transductions were plotted, respectively. Error was calculated using standard error of the mean
678 (fluorescent/luminescent reporters) or standard deviation (knockdowns).

679

680 **Protein isolation and western blotting**

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

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

698 **Luminescent reporters for *in vitro* assays**

699 A pNL-globine vector derived from pNL1.1 vector (Promega), with a β-globin 5′UTR
700 addition before nanoluciferase (NLuc) coding sequence was used (Shuvalov et al., 2021).
701 Additional constructs were created based on pNL-globin, encoding NLuc with various
702 substitutions of the codon before the stop codon using the QuikChange Site-Directed
703 Mutagenesis Kit (Agilent Technologies, cat. 200518-5) was used. Primers were selected in the
704 manufacturer's recommended web-based QuikChange Primer Design Program
705 (www.agilent.com/genomics/qcpd). For obtaining mRNA, the fragment of the plasmid were
706 amplified using RV3L (CTAGCAAAATAGGCTGTCCCCAG) and FLA50
707 (TTTT

721 His-tagged Ulp1 protease. Untagged eRF1 was further purified by anion-exchange
722 chromatography (HiTrap Q HP, Cytiva). Fractions enriched with eRF1 were collected, dialyzed in
723 storage buffer, frizzed in liquid nitrogen, and stored at −70 ºC. Recombinant eRF3A cloned into
724 baculovirus vector EMBacY from a MultiBac expression system was expressed in the insect cell
725 line Sf21. Following this, recombinant proteins were purified using Ni-NTA agarose and ion-
726 exchange chromatography, as described previously (Ivanov et al., 2016).

727 **Expression, purification, and determination of luminescence or recombinant**
728 **nanoluciferase**

729 To obtain recombinant NLuc with varying amino acids in the position -1 to the stop codon,
730 corresponding constructs of pNL-globine were created using the petSUMO vector (Invitrogen, cat.
731 K30001). Each vector was assembled via Gibson Assembly (NEB, cat. E2611L) using PCR
732 products of the petSUMO backbone (forward primer:
733 AGCTTAGGTATTTATTCGGCGCAAAGTG; reverse primer:
734 ACCACCAATCTGTTCTCTGTGAGC) and pNL-globine inserts (forward primer:
735 GAGAACAGATTGGTGGTATGGTCTTCACACTCGAAGATTTCGTTGG; reverse primer:
736 CCGAATAAATACCTAAGCTTACGCCAGAATGCGTTCGCA) amplified using Q5 High-Fidelity
737 DNA Polymerase (NEB, cat. M0491L). The resulting plasmids were expressed in *E. coli*
738 BL21(DE3) and His-SUMO-Nluc was purified using Ni-affinity chromatography. Recombinant
739 6xHis-ULP1 protease was added to purify 6xHisSUMO-Nluc and incubated for 2 hours. The
740 6xHisSUMO fragment and 6xHis-ULP1 were then removed with Ni-NTA agarose to obtain purified
741 Nluc without tags. 0.1 femtomole of recombinant Nluc was incubated in storage buffer, with the
742 addition of BSA up to 500 µg/mL and 0.5 % NanoGlo (Promega). Luminescence was measured
743 at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) in a 40 min period. The
744 luminescence of NLuc was calculated as a maximum of relative luminescence units (RLU_{max}).

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

746 For the Termi-Luc assay, preTCs translating NLuc (preTC-NLuc) were purified using
747 previously published methods (Shuvalov et al., 2021; Susorov et al., 2020). 100% RRL lysate
748 was preincubated in a mixture containing 1 mM $CaCl_2$ and 3 U/µL Micrococcal nuclease
749 (Fermentas) at 30°C for 10 min, followed by the addition of EGTA to a final concentration of 4
750 mM. The lysate was then diluted to 70% (v/v) and supplemented with 20 mM HEPES-KOH (pH
751 7.5), 80 mM KOAc, 0.5 mM $Mg(OAc)_2$, 0.3 mM ATP, 0.2 GTP, 0.04 mM of each of 20 amino acids
752 (Promega), 0.5 mM spermidine, 0.45 µM aminoacylated total rabbit tRNA, 10 mM creatine

753 phosphate, 0.003 U/μL creatine kinase (Sigma), 2 mM DTT, and 0.2 U/μL Ribolock
754 (ThermoFisher) (70% RRL mix).

755 For preTC-NLuc assembly, 220 μL of 70% RRL mix was preincubated in the presence of
756 1.7 μM ERF1 G183A mutant (Alkalaeva et al., 2006) at 30° C for 10 min, followed by the addition
757 of 10.5 pmol of NLuc mRNA. The mixture was incubated at 30°C for 40 min. The KOAc
758 concentration was then adjusted to 300 mM and the mixture was layered on 5 ml of a 10–35%
759 linear sucrose gradient in a buffer containing 50 mM HEPES-KOH, pH 7.5, 7.5 mM $Mg(OAc)_2$,
760 300 mM KOAc, 2 mM DTT. The gradient was centrifuged in a SW55-Ti (Beckman Coulter) rotor
761 at 55 000 rpm (367 598 g_{max}) for 1 h. The gradient was fractionated in 15 fractions of 150 μL from
762 bottom to top and the remaining sucrose was collected separately. Fractions of the first peak,
763 enriched with the preTC-Nluc were analyzed by the Termi-Luc assay, merged, flash-frozen in
764 liquid nitrogen, and stored at −70°C.

765 A peptide release assay with the preTC-NLuc (Termi-Luc) was performed as previously
766 described with some modifications (Susorov et al., 2020). Peptide release was conducted in a
767 solution containing 1.5 pM preTC-Nluc, 45 mM HEPES-KOH pH 7.5, 1.4 mM Mg_2OAc, 56 mM
768 KOAc pH 7.0, 1 mM DTT, 177 μM spermidine, 1.5 % (w/w) sucrose, 0.8 mM $MgCl_2$, 0.2 mM GTP
769 supplemented with equimolar $MgCl_2$, and 0.5 % NanoGlo (Promega), in the absence or the
770 presence of release factors eRF1 and eRF3a at various concentrations. Luminescence was
771 measured at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) for 12 min.
772 The translation efficiency was calculated as the maximal derivative of the growing linear section
773 of the luminescence curve (v_0, RLU/min). For comparison of different preTCs, a single working
774 concentration of 3 nM of eRF1-eRF3A was chosen in the linear section. The instability of the
775 preTC-NLuc was evaluated in Termi-Luc assay in the absence of release factors.

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

781 **Translation of NLuc in RRL lysates**

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

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

787

788 **QUANTIFICATION AND STATISTICAL ANALYSIS**

789 **Data analysis, statistical tests, and visualization**

790 All experiments were performed with a sample size of at least n=3. For transient
791 transfections, each replicate represents an independent transfection. For stably integrated cells,
792 the same cell line was plated in 3 different wells, from which RNA and protein were harvested.
793 Statistical significance between various populations was calculated using a student's t-test in R
794 and p-values were two-sided (*$p<0.05$, **$p<0.005$, ***$p<0.0005$). Statistical details of specific
795 experiments can be found in the Results, Methods, and/or Figure Legends. Plots were generated
796 using R plotting functions and/or the ggplot2 package. The exact code and packages used for
797 these analyses can be found in this repository: https://github.com/jagannathan-lab/2024-
798 kolakada_et_al.

799

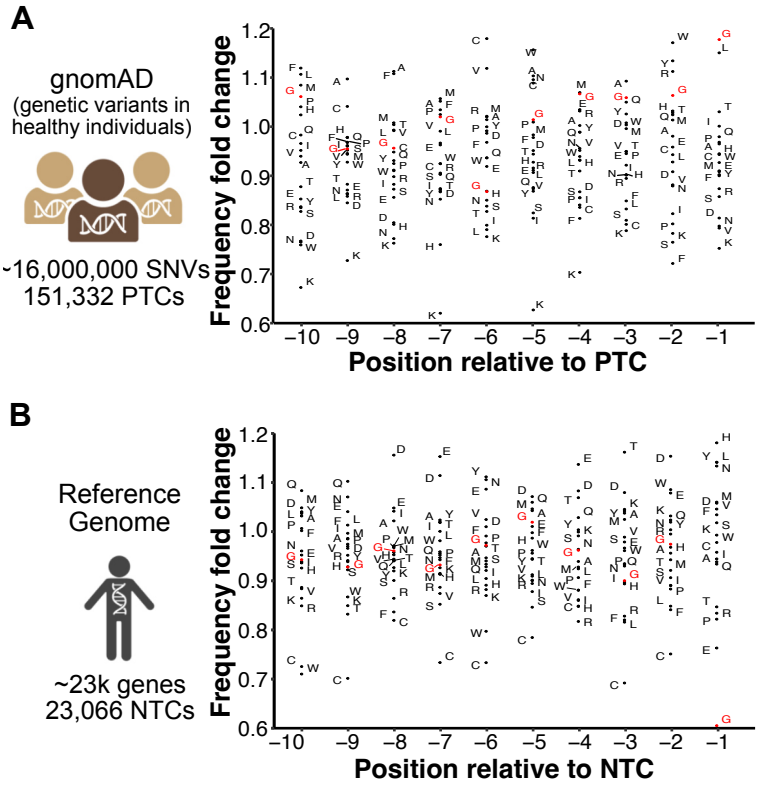
800 **SUPPLEMENTAL INFORMATION**

801

802 **SUPPLEMENTARY FIGURES**

803 **Figure S1**

804



805
806
807

808 **Figure S1. Amino acids enriched from the -10 to the -1 position before a termination codon.**
809 (**A**) Left: Schematic to indicate PTCs analyzed in the gnomAD database, right: amino acid
810 enrichment before the PTC. (**B**) Left: Schematic to indicate NTCs analyzed in the reference
811 genome, right: amino acid enrichment before the NTC. Gly is colored in red.
812

Figure S2

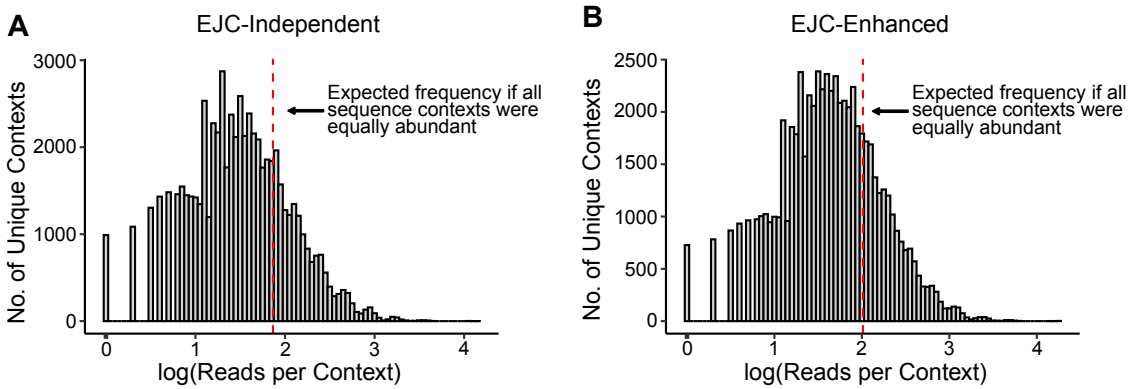
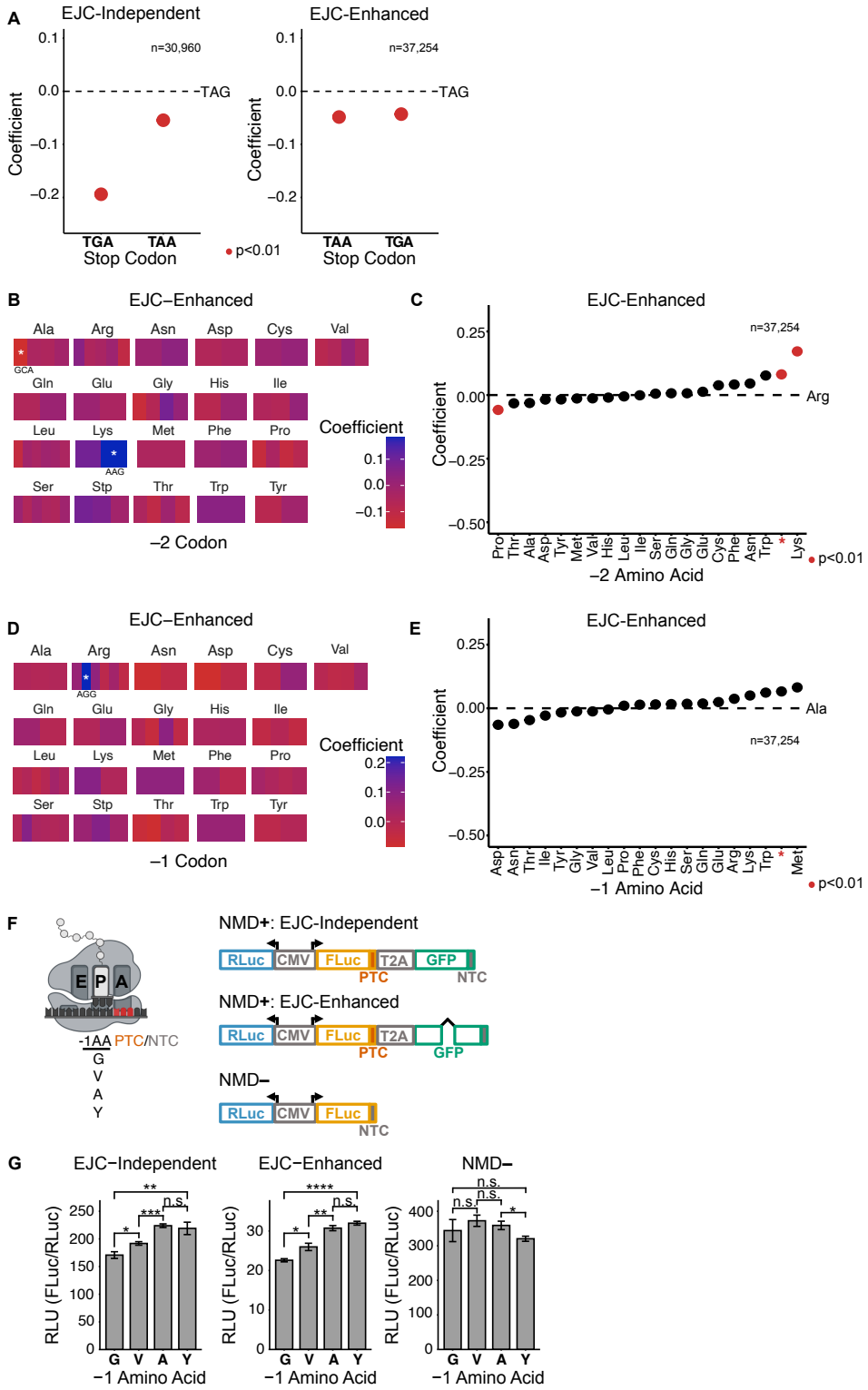


Figure S2. MPRA library representation. (**A**) Plasmid library representation for the EJC-independent NMD**+** reporter library. (**B**) The same as (A) but for the EJC-enhanced NMD**+** reporter library.

835 **Figure S3**



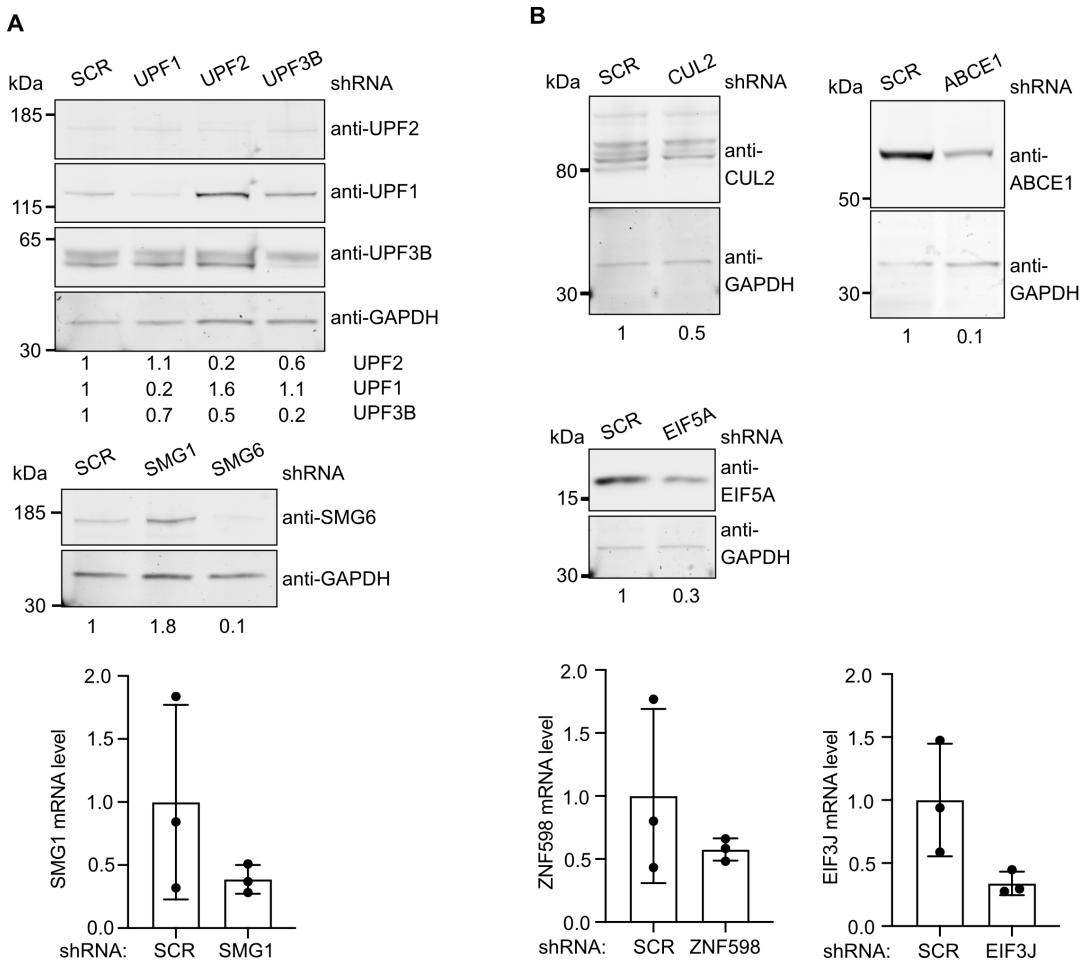
836

837 **Figure S3. The codon identity at the -1 position of the PTC alters NMD activity of EJC-**
838 **enhanced luminescence-based reporters.** (**A**) Effect of stop codon identity on EJC-

839 independent and EJC-enhanced NMD without TGG. (**B-E**) Plots depicting a variable's effect on
840 EJC-enhanced NMD: (**B**) heat map for the -2 codon, (**C**) dot plot for the -2 amino acid, (**D**) heat
841 map for them -1 codon, and (**E**) dot plot for the -1 amino acid. (**F**) Schematic of luminescence
842 based NMD reporters used (right) as well as the amino acids tested in the -1 position to the
843 PTC/NTC (left). (**G**) Luciferase activity for each of the reporters in EJC-independent, EJC-
844 enhanced, and NMD– contexts.

845

846 **Figure S4**



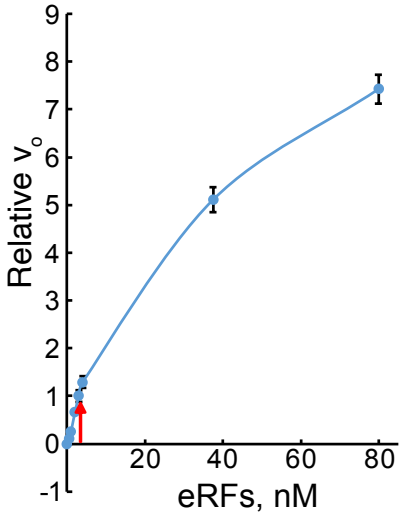
847

848 **Figure S4. Knockdown validations for all shRNAs used.** (**A**) Top: western blot validating
849 UPF1, UPF2, and UPF3B knockdowns; middle: western blot validating SMG6 knockdown;
850 bottom: qPCR validation of SMG1 knockdown. (**B**) Top left: western blot for CUL2 knockdown
851 validation; top right: western blot for ABCE1 knockdown validation; middle left: western blot for
852 EIF5A knockdown validation; bottom left: qPCR validation of ZNF598 knockdown; bottom right:
853 qPCR validation for EIF3J knockdown. GAPDH was the loading control for all western blots.
854 RPL27 was the housekeeping gene for all qPCRs. qPCRs were performed in triplicate.

855

856

857 **Figure S5.**



858

859 **Figure S5. Determining the optimum concentration of release factors to use for the Termi-**
860 **Luc assay.**

861

862

863 **SUPPLEMENTARY TABLES**

864 **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

865

866 **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

867 **SUPPLEMENTARY DATA**

868

869 **Supplementary Data Table 1.** DEseq2 results for the EJC-independent library of sequences
870 representing the fold change of DMSO over SMG1i treatment of cells.
871
872 **Supplementary Data Table 2.** DEseq2 results for the EJC-enhanced library of sequences
873 representing the fold change of DMSO over SMG1i treatment of cells.

874 **REFERENCES**

875

876 Alkalaeva, E. Z., Pisarev, A. V., Frolova, L. Y., Kisselev, L. L., & Pestova, T. V. (2006, Jun 16).
877 In vitro reconstitution of eukaryotic translation reveals cooperativity between release
878 factors eRF1 and eRF3. *Cell, 125*(6), 1125-1136.
879 https://doi.org/10.1016/j.cell.2006.04.035
880

881 Arias-Palomo, E., Yamashita, A., Fernandez, I. S., Nunez-Ramirez, R., Bamba, Y., Izumi, N.,
882 Ohno, S., & Llorca, O. (2011, Jan 15). The nonsense-mediated mRNA decay SMG-1
883 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes*
884 *Dev, 25*(2), 153-164. https://doi.org/10.1101/gad.606911
885

886 Baker, S. L., & Hogg, J. R. (2017). A system for coordinated analysis of translational
887 readthrough and nonsense-mediated mRNA decay | PLOS ONE. *PLoS ONE, 12(3):*
888 *e0173980*. https://doi.org/10.1371/journal.pone.0173980
889

890 Bateman, J. F., Freddi, S., Nattrass, G., & Savarirayan, R. (2003, Feb 1). Tissue-specific RNA
891 surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in
892 Schmid metaphyseal chondrodysplasia cartilage. *Hum Mol Genet, 12*(3), 217-225.
893 https://www.ncbi.nlm.nih.gov/pubmed/12554676
894

895 Bazzini, A. A., Del Viso, F., Moreno-Mateos, M. A., Johnstone, T. G., Vejnar, C. E., Qin, Y., Yao,
896 J., Khokha, M. K., & Giraldez, A. J. (2016, Oct 4). Codon identity regulates mRNA
897 stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J,*
898 *35*(19), 2087-2103. https://doi.org/10.15252/embj.201694699
899

900 Bonetti, B., Fu, L., Moon, J., & Bedwell, D. M. (1995, Aug 18). The efficiency of translation
901 termination is determined by a synergistic interplay between upstream and downstream
902 sequences in Saccharomyces cerevisiae. *J Mol Biol, 251*(3), 334-345.
903 https://doi.org/10.1006/jmbi.1995.0438

Buhler, M., Steiner, S., Mohn, F., Paillusson, A., & Muhlemann, O. (2006, May). EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat Struct Mol Biol, 13*(5), 462-464. https://doi.org/10.1038/nsmb1081

Chang, J. C., & Kan, Y. W. (1979, Jun). beta 0 thalassemia, a nonsense mutation in man. *Proc Natl Acad Sci U S A, 76*(6), 2886-2889. https://www.ncbi.nlm.nih.gov/pubmed/88735 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC383714/pdf/pnas00006-0406.pdf

Chu, V., Feng, Q., Lim, Y., & Shao, S. (2021, Dec 1). Selective destabilization of polypeptides synthesized from NMD-targeted transcripts. *Mol Biol Cell, 32*(22), ar38. https://doi.org/10.1091/mbc.E21-08-0382

Cridge, A. G., Crowe-McAuliffe, C., Mathew, S. F., & Tate, W. P. (2018, Feb 28). Eukaryotic translational termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA channel. *Nucleic Acids Res, 46*(4), 1927-1944. https://doi.org/10.1093/nar/gkx1315

Egorova, T., Biziaev, N., Shuvalov, A., Sokolova, E., Mukba, S., Evmenov, K., Zotova, M., Kushchenko, A., Shuvalova, E., & Alkalaeva, E. (2021, Nov 8). eIF3j facilitates loading of release factors into the ribosome. *Nucleic Acids Res, 49*(19), 11181-11196. https://doi.org/10.1093/nar/gkab854

Frolova, L., Seit-Nebi, A., & Kisselev, L. (2002, Feb). Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA, 8*(2), 129-136. https://doi.org/10.1017/s1355838202013262

931 Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V.
932 M., Tatkov, S. I., & Kisselev, L. L. (1999, Aug). Mutations in the highly conserved GGQ
933 motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger
934 peptidyl-tRNA hydrolysis. *RNA, 5*(8), 1014-1020.
935 https://doi.org/10.1017/s135583829999043x
936

937 Grosjean, H., & Westhof, E. (2016, Sep 30). An integrated, structure- and energy-based view of
938 the genetic code. *Nucleic Acids Res, 44*(17), 8020-8040.
939 https://doi.org/10.1093/nar/gkw608
940

941 Hellen, C. U. T. (2018, Oct 1). Translation Termination and Ribosome Recycling in Eukaryotes.
942 *Cold Spring Harb Perspect Biol, 10*(10). https://doi.org/10.1101/cshperspect.a032656
943

944 Hogg, J. R., & Goff, S. P. (2010, Oct 29). Upf1 senses 3'UTR length to potentiate mRNA decay.
945 *Cell, 143*(3), 379-389. https://doi.org/10.1016/j.cell.2010.10.005
946

947 Inacio, A., Silva, A. L., Pinto, J., Ji, X., Morgado, A., Almeida, F., Faustino, P., Lavinha, J.,
948 Liebhaber, S. A., & Romao, L. (2004, Jul 30). Nonsense mutations in close proximity to
949 the initiation codon fail to trigger full nonsense-mediated mRNA decay. *J Biol Chem,*
950 *279*(31), 32170-32180. https://doi.org/10.1074/jbc.M405024200
951

952 Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A.,
953 Schaffitzel, C., & Alkalaeva, E. (2016, Sep 19). PABP enhances release factor
954 recruitment and stop codon recognition during translation termination. *Nucleic Acids*
955 *Res, 44*(16), 7766-7776. https://doi.org/10.1093/nar/gkw635
956

Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2012). Termination and post-termination events in eukaryotic translation. *Adv Protein Chem Struct Biol, 86*, 45-93. https://doi.org/10.1016/B978-0-12-386497-0.00002-5

Jagannathan, S., & Bradley, R. K. (2016, Dec). Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Res, 26*(12), 1639-1650. https://doi.org/10.1101/gr.205070.116

Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V., & Hegde, R. S. (2018, Nov 1). ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Mol Cell, 72*(3), 469-481 e467. https://doi.org/10.1016/j.molcel.2018.08.037

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., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome Aggregation Database, C., Neale, B. M., Daly, M. J., & MacArthur, D. G. (2020, May). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581*(7809), 434-443. https://doi.org/10.1038/s41586-020-2308-7

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., & Ohno, S. (2006, Feb 1). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and

986 nonsense-mediated mRNA decay. *Genes Dev, 20*(3), 355-367.
987 https://doi.org/10.1101/gad.1389006
988

989 Kazazian, H. H., Jr., Dowling, C. E., Hurwitz, R. L., Coleman, M., Stopeck, A., & Adams, J. G.,
990 3rd. (1992, Jun 1). Dominant thalassemia-like phenotypes associated with mutations in
991 exon 3 of the beta-globin gene. *Blood, 79*(11), 3014-3018.
992 https://www.ncbi.nlm.nih.gov/pubmed/1586746
993 http://www.bloodjournal.org/content/bloodjournal/79/11/3014.full.pdf
994

995 Keeling, K. M., Lanier, J., Du, M., Salas-Marco, J., Gao, L., Kaenjak-Angeletti, A., & Bedwell, D.
996 M. (2004, Apr). Leaky termination at premature stop codons antagonizes nonsense-
997 mediated mRNA decay in S. cerevisiae. *RNA, 10*(4), 691-703.
998 https://www.ncbi.nlm.nih.gov/pubmed/15037778
999

1000 Kerr, T. P., Sewry, C. A., Robb, S. A., & Roberts, R. G. (2001, Oct). Long mutant dystrophins
1001 and variable phenotypes: evasion of nonsense-mediated decay? *Hum Genet, 109*(4),
1002 402-407. https://doi.org/10.1007/s004390100598
1003

1004 Khandelia, P., Yap, K., & Makeyev, E. V. (2011, Aug 2). Streamlined platform for short hairpin
1005 RNA interference and transgenesis in cultured mammalian cells. *Proc Natl Acad Sci U S*
1006 *A, 108*(31), 12799-12804. https://doi.org/10.1073/pnas.1103532108
1007

1008 Kolakada, D., Campbell, A. E., Galvis, L. B., Li, Z., Lore, M., & Jagannathan, S. (2024, Feb 20).
1009 A system of reporters for comparative investigation of EJC-independent and EJC-
1010 enhanced nonsense-mediated mRNA decay. *Nucleic Acids Res*.
1011 https://doi.org/10.1093/nar/gkae121
1012

1013 Koren, I., Timms, R. T., Kula, T., Xu, Q., Li, M. Z., & Elledge, S. J. (2018, Jun 14). The
1014 Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons.
1015 *Cell, 173*(7), 1622-1635 e1614. https://doi.org/10.1016/j.cell.2018.04.028
1016

1017 Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J.,
1018 Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A.,
1019 McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., Kattman, B. L., & Maglott,
1020 D. R. (2018, Jan 4). ClinVar: improving access to variant interpretations and supporting
1021 evidence. *Nucleic Acids Res, 46*(D1), D1062-D1067.
1022 https://doi.org/10.1093/nar/gkx1153
1023

1024 Langmead, B., & Salzberg, S. L. (2012, Mar 4). Fast gapped-read alignment with Bowtie 2. *Nat*
1025 *Methods, 9*(4), 357-359. https://doi.org/10.1038/nmeth.1923
1026

1027 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1028 Durbin, R., & Genome Project Data Processing, S. (2009, Aug 15). The Sequence
1029 Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.
1030 https://doi.org/10.1093/bioinformatics/btp352
1031

1032 Lin, H. C., Yeh, C. W., Chen, Y. F., Lee, T. T., Hsieh, P. Y., Rusnac, D. V., Lin, S. Y., Elledge,
1033 S. J., Zheng, N., & Yen, H. S. (2018, May 17). C-Terminal End-Directed Protein
1034 Elimination by CRL2 Ubiquitin Ligases. *Mol Cell, 70*(4), 602-613 e603.
1035 https://doi.org/10.1016/j.molcel.2018.04.006
1036

1037 Linde, L., Boelz, S., Neu-Yilik, G., Kulozik, A. E., & Kerem, B. (2007, Nov). The efficiency of
1038 nonsense-mediated mRNA decay is an inherent character and varies among different
1039 cells. *Eur J Hum Genet, 15*(11), 1156-1162. https://doi.org/10.1038/sj.ejhg.5201889
1040

1041 Lindeboom, R. G., Supek, F., & Lehner, B. (2016, Oct). The rules and impact of nonsense-
1042 mediated mRNA decay in human cancers. *Nat Genet, 48*(10), 1112-1118.
1043 https://doi.org/10.1038/ng.3664
1044

1045 Lindeboom, R. G. H., Vermeulen, M., Lehner, B., & Supek, F. (2019, Nov). The impact of
1046 nonsense-mediated mRNA decay on genetic disease, gene editing and cancer
1047 immunotherapy. *Nat Genet, 51*(11), 1645-1651. https://doi.org/10.1038/s41588-019-
1048 0517-5
1049

1050 Livak, K. J., & Schmittgen, T. D. (2001, Dec). Analysis of relative gene expression data using
1051 real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-
1052 408. https://doi.org/10.1006/meth.2001.1262
1053

1054 Loughran, G., Chou, M. Y., Ivanov, I. P., Jungreis, I., Kellis, M., Kiran, A. M., Baranov, P. V., &
1055 Atkins, J. F. (2014, Aug). Evidence of efficient stop codon readthrough in four
1056 mammalian genes. *Nucleic Acids Res, 42*(14), 8928-8938.
1057 https://doi.org/10.1093/nar/gku608
1058

1059 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
1060 dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550.
1061 https://doi.org/10.1186/s13059-014-0550-8
1062

1063 MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins,
1064 L., Habegger, L., Pickrell, J. K., Montgomery, S. B., Albers, C. A., Zhang, Z. D., Conrad,
1065 D. F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M. A., Banks, E., Hu, M., Handsaker, R.
1066 E., Rosenfeld, J. A., Fromer, M., Jin, M., Mu, X. J., Khurana, E., Ye, K., Kay, M.,
1067 Saunders, G. I., Suner, M. M., Hunt, T., Barnes, I. H., Amid, C., Carvalho-Silva, D. R.,
1068 Bignell, A. H., Snow, C., Yngvadottir, B., Bumpstead, S., Cooper, D. N., Xue, Y.,
1069 Romero, I. G., Genomes Project, C., Wang, J., Li, Y., Gibbs, R. A., McCarroll, S. A.,

1070 Dermitzakis, E. T., Pritchard, J. K., Barrett, J. C., Harrow, J., Hurles, M. E., Gerstein, M.
1071 B., & Tyler-Smith, C. (2012, Feb 17). A systematic survey of loss-of-function variants in
1072 human protein-coding genes. *Science, 335*(6070), 823-828.
1073 https://doi.org/10.1126/science.1215040
1074

1075 Mangkalaphiban, K., He, F., Ganesan, R., Wu, C., Baker, R., & Jacobson, A. (2021, Apr).
1076 Transcriptome-wide investigation of stop codon readthrough in Saccharomyces
1077 cerevisiae. *PLoS Genet, 17*(4), e1009538. https://doi.org/10.1371/journal.pgen.1009538
1078

1079 Martin, M. (2011, 2011-05-02). Cutadapt removes adapter sequences from high-throughput
1080 sequencing reads [next generation sequencing; small RNA; microRNA; adapter
1081 removal]. *2011, 17*(1), 3. https://doi.org/10.14806/ej.17.1.200
1082

1083 Mazo-Vargas, A., Park, H., Aydin, M., & Buchler, N. E. (2014, Nov 5). Measuring fast gene
1084 dynamics in single cells with time-lapse luminescence microscopy. *Mol Biol Cell, 25*(22),
1085 3699-3708. https://doi.org/10.1091/mbc.E14-07-1187
1086

1087 McCaughan, K. K., Brown, C. M., Dalphin, M. E., Berry, M. J., & Tate, W. P. (1995, Jun 6).
1088 Translational termination efficiency in mammals is influenced by the base following the
1089 stop codon. *Proc Natl Acad Sci U S A, 92*(12), 5431-5435.
1090 https://doi.org/10.1073/pnas.92.12.5431
1091

1092 Miller, J. N., & Pearce, D. A. (2014, Oct-Dec). Nonsense-mediated decay in genetic disease:
1093 friend or foe? *Mutat Res Rev Mutat Res, 762*, 52-64.
1094 https://doi.org/10.1016/j.mrrev.2014.05.001
1095

1096 Neu-Yilik, G., Amthor, B., Gehring, N. H., Bahri, S., Paidassi, H., Hentze, M. W., & Kulozik, A. E.
1097 (2011, May). Mechanism of escape from nonsense-mediated mRNA decay of human

1098 beta-globin transcripts with nonsense mutations in the first exon. *RNA, 17*(5), 843-854.
1099 https://doi.org/10.1261/rna.2401811
1100

1101 Nguyen, L. S., Wilkinson, M. F., & Gecz, J. (2014, Oct). Nonsense-mediated mRNA decay:
1102 inter-individual variability and human disease. *Neurosci Biobehav Rev, 46 Pt 2*, 175-186.
1103 https://doi.org/10.1016/j.neubiorev.2013.10.016
1104

1105 Pelechano, V., & Alepuz, P. (2017, Jul 7). eIF5A facilitates translation termination globally and
1106 promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic*
1107 *Acids Res, 45*(12), 7326-7338. https://doi.org/10.1093/nar/gkx479
1108

1109 Pereira, F. J., Teixeira, A., Kong, J., Barbosa, C., Silva, A. L., Marques-Ramos, A., Liebhaber,
1110 S. A., & Romao, L. (2015, Jul 27). Resistance of mRNAs with AUG-proximal nonsense
1111 mutations to nonsense-mediated decay reflects variables of mRNA structure and
1112 translational activity. *Nucleic Acids Res, 43*(13), 6528-6544.
1113 https://doi.org/10.1093/nar/gkv588
1114

1115 Pierson, W. E., Hoffer, E. D., Keedy, H. E., Simms, C. L., Dunham, C. M., & Zaher, H. S. (2016,
1116 Sep 27). Uniformity of Peptide Release Is Maintained by Methylation of Release Factors.
1117 *Cell Rep, 17*(1), 11-18. https://doi.org/10.1016/j.celrep.2016.08.085
1118

1119 Popp, M. W., & Maquat, L. E. (2016, Jun 2). Leveraging Rules of Nonsense-Mediated mRNA
1120 Decay for Genome Engineering and Personalized Medicine. *Cell, 165*(6), 1319-1322.
1121 https://doi.org/10.1016/j.cell.2016.05.053
1122

1123 Schrader, R., Young, C., Kozian, D., Hoffmann, R., & Lottspeich, F. (2006, Nov 17).
1124 Temperature-sensitive eIF5A mutant accumulates transcripts targeted to the nonsense-

1125 mediated decay pathway. *J Biol Chem, 281*(46), 35336-35346.
1126 https://doi.org/10.1074/jbc.M601460200
1127

1128 Schuller, A. P., Wu, C. C., Dever, T. E., Buskirk, A. R., & Green, R. (2017, Apr 20). eIF5A
1129 Functions Globally in Translation Elongation and Termination. *Mol Cell, 66*(2), 194-205
1130 e195. https://doi.org/10.1016/j.molcel.2017.03.003
1131

1132 Shuvalov, A., Shuvalova, E., Biziaev, N., Sokolova, E., Evmenov, K., Pustogarov, N.,
1133 Arnautova, A., Matrosova, V., Egorova, T., & Alkalaeva, E. (2021, Nov 12). Nsp1 of
1134 SARS-CoV-2 stimulates host translation termination. *RNA Biol, 18*(sup2), 804-817.
1135 https://doi.org/10.1080/15476286.2021.1999103
1136

1137 Singh, G., Rebbapragada, I., & Lykke-Andersen, J. (2008, Apr 29). A competition between
1138 stimulators and antagonists of Upf complex recruitment governs human nonsense-
1139 mediated mRNA decay. *PLoS Biol, 6*(4), e111.
1140 https://doi.org/10.1371/journal.pbio.0060111
1141

1142 Smith, T., Heger, A., & Sudbery, I. (2017, Mar). UMI-tools: modeling sequencing errors in
1143 Unique Molecular Identifiers to improve quantification accuracy. *Genome Res, 27*(3),
1144 491-499. https://doi.org/10.1101/gr.209601.116
1145

1146 Susorov, D., Egri, S., & Korostelev, A. A. (2020, Dec). Termi-Luc: a versatile assay to monitor
1147 full-protein release from ribosomes. *RNA, 26*(12), 2044-2050.
1148 https://doi.org/10.1261/rna.076588.120
1149

1150 Tate, W. P., & Mannering, S. A. (1996, Jul). Three, four or more: the translational stop signal at
1151 length. *Mol Microbiol, 21*(2), 213-219. https://doi.org/10.1046/j.1365-
1152 2958.1996.6391352.x

1153

1154 Tork, S., Hatin, I., Rousset, J. P., & Fabret, C. (2004). The major 5' determinant in stop codon
1155 read-through involves two adjacent adenines. *Nucleic Acids Res, 32*(2), 415-421.
1156 https://doi.org/10.1093/nar/gkh201
1157

1158 van Engelen, B. G., Muchir, A., Hutchison, C. J., van der Kooi, A. J., Bonne, G., & Lammens, M.
1159 (2005, Jan 25). The lethal phenotype of a homozygous nonsense mutation in the lamin
1160 A/C gene. *Neurology, 64*(2), 374-376.
1161 https://doi.org/10.1212/01.WNL.0000149763.15180.00
1162

1163 Wallmeroth, D., Lackmann, J. W., Kueckelmann, S., Altmuller, J., Dieterich, C., Boehm, V., &
1164 Gehring, N. H. (2022, May 16). Human UPF3A and UPF3B enable fault-tolerant
1165 activation of nonsense-mediated mRNA decay. *EMBO J, 41*(10), e109191.
1166 https://doi.org/10.15252/embj.2021109191
1167

1168 Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R.,
1169 Morita, T., Iwamatsu, A., Hachiya, T., Kurata, R., Hirano, H., Anderson, P., & Ohno, S.
1170 (2009, May 1). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate
1171 remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay.
1172 *Genes Dev, 23*(9), 1091-1105. https://doi.org/10.1101/gad.1767209
1173

1174 Zetoune, A. B., Fontaniere, S., Magnin, D., Anczukow, O., Buisson, M., Zhang, C. X., &
1175 Mazoyer, S. (2008, Dec 5). Comparison of nonsense-mediated mRNA decay efficiency
1176 in various murine tissues. *BMC Genet, 9*, 83. https://doi.org/10.1186/1471-2156-9-83
1177

1178 Zhang, J., & Maquat, L. E. (1997, Feb 17). Evidence that translation reinitiation abrogates
1179 nonsense-mediated mRNA decay in mammalian cells. *EMBO J, 16*(4), 826-833.
1180 https://doi.org/10.1093/emboj/16.4.826

Zhu, X., Zhang, H., & Mendell, J. T. (2020, Jul 14). Ribosome Recycling by ABCE1 Links Lysosomal Function and Iron Homeostasis to 3' UTR-Directed Regulation and Nonsense-Mediated Decay. *Cell Rep, 32*(2), 107895. https://doi.org/10.1016/j.celrep.2020.107895