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

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

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

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

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

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

[5]Lead contact

*Corresponding Author: sujatha.jagannathan@cuanschutz.edu

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

Keywords: NMD, PTC, translation, termination, splicing

SUMMARY

Nonsense-mediated mRNA decay (NMD) is a quality control mechanism that prevents the accumulation of harmful truncated proteins by degrading transcripts with premature termination codons (PTCs). NMD activity varies across many contexts, with the variation having critical consequences during both normal development and disease. However, the factors that influence this variability remain poorly understood. Here, we investigate the relationship between the immediate sequence context of a PTC and NMD activity. We find an enrichment of glycine (Gly) codons preceding a PTC in common nonsense variants in contrast with a depletion of Gly codons preceding a normal termination codon (NTC). Furthermore, Gly-PTC enrichment is most pronounced in genes that tolerate loss-of-function variants, suggesting a biological impact for Gly-PTC contexts. Consistent with this idea, Gly-PTC contexts have higher NMD activity compared to an alanine-PTC context. We used a massively parallel reporter assay to comprehensively assess the effect of PTC sequence context on NMD activity, which revealed that peptidyl-tRNA hydrolysis rate during translation termination was the most important feature in discriminating high and low NMD activity. We show with biochemical assays in a eukaryotic system that Gly-TC contexts have the slowest termination rate compared to other codons. Based on these findings, we propose that NMD activity is modulated by the "window of opportunity" offered by peptidyl-tRNA hydrolysis rate and thus, translation termination kinetics. Gly-PTC contexts may have coevolved with NMD to ensure robust elimination of potentially toxic truncated proteins from non-essential genes.

INTRODUCTION

RNA quality control mechanisms ensure fidelity of gene expression in all eukaryotes. One such mechanism, nonsense-mediated mRNA decay (NMD), eliminates transcripts that contain premature termination codons (PTCs), often causing loss of gene function. PTC-induced loss of gene function underlies many Mendelian diseases, including cystic fibrosis, Duchenne muscular dystrophy, and β-thalassemia (Miller & Pearce, 2014; Mort et al., 2008). Much of our understanding of NMD comes from studying disease-causing PTCs, including the "50-55 nt rule", which posits that PTCs trigger NMD when present >50-55 nucleotides (nts) upstream of the 3'-most exon-exon junction (Nagy & Maquat, 1998). When the terminating ribosome fails to evict the exon-junction complex (EJC), it serves as a potent trigger for NMD (Gerbracht et al., 2020; Lykke-Andersen et al., 2001). An alternative mode of NMD is triggered when a PTC creates a long 3′ untranslated region (UTR) where, in the absence of EJCs, RNA binding proteins (RBPs) such as UPF1 help to recruit decay machinery (Hogg & Goff, 2010; Kurosaki & Maquat, 2013). In both cases, the terminating ribosome, along with downstream RBPs in the 3′ UTR, serve as the platform for the assembly of NMD factors and are essential to trigger NMD.

Studies over the years have shown that the "50-55 nt rule" does not encompass the breadth of NMD activity observed across transcripts and even across different PTCs within the same transcript (Neu-Yilik et al., 2011; Sato & Singer, 2021; Zhang & Maquat, 1997). For example, the 50-55 nt rule only predicted the fate of ~50% of PTCs in the Cancer Genome Atlas database; while additional factors such as exon length could account for another ~20% of NMD variability, ~30% of NMD variability is still unexplained by any known rule (Lindeboom et al., 2016; Lindeboom et al., 2019). Another study that investigated 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?

One potential factor which varies and likely influences NMD is translation termination efficiency. 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

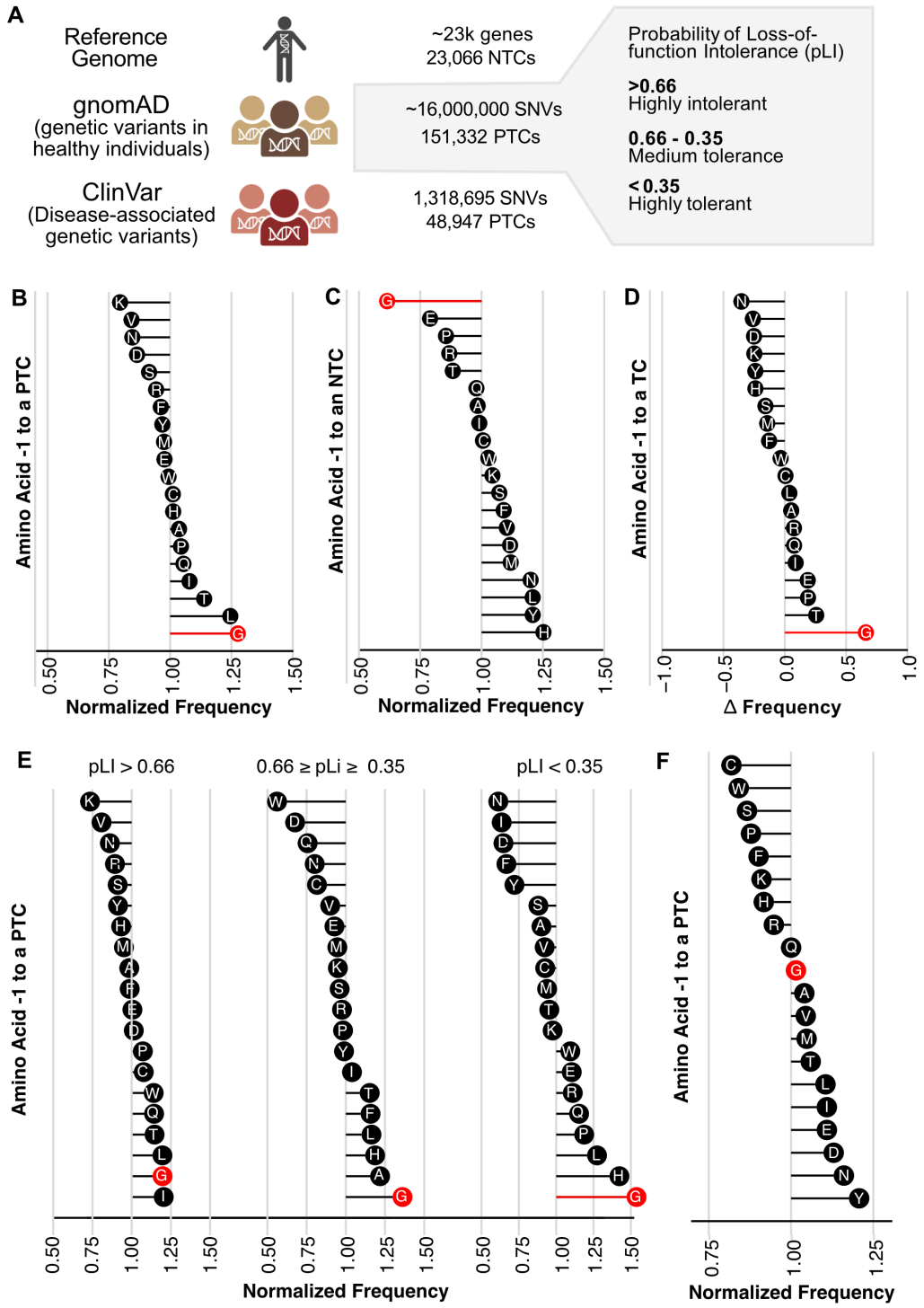
sequence context, as well as certain viral structures, can reduce termination fidelity (Baker & Hogg, 2017; Bonetti et al., 1995; Cridge et al., 2018; Loughran et al., 2014; Mangkalaphiban et al., 2021; Tork et al., 2004). Additionally, the identity of the nucleotide immediately after the stop codon 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, most of these studies focus on normal termination, which has been proposed to be distinct from premature termination (Amrani et al., 2004; Behm-Ansmant et al., 2007; Ivanov et al., 2016; Ivanov et al., 2008), and thus, how various cis-acting elements influence NMD is unknown.

In this study, we hypothesized that investigating sequence contexts of PTCs that occur in healthy individuals may offer novel insight into how these sequences influence NMD. Indeed, we found that Gly-PTC, a context highly enriched in healthy individuals, enhances NMD activity. To systematically assess the impact of PTC sequence context on NMD activity, we performed a massively parallel reporter assay (MPRA) and found a broad range of NMD activities associated with different sequences. Statistical modeling of the MPRA data identified peptidyl-tRNA hydrolysis rate as the major predictor of NMD activity. We propose that variable peptidyl-tRNA hydrolysis rates alter NMD activity by modulating the kinetic window available for NMD factors to bind and enact NMD during translation termination.

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 highly frequent PTCs (minor allele frequency (MAF) > 5%) are enriched for NMD escape events (Jagannathan & Bradley, 2016). Here, we focused on rare variants with MAF $\leq$ 1% among healthy individuals that would be expected to undergo NMD but perhaps with varying efficiencies. We analyzed enrichment of different sequences in PTC contexts of these variants within 10 amino acids preceding PTCs found in the gnomAD database (n=151,332, v2.1.1), which contains genetic variant data for >700,000 healthy humans (Karczewski et al., 2020) (**Fig. 1A, Supp. Fig. 1A)**. We compared the enrichment of these amino acids to that of normal termination codon (NTC) contexts in the reference genome (n=23,066, **Fig. 1A, Supp. Fig. 1B**). Gly codons stood out as the most enriched at the -1 position of PTCs and least enriched before the NTC (**Fig. 1B-C,** and **Supp. Fig. 1**). Threonine (Thr) and proline (Pro) were the next most enriched in PTC compared to NTC (**Fig. 1D**). Other amino acids show the opposite trend –enriched before an NTC and depleted

before a PTC – including asparagine (Asn), valine (Val), and aspartic acid (Asp). Clearly, the strongest effect was seen for Gly-NTC versus Gly-PTC contexts (**Fig. 1D**). These data suggest that the enrichment of Gly immediately upstream of a PTC (Gly-PTC) has a functional impact that allowed it to rise to high frequency in healthy individuals.



113 **Figure 1. Gly is enriched preceding a PTC in non-essential genes.** (**A**) Schematic for the analysis of
114 LoF scores for rare PTCs in healthy individuals as well as PTCs in disease contexts. (**B-C**) Amino acid
115 enrichment -1 to a PTC (B) or an NTC (C). (**D**) The frequency delta (enrichment before a PTC-NTC) for
116 each amino acid at the -1 position. (**E**) The -1 amino acid enrichment of PTCs binned into 3 categories of
117 probability of loss-of-function intolerance (pLI). The lower the pLI, the more tolerant the PTCs are to loss of
118 function. (**F**) The -1 amino acid enrichment of PTCs in disease contexts.

119

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

131

132 **Gly codons that precede a PTC enhance NMD activity compared to Ala.** A previous study
133 showed that Gly acts as a C-terminal degron and is rarely found at the C-terminal end of the
134 normal proteome, suggesting that Gly-PTC enrichment in non-essential genes may be due to
135 efficient Gly-end truncated protein degradation (Koren et al., 2018; Lin et al., 2018). It is also
136 possible that Gly-PTC stimulates PTC readthrough, leading to NMD escape and some
137 functionality of the protein product (Loughran et al., 2023), or that Gly-PTC enhances NMD and
138 elimination of the mRNA transcript. Readthrough would confer a benefit to the organism as it
139 would preserve gene function, while elimination of the truncated protein or encoding transcript
140 might have a neutral effect if the gene is otherwise non-essential.

141 To distinguish between these hypotheses, we used a set of NMD reporters (Kolakada et al.,
142 2024) that enabled the study of EJC-independent and EJC-enhanced forms of NMD, as well as
143 NMD-associated mechanisms such as PTC readthrough and truncated protein degradation. The
144 NMD-positive (NMD**+**) reporter consists of a bi-directional promoter that drives BFP as a
145 transfection control, in addition to an RFP and GFP containing transcript (**Fig 2A**, left). A stop
146 codon between RFP and GFP creates a long 3′ UTR on GFP, making the transcript a target of

147 the EJC-independent NMD. In another version of the NMD**+** reporter, an intron within the GFP
148 sequence renders the transcript an ideal EJC-enhanced NMD substrate (**Supp. Fig 2A**). A T2A
149 signal after the stop ensures independent expression of RFP and GFP proteins in the event of a
150 readthrough of the stop codon. Using this reporter, we varied the codon at the -1 position of the
151 PTC (TAA), testing two different glycine codons or an alanine codon. We also tested these
152 sequences at the NTC of an NMD-negative (NMD**–**) reporter, which lacks the GFP 3′ UTR and so
153 behaves as a normal transcript, allowing us to test the effect of glycine as a C-end degron
154 independent of NMD. (**Fig 2A**, middle).

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

164 As shown in Fig 2B, all reporters yielded robust BFP expression. While the no-PTC control
165 construct produced GFP, there was no GFP protein detected with either the glycine or alanine
166 PTC contexts (**Fig 2B**), ruling out the hypothesis that a Gly-PTC context promotes readthrough.
167 Notably, both EJC-independent NMD**+** reporters with Gly-PTC resulted in less RFP protein
168 compared to the Ala-PTC reporter, a difference that could be recovered upon treatment with the
169 NMD inhibitor, SMG1i (quantified in **Fig 2C**). In comparison, only a modest reduction in RFP
170 protein was observed for the Gly-PTC NMD**–** reporters compared to the Ala-PTC reporter (**Fig**
171 **2B**), suggesting that a C-end glycine degron was not sufficient for the protein depletion we
172 observed in the NMD+ reporters. No RFP was detected from any of the EJC-enhanced NMD
173 reporters, consistent with their high NMD activity (**Supp. Fig 1B-C**). These data demonstrate that
174 varying the codon preceding the PTC can inherently influence NMD activity through a hitherto
175 unknown mechanism.

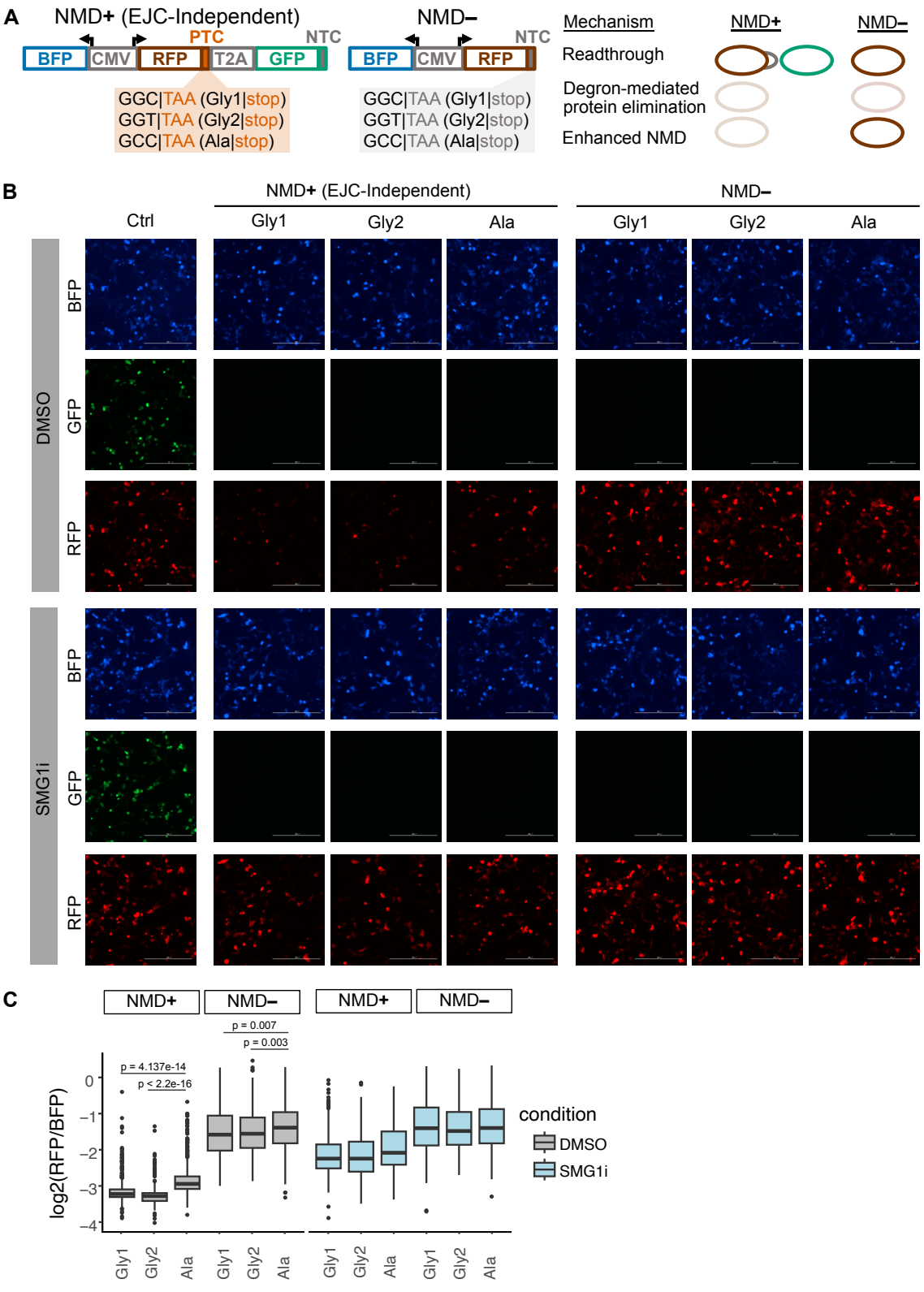


Figure 2. Gly-PTC contexts enhance NMD activity compared to Ala-PTC. (A) Schematic of NMD reporter system used. Left: EJC-independent NMD+ reporter, middle: NMD– reporter; right: expected outcomes. **(B)** Fluorescence imaging of BFP, GFP, and RFP from cells transfected with NMD reporters

180 containing Gly-PTC or Ala-PTC reporters, treated with a DMSO as a vehicle control or SMG1i. Ctrl refers
181 to a construct that robustly expresses all three fluorescent proteins. Gly1 and Gly2 are the GGC and GGT
182 codons for Gly. The codon used for Ala was GCC. **(C)** Quantification of fluorescence from panel (B).

183

184 **Massively parallel reporter assay (MPRA) captures widespread sequence-dependent**
185 **variation in NMD activity.** Having established that Gly-PTC causes enhanced NMD compared
186 to Ala-PTC, we next sought to systematically examine the relationship between NMD activity and
187 the PTC sequence context. We used a massively parallel reporter assay wherein we varied the 6
188 nucleotides before the PTC, the PTC itself, and 1 nucleotide after the PTC in an unbiased fashion
189 (N_6TRRN; corresponding to 65,536 unique sequences; **Fig. 3A**) and asked how these sequences
190 impacted NMD activity in both EJC-independent and EJC-enhanced reporters. Because we
191 employed stop codons with "TRR" for array-based oligonucleotide synthesis, 25% of our
192 sequence library consists of a tryptophan (Trp) codon (UGG) instead of a PTC and serve as
193 control sequences that do not undergo NMD. All sequences were incorporated into both EJC-
194 independent and EJC-enhanced fluorescent reporter backbones (**Supp. Fig. 3A-B**) and
195 integrated into HEK293T cells via Cre-Lox mediated recombination (Khandelia et al., 2011).
196 These cells were then treated with either DMSO or SMG1i, harvested for RNA, and subjected to
197 targeted sequencing. Sequencing reads were analyzed by counting the occurrences of each
198 reporter within a sample. The NMD activity of an individual reporter sequence was the $\log_2$
199 transcript fold change between DMSO-treated cells (NMD active) and SMG1i-treated cells (NMD-
200 inhibited) calculated with DESeq2 (Love et al., 2014), thus a lower (negative) value means higher
201 NMD activity, and vice versa. The ~1/4th of control sequences containing the Trp codon rather
202 than a stop codon were used for normalization.

203 The NMD activity as measured by transcript fold changes for both the EJC-independent and
204 EJC-enhanced reporters had a bimodal distribution, corresponding to the stop codon identity of
205 the sequences in each population (**Fig. 3B**): sequences containing a stop codon were left-shifted
206 compared to the sequences containing the Trp control, indicating that the former were targets of
207 NMD. In the population of sequences undergoing NMD (i.e. those containing a stop codon) there
208 was a wide range of NMD activity for both the EJC-independent and the EJC-enhanced reporters.
209 EJC-enhanced reporter expression was reduced by 4.8 fold (log2FC: -2.27), which is greater than
210 the 3.3 fold decrease (log2FC: -1.71) for EJC-independent reporters (**Fig. 3B**), consistent with
211 EJC-enhanced targets undergoing more efficient NMD compared to EJC-independent targets
212 (Buhler et al., 2006; Chu et al., 2021; Singh et al., 2008). These data indicate that NMD activity
213 indeed varies based on PTC sequence context.

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

224 As expected, our modeling showed that all stop codons caused significantly stronger NMD
225 activity compared to the Trp control for both EJC-independent and EJC-enhanced reporters (**Fig.**
226 **3C**). Notably, the TGA stop codon increased NMD activity by ~20% compared to the TAG stop
227 codon for the EJC-independent reporters. However, this difference was not observed for the EJC-
228 enhanced reporters where all stop codons elicit similar levels of NMD activity (**Fig. 3C**). For
229 subsequent analyses, only transcripts expected to undergo NMD (i.e. transcripts with stop
230 codons) were analyzed. When determining the effect of the +4 nt identity on NMD, we found that
231 T, G, and C promoted weaker (~6% decrease) NMD activity compared to A for the EJC-
232 independent reporters. A similar effect was present only for G compared to A for the EJC-
233 enhanced reporters (**Fig. 3D**). Overall, the nucleotide after the PTC does not critically impact the
234 extent of NMD activity.

235 The identity of the codon at the –2 position had a wide range of effects on NMD activity for
236 the EJC-independent reporters. As the –2 codon position is coded by randomized nucleotides
237 (NNN), it can include all three stop codons as well. We find that in fact a stop codon at the –2
238 position causes the strongest increase (~40%) in NMD activity compared to the control variable
239 Ala (**Fig. 3E**). Codons for amino acids caused smaller changes in NMD activity, increasing NMD
240 activity up to ~10% (Val) or reducing it up to 20% (Arg) compared to Ala. For the EJC-enhanced
241 reporters, most amino acid-encoding codons did not affect the efficiency of NMD. Interestingly,
242 the presence of a stop codon in the –2 position weakened NMD by ~10% compared to the control
243 variable Arg, opposite in trend to the EJC-independent reporters. Lys at the -2 position had the
244 largest effect compared to Arg, weakening EJC-enhanced NMD by ~20% (**Fig. 3E**). Altogether,
245 the codon at the -2 position can influence NMD activity, impacting the EJC-independent reporters
246 more than the EJC-enhanced.

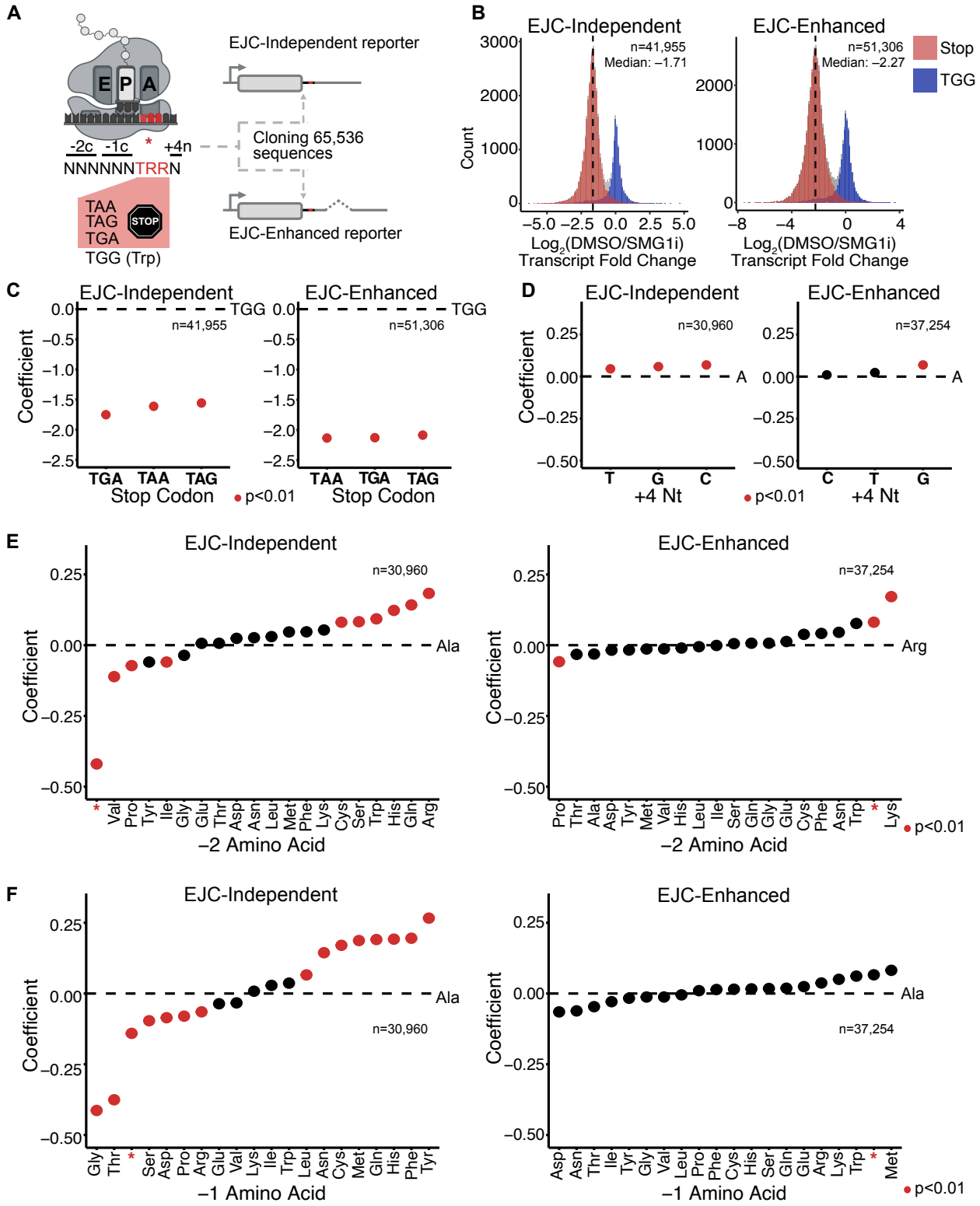


Figure 3. MPRA shows a wide range of PTC context-dependent NMD efficiencies. (A) Design of the MPRA. **(B)** Histogram of transcript fold change EJC-independent and EJC-enhanced NMD+ reporters comparing DMSO treated cells to SMG1i treated cells. **(C-F)** Dot plots of the beta coefficient representing

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

253

254 As with the -2 position, the codon at the -1 position is also encoded by randomized nucleotides
255 (NNN) and can thus include stop codons. Remarkably, when examining the effect of the codon at
256 the -1 position on NMD activity, we found that Gly-PTC (the same context enriched in the gnomAD
257 database; **Fig 1**) has the strongest effect on EJC-independent NMD activity, reducing it by 40%
258 compared to Ala (**Fig. 3F**). Thr followed closely after and both Gly-PTC and Thr-PTC cause
259 stronger NMD than a stop codon in the -1 position. In contrast, Tyr-PTC had the least efficient
260 NMD, weakening NMD by ~25% compared to Ala-PTC. These effects were not observed for the
261 EJC-enhanced reporters where the -1 amino acid identity did not cause any apparent effects (**Fig.**
262 **3F**).

263 The discrepancy in results between the EJC-independent and EJC-enhanced reporters could
264 stem from two causes: the highly efficient EJC-enhanced NMD is insensitive to other modifying
265 factors; or the PTC sequence context effect on NMD is transcript-specific and limited to the
266 fluorescent reporters. To distinguish between these possibilities, we switched to an independent,
267 luciferase-based reporter system (Kolakada et al., 2024). This system is identical to our
268 fluorescent reporter system except that the transfection control is Renilla luciferase and the NMD
269 readout is provided by Firefly luciferase (**Supp. Fig. 4A**).

270 To test whether the codon at the -1 position also affected NMD within the luciferase reporter
271 system we varied the amino acid in the -1 position of either the PTC or the NTC, choosing Gly,
272 Val, Ala, and Tyr, which display a range of NMD efficiencies from highest to lowest (**Supp. Fig.**
273 **4A**). As expected, luciferase activity was ~10-fold lower for the EJC-enhanced reporters
274 compared to EJC-independent (**Supp. Fig. 4B**). Notably, the relative trend in -1 amino acid-
275 dependent NMD variability was the same for both the EJC-independent and EJC-enhanced
276 reporters. The NMD activity varied depending on the -1 amino acid, with Gly having ~22% or
277 ~29% more efficient NMD compared to Tyr for the EJC-independent and EJC-enhanced
278 reporters, respectively (**Supp. Fig. 4B**). Compared to the EJC-independent and EJC-enhanced
279 reporters, the NMD**–** reporters did not vary with the change in the -1 amino acid (**Supp. Fig. 4B**).
280 Taken together, these data demonstrate that the -1 amino acid effect can extend to both forms of
281 NMD, although it is unclear why it is not observed in the MPRA data with the fluorescent reporters.

282 Next, we asked if there is codon-level specificity to the effect of amino acids on EJC-
283 independent NMD activity. We found that the effects of the codons in the -2 and -1 positions were

284 largely grouped by the identity of the encoded amino acid, with only a few that were significantly
285 different (**Supp. Fig. 5A-B**, all codons compared to Lys-AAA). An exception to this trend was Leu
286 at the -1 position whose codon TTA seems to have the opposite effect on NMD compared to the
287 rest of the Leu codons. Reassuringly, all the effects at the amino acid level were replicated at the
288 codon level with the stop codons having the strongest NMD activity at the -2 position and Arg the
289 weakest (**Supp. Fig. 5A**), and Gly having the strongest NMD activity at the -1 position and Tyr
290 the weakest (**Supp. Fig. 5B**). Taken together these data suggest that the amino acids encoded
291 primarily by the codon in the -1 position to the PTC are key to determining the efficiency of EJC-
292 independent NMD.

293

294 **tRNA-peptidyl hydrolysis is the most important feature for EJC-independent NMD**
295 **substrate activity.** To discover other sequence-dependent factors that could influence NMD
296 activity, we expanded the set of biological variables and constructed a random forest classifier
297 trained to discriminate between the high and low NMD activity (**Fig. 4A**). We selected the 3000
298 sequences with the highest (lowest Log2 FC) and lowest (highest Log2 FC) NMD activity for the
299 EJC-independent and EJC-enhanced reporters.

300 We introduced additional variables such as the identities of the codons as well as their
301 properties such as usage frequency (GenScript), tRNA binding strength, and optimality (Bazzini
302 et al., 2016; Grosjean & Westhof, 2016); the identities of the amino acids as well as their physical
303 properties such as charge, solubility, dissociation constants, and hydropathy (ThermoFisher
304 Scientific); the GC and pyrimidine content of each of the sequences; the prokaryotic peptide
305 release rates based on the identity of the amino acid encoded by the -1 codon, as these values
306 are not available for eukaryotic systems (Pierson et al., 2016).

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

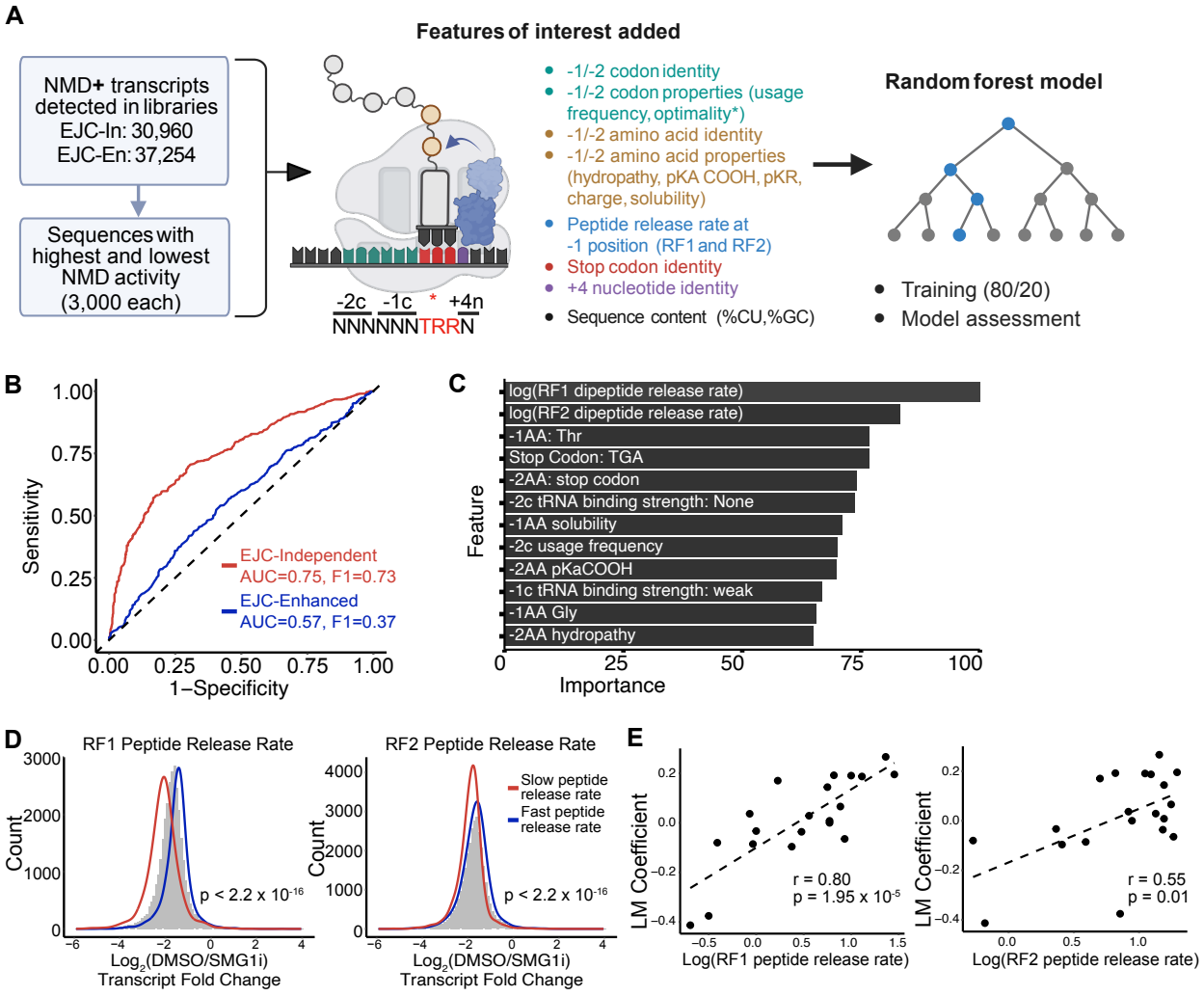


Figure 4. Peptide release rates predict NMD activity (A) Data preparation and variables input into the Random Forest model. **(B)** ROC curves, AUC values, and F1 scores for the EJC-independent and EJC-enhanced classifiers. **(C)** Ranked feature importance for the EJC-independent model. **(D)** Density 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, fast release rates is defined as log (RF1 release rate) > 1. For RF2, slow release rates are defined as log (RF2 release rate) £ 0, fast release rates is defined as log (RF2 release rate) > 1.25. 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.

We focused on the features that were most important for the EJC-independent model. Prokaryotic release factor 1 (RF1) and release factor 2 (RF2) peptide release rates emerged as the most important predictor of NMD activity (**Fig 4C**). Although these rates were observed in a prokaryotic context, the similarity in peptide release mechanisms between prokaryotes and eukaryotes suggests that these findings may also apply to eukaryotic systems (Hellen, 2018). To examine the effect of peptide release rates on NMD activity, we compared the subpopulation of

334 sequences with the fastest and slowest dipeptide release rates versus all NMD targets (**Fig. 4D**).
335 The two populations were well-separated for RF1 peptide release rates and less robustly for the
336 RF2 peptide release rates (**Fig. 4D**). To examine this relationship more quantitatively we
337 compared the beta coefficients of the effect on NMD activity of the -1 amino acid identity to the
338 peptide release rates of RF1 and RF2 (**Fig 4E**). We observed a strong correlation for the RF1
339 peptide release rates ($r = 0.8$; $p = 1.95 \times 10^{-5}$) and a weaker but significant correlation for the RF2
340 peptide release rates ($r = 0.55$; $p = 0.01$) (**Fig 4E**). Taken together these data suggest that peptide
341 release rate may be a major effector of the -1 amino acid-dependent effect on NMD activity.

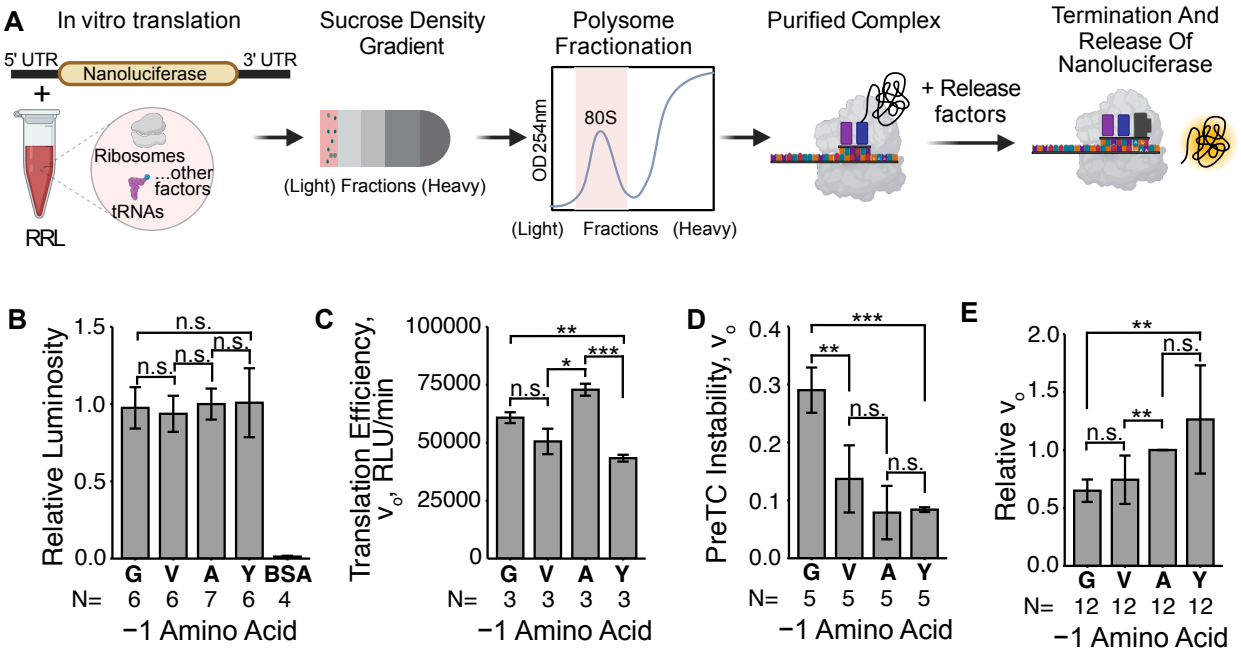
342

343 **Eukaryotic translation termination efficiency varies with the identity of the amino acid**
344 **upstream of the stop codon *in vitro*.** The peptide release rate values used in the Random
345 Forest classifier (**Fig. 4**) came from a prokaryotic system and we wanted to see if it reflected the
346 peptidyl tRNA hydrolysis rate in eukaryotic cells (Pierson et al., 2016). To determine if the rate of
347 peptide release varied based on the identity of the -1 amino acid in a eukaryotic system, we used
348 the *in vitro* Termi-Luc assay (**Fig. 5A**) in the excess of mutated eukaryotic release factor 1
349 (eRF1(AGQ)). This protein is able to recognize stop codons, but unable to induce peptidyl-tRNA
350 hydrolysis (Frolova et al., 1999). eRF1(AGQ) binding allows to freeze translation at the
351 termination stage and obtain pretermination ribosomal complexes (preTCs). PreTCs are then
352 purified from eRF1(AGQ) by sucrose gradient centrifugation, followed by the addition of release
353 factors (working concentrations determined in **Supp. Fig. 6A**). The real-time release kinetics of
354 the synthesized NLuc are then measured to determine the translation termination rate of the
355 transcript. We used this system to test the termination rates when Gly, Val, Ala, or Tyr codons
356 (which represent the full range of observed NMD efficiencies from high to low) are present at the
357 -1 position to the stop codon. If peptide release rate influences the time of residence of the
358 terminating ribosome, we would expect to see a difference in termination rate when different
359 amino acids are present upstream of the stop codon.

360 First, control experiments were performed to ensure that substitution of an amino acid residue
361 did not affect the activity, translation efficiency, or preTC stability of NLuc. To determine whether
362 NLuc activity was impacted by altering the last amino acid, we purified all tested peptide variants
363 as recombinant proteins and determined their maximum luminescence. All NLuc peptide variants
364 had the same activity (**Fig. 6B**). To determine whether translation efficiency was impacted by
365 changing the last amino acid, we calculated the maximal derivative (slope) of the linear portion of
366 the luminescence production curve. The identity of the last amino acid affected translation

367 efficiency with Tyr having the lowest and Ala having the highest efficiency (**Fig. 6C**). To determine
368 preTC instability, the rate of peptide release of each peptide variant was calculated without the
369 addition of any release factors. These data showed that Gly had the most unstable preTC, while
370 Tyr had the most stable preTC (**Fig. 6D**). Finally, we performed the Termi-Luc assay on each
371 peptide variant starting with equal amounts of preTC complexes to account for the variability in
372 translation efficiency and preTC stability. We found that Gly had the slowest release rate, and Tyr
373 had the fastest release rate, a difference which was statistically significant and correlated with
374 high and low NMD activity, respectively (**Fig. 6E**). Taken together, these data show that peptide
375 release rates indeed vary based on the identity of the last amino acid in a eukaryotic system, with
376 glycine showing the slowest termination rate and tyrosine the fastest, among the amino acids
377 tested.

378



379
380 **Figure 5. Peptide release rate in an *in vitro* eukaryotic system varies with the identity of the amino**
381 **acid upstream of the stop codon. (A)** Schematic of the Termi-Luc peptide release assay. This assay was
382 performed for nanoluciferase reporters with different amino acids -1 to the stop codon. **(B)** Tested *in vitro*
383 luminescence activity of recombinant NLuc variants. **(C)** Translation efficiency of each reporter in the *in*
384 *vitro* system. **(D)** Relative rate of peptide release when no release factors are added to the termination
385 reaction i.e., preTC instability. **(E)** Relative peptide release rates for each reporter.

386

387 **Key NMD and translation factors do not influence the -1 amino acid dependent variation in**
388 **NMD activity.** Finally, we wanted to ask whether any trans factors played a role in the -1 amino

acid-dependent effect on NMD activity. To this end, we utilized shRNAs to knock down various NMD- and translation-related factors and test the effect of the knockdown on reporter levels using the EJC-independent reporters.

We investigated the UPF proteins (UPF1, UPF2, and UPF3B), which are key effectors of NMD (Kashima et al., 2006; Yamashita et al., 2009); SMG1, which phosphorylates UPF1 to trigger NMD; and SMG6, which is responsible for cleaving the transcript when NMD is initiated (Arias-Palomo et al., 2011; Yamashita et al., 2009) (**Fig. 6A-B**). Reporter protein levels increased upon knockdown of all NMD factors except for UPF3B, likely because UPF3B is known to be redundant to UPF3A (Wallmeroth et al., 2022). However, the relative difference in expression of the reporters based on -1 amino acid identity largely remained the same (**Fig. 6B**; **Supp. Fig. 7A**).

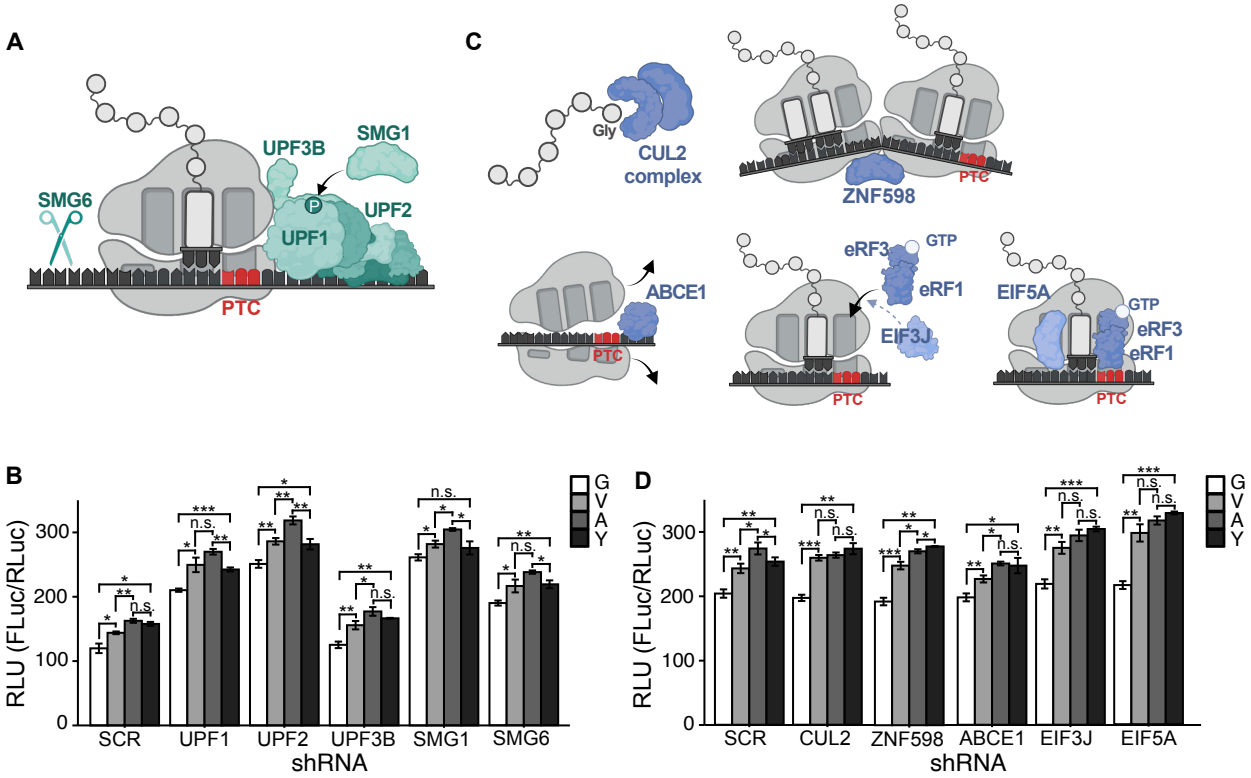


Figure 6. Key NMD and translation factors do not influence the -1 amino acid dependent variation in NMD activity. (A) Model of NMD factors tested. **(B)** shRNA knockdown of key NMD factors and their effect on the luminescence-based EJC-independent reporters. **(C)** Models of the mechanisms of translation 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.

Next, we investigated key translation and protein quality control related factors, CUL2, ZNF598, ABCE1, EIF3J, and EIF5A using the same approach (**Fig. 6C-D**). CUL2 forms a key part of the protein complex that recognizes C-terminal Gly degrons (Koren et al., 2018; Lin et al., 2018). ZNF598 recognizes the interface of collided disomes, thus playing a role in ribosome-

409 associated quality control substrate recognition (Juszkiewicz et al., 2018). It is possible that some
410 sequences would promote ribosome collision at the end of a transcript, which would then influence
411 NMD. ABCE1 is a ribosome recycling factor that acts during normal and premature termination
412 (**Fig. 6C**) (Hellen, 2018; Jackson et al., 2012; Zhu et al., 2020). EIF3J facilitates the binding of the
413 release factors to the ribosome, and thus stimulates peptide release (Egorova et al., 2021).
414 Finally, EIF5A is an elongation factor that aids the ribosome in translating Gly/Pro sequences.
415 EIF5A accumulates at the ribosome during translation termination and is hypothesized to bind at
416 the E-site when the deacylated tRNA leaves (Pelechano & Alepuz, 2017; Schrader et al., 2006;
417 Schuller et al., 2017). EIF5A binding to the termination complex may be influenced by the
418 transcript sequence within the ribosome. Similar to the NMD factors, none of these conditions
419 eliminated the differences in reporter level expression based on -1 amino acid identity (**Fig. 6D**;
420 knockdown validation **Supp. Fig. 7B**), suggesting that observed variability is the result of the
421 inherent termination kinetics at the PTC.

422 Taken together, our data supports a model where slower peptide release rate offers a longer
423 kinetic window for NMD factors to assemble and degrade a transcript, thus causing highly efficient
424 NMD (**Fig. 7**).
425

426 **DISCUSSION**

427 NMD is often considered a binary outcome that relies on the presence of triggering factors
428 downstream of the PTC. However, recent results increasingly suggest that NMD is tunable
429 (Jagannathan & Bradley, 2016; Lindeboom et al., 2019). However, the factors influencing this
430 variability are incompletely understood. While translation termination is a pre-requisite for NMD,
431 it is not thought to quantitatively influence NMD activity. Our data support a window of opportunity
432 model where the time of residence of the terminating ribosome, which serves as the platform for
433 the assembly of NMD machinery, determines the effectiveness of NMD (**Fig. 7**). Consistent with
434 this model, slower peptidyl-tRNA hydrolysis rate enhances NMD activity, while faster rate leads
435 to less efficient NMD. Together with the RBP cues that trigger NMD, termination kinetics could be
436 a key modifier of the probability with which a prematurely terminating ribosome leads to NMD.

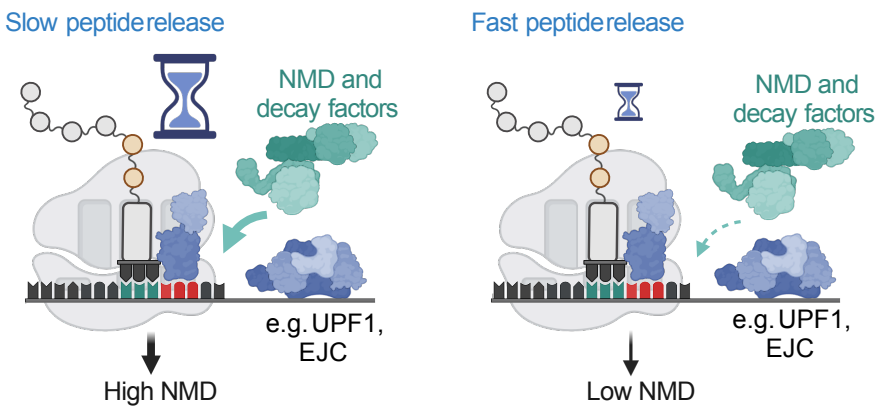


Figure 7. A kinetic model for the variability in NMD activity dependent upon the codon preceding the PTC. Window of opportunity model for how NMD activity is influenced by peptide release rates. A slower peptide release rate gives NMD factors a larger window to bind an act on a target, and vice versa.

Several lines of evidence point to the significance of the Gly-PTC context leading to higher efficiency of NMD. We show that Gly codons are most enriched before a PTC and least enriched before an NTC, cause the most efficient NMD, and have the slowest associated peptide release rate. Further, PTCs with Gly-PTC contexts are enriched among variants that are tolerant to loss of function in healthy individuals, while they are not enriched among disease-causing variants. Gly is also a C-terminal degron, which suggests that Gly-end proteins, truncated or otherwise, are more efficiently cleared (Koren et al., 2018; Lin et al., 2018). Thus, the preponderance of glycine codons at the ends of truncated proteins may be the result of co-evolution of PTCs alongside processes that efficiently eliminate Gly-end proteins at both the RNA and protein levels.

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 cause efficient NMD (less efficient than Gly but more efficient than most other amino acids) and have slow peptide release rates. On the other end of the spectrum, Tyr which had the lowest NMD activity in our experiments, was enriched before an NTC but not a PTC. It also has the fastest peptide release rate. PTCs with Tyr at the -1 of a PTC are depleted among variants that are tolerant to loss of function and are enriched in disease contexts.

Surprisingly, we also find that the stop codon TGA caused 20% more efficient EJC-independent NMD compared to 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 lowest fidelity i.e. it is most likely to be readthrough, which usually stabilizes an NMD target (Tate & Mannering, 1996). Our results suggest there are kinetic

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

In summary, our results point to a key role for the kinetics of translation in determining NMD activity, which may also extend to other translation-associated quality control processes. Moreover, the effect of various amino acids upstream of the PTC on NMD activity and their pattern of enrichment in healthy and diseased populations suggest that NMD has played a key role in shaping the evolution of nonsense variants. These findings can be used to better model the phenotypic outcomes of PTCs and thus, develop more effective strategies to counteract PTC-associated diseases.

LIMITATIONS OF THE STUDY

Linear modeling and the Random Forest classifier of the MPRA data suggested that PTC-sequence context mattered more for EJC-independent NMD compared to EJC-enhanced NMD. In contrast, individual luciferase reporters demonstrated that the -1 amino acid effect applied to both EJC-independent and EJC-enhanced forms of NMD. The cause for this apparent discrepancy in the results from these two analyses is unclear. We know that EJC-independent NMD is weaker than EJC-enhanced NMD (Buhler et al., 2006; Chu et al., 2021; Kolakada et al., 2023; Singh et al., 2008) and has distinct triggers. Thus, the window of opportunity during termination may matter more for a weaker, EJC-independent target, than a stronger EJC-enhanced target. It is also possible that different reporter genes respond distinctly to PTC sequence context in their NMD activity. Further studies are needed to understand what other variables may influence the window of opportunity model as it applies to the two modes of NMD.

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.

510 **STAR METHODS**

511

512 **RESOURCE AVAILABILITY**

513

514 **Lead contact**

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

517

518 **Materials availability**

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

521

522 **Data and code availability**

523 • All sequencing data have been deposited at GEO (GSE261151) and will be publicly
524 available as of the date of publication.
525 • All original code is available at the GitHub repository: https://github.com/jagannathan-
526 lab/2024-kolakada_et_al.
527 • Any additional information required to reanalyze the data reported in this paper is available
528 from the lead contact upon request.

529

530 **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

531

532 **Cell lines and culture conditions**

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

538 **METHOD DETAILS**

539 **gnomAD and ClinVar analysis**

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

552 **Cloning**

553 The NMD reporter plasmids have been previously described (Kolakada et al., 2024). For
554 individual fluorescent and luminescent reporters containing Gly, Val, Ala, and Tyr amino acids
555 before the PTC, oligos containing these sequences (Integrated DNA Technologies) and matching
556 the EcoRI (NEB, R3101S) and XhoI (NEB, R0146S) restriction sites were synthesized. These
557 oligos were annealed and ligated using the Quick Ligation Kit (NEB, M2200L) into EJC-
558 independent and EJC-enhanced fluorescent and luminescent backbones digested with EcoRI
559 and XhoI. To make the NMD– reporters for each sequence, each EJC-independent reporter was
560 digested with EcoRI and MfeI (NEB, R3589S), the sticky ends were filled in using Klenow
561 Fragment (NEB, M0212S) and the blunt ends were ligated together using the Quick Ligation Kit
562 (NEB), removing the GFP 3′ UTR. The oligos synthesized for each sequence tested are as
563 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

564

565 **MPRA plasmid library**

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

577 The EJC-independent and EJC-enhanced fluorescent reporter backbones were linearized
578 using EcoRI (NEB) and XhoI (NEB) at room temperature overnight. Digested plasmid DNA was
579 gel purified using the NucleoSpin Gel and PCR Clean-up kit (Takara Bio, 740609.250). The oligo
580 library was then cloned into the digested backbones using the Gibson Assembly Master Mix (NEB,
581 E2611L) using an insert:vector molar ratio of 7:1 for the EJC-independent backbone and 3:1 for
582 the EJC-enhanced backbone. The reactions were incubated at 50°C for 1 h. The reactions were
583 then purified using AMPure XP beads to remove excess salts and a total of 200 ng of plasmid
584 DNA was transformed into MegaX DH10B T1R Electrocompetent Cells (ThermoFisher,
585 C640003), using a Biorad GenePulser electroporator. The transformed cells were recovered in
586 the recovery medium provided with the electrocompetent cells at 37°C for 1 h, then plated on 33
587 pre-warmed 15 cm Luria broth (LB) agar-Carbenicillin (RPI, C46000-5.0) plates and incubated at
588 37°C overnight. The next day colonies were collected by scraping plates, spun-down, and then
589 midi-prepped (NucleoBond Xtra Midi EF kit, 740420.50) to extract the plasmid DNA.

590

591 **Plasmid transfections**

592 Transfections for fluorescent and luminescent reporters were performed differently since
593 fluorescent proteins have longer maturation rates than luciferase proteins (Mazo-Vargas et al.,
594 2014). For the fluorescent reporters, cells were split 24 h after transfection equally into two new
595 6-wells. Forty-eight hours after transfection, one set of cells was harvested for RNA and the other
596 set for flow cytometry. For the luminescent reporters, transfections were performed in 24-well
597 plates at a 40% confluency with 500 ng of plasmid DNA using Lipofectamine 2000 (ThermoFisher
598 Scientific, 11668027), as per manufacturer's instructions. All cells were harvested for luciferase
599 assays.

600 **Fluorescence imaging**

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

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

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

616 **Targeted sequencing of MPRA libraries**

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

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

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

639 **MPRA data analysis**

640 To process the MPRA sequencing libraries, the UMI-tools repository was used to add the
641 UMIs to the name of each read in the FASTA files (Smith et al., 2017). Following this, Cutadapt
642 was used to trim the 5′ and 3′ends of the reads in a sequence dependent manner, up to the 10 nt
643 context of interest (Martin, 2011); ~25% of reads were lost in this process. SeqPrep was used to
644 merge sequencing reads 1 and 2, only accepting reads that perfectly align
645 (https://github.com/jstjohn/SeqPrep); ~5% of reads were lost in this process. Bowtie was used to
646 map the reads to a reference file of all sequence contexts (Langmead & Salzberg, 2012); 1% of
647 reads were lost in this process. Samtools was used to create a BAM file of mapped reads, followed
648 by deduplicating the BAM files based on UMIs and converting the BAM files to BED files (Li et al.,
649 2009). Around 10% of reads were lost upon deduplication of UMIs. A custom python script was
650 used to remove reads on the negative strand and to get a bed file of reads in the correct
651 orientation.

652 The NMD activity of an individual reporter transcript was the log2-fold change in between
653 DMSO-treated cells to SMG1i-treated cells calculated with DESeq2 (Love et al., 2014). The
654 reporter transcripts that do not undergo NMD, i.e. Trp codon rather than a stop codon at the PTC
655 position, were used similar to housekeeping normalization controls within DESeq2. Both linear
656 models and random forest classifiers were performed in R. Random Forest models were built

657 utilizing the caret R package (doi:10.18637/jss.v028.i05). First, features with variance close to
658 zero were eliminated, after which the linear correlation for all numeric features was calculated.
659 For the features with correlation ≥ 0.85 one feature was randomly selected, and the others
660 discarded from the model. Data were divided into training and validation set using 0.8:0.2 ratio.
661 Hyperparameters tuning was performed for train set using 5x train-test cross validation for the
662 following model parameters: mtry, maxnodes and ntree. Performance of final model was
663 measured using F1 scores and on validation set. The code and packages used for these analyses
664 can be found in this repository: https://github.com/jagannathan-lab/2024-kolakada_et_al.

665 **Luciferase assays**

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

674 **Lentiviral transductions**

675 Knockdowns were conducted using shRNAs. Plasmids containing shRNAs were first
676 packaged into lentivirus. To do this, HEK293T cells were transfected with 4 µg of the relevant
677 shRNA vector (Functional Genomics Core, CU Anschutz), 3 µg of psPAX2 (Addgene plasmid
678 #12260), and 1 µg of pMD2.G (Addgene plasmid #12259), in 80% confluent 6 cm plates. The
679 shRNAs used were for scrambled (SCH016), UPF1 (TRCN0000022254), UPF2
680 (TRCN0000151381), UPF3B (TRCN0000152769), SMG1 (TRCN0000194827), SMG6
681 (TRCN0000040014), ZNF598 (TRCN0000073159), EIF5A (TRCN0000062552), EIF3J
682 (TRCN0000062013), CUL2 (TRCN0000006523), and ABCE1 (TRCN0000158417). Twenty-four
683 hours after transfection the media was replaced with 3 mL of fresh media. Forty-eight hours after
684 transfection media containing lentivirus was harvested and spun down at 1200 rpm for 5 minutes.
685 Aliquots of the supernatant were frozen at -80°C for later use.

686 For the knockdown experiments with the luciferase reporters, cells were seeded at a 40%
687 confluency in 6-wells and infected with 50 µL of virus. Twenty-four hours post transduction, the
688 cells were split into 24-wells at a 20% confluency. Forty-eight hours post transduction, the

689 luciferase reporters were transfected into cells as described above. Seventy-two hours post-
690 transduction the cells were harvested for luciferase assays. For knockdown validation
691 experiments, the same procedure was followed without the transfection of luciferase reporters.

692 **RNA extraction and RT-qPCR**

693 For RT-qPCR used to validate knockdowns of SMG1, ZNF598, and EIF3J, RNA was
694 extracted from cells using TRIzol Reagent (ThermoFisher Scientific). DNase digestion was
695 performed with DNaseI (ThermoFisher Scientific, 18068-015) on 1 µg of extracted RNA. This was
696 followed by cDNA synthesis using the SuperScript III First-Strand Synthesis kit, using random
697 hexamers. A no-RT sample was included as a control to make sure there was no genomic DNA
698 contamination. The cDNA was then diluted 1:4 and 2 µL were used per 10 µL qPCR reaction.
699 qPCR was performed using iTaq Universal SYBR Green Supermix and primers specific for
700 SMG1, ZNF598, EIF3J, and the housekeeping gene RPL27, were used at a final concentration
701 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

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

709

710 **Protein isolation and western blotting**

711 Cells were harvested in RIPA buffer supplemented with complete protease inhibitor
712 (Roche, 11836170001). Protein was run on NuPAGE Bis-Tris precast polyacrylamide gels
713 (Thermo Fisher Scientific, NP0323BOX) alongside PageRuler Plus Prestained Protein Ladder

714 (Thermo Fisher Scientific, 26619) and transferred to Odyssey nitrocellulose membrane (LI-COR
715 Biosciences, 926-31092). Membranes were blocked in Intercept (PBS) Blocking Buffer (LI-COR
716 Biosciences, 927-70001) before overnight incubation at 4°C with primary antibodies diluted in
717 Blocking Buffer containing 0.2% Tween 20. Membranes were incubated with IRDye-conjugated
718 secondary antibodies (LI-COR Biosciences) for 1 h and fluorescent signal visualized using a
719 Sapphire Biomolecular Imager (Azure Biosystems) and Sapphire Capture software (Azure
720 Biosystems). When appropriate, membranes were stripped with Restore Western Blot Stripping
721 Buffer (ThermoFisher Scientific) before being re-probed. Primary antibodies include anti-UPF1
722 (Abcam, ab109363), anti-UPF2 (Cell Signaling, 11875S), anti-UPF3B (Abcam, ab134566), anti-
723 SMG6 (ABclonal, A10141), anti-CUL2 (Bethyl Laboratories, A302-476A), anti-ABCE1 (Abcam,
724 ab185548), anti-EIF5A (Abcam, ab32407), and anti-GAPDH (Abcam, ab9484). Secondary
725 antibodies used include IRDye 650 Goat anti-Mouse IgG Secondary Antibody (LI-COR
726 Biosciences, 926-65010) and IRDye 800CW Goat anti-Rabbit IgG Secondary Antibody (LI-COR
727 Biosciences, 926-32211).

728 **Luminescent reporters for *in vitro* assays**

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

742 **Expression and purification of eRF1 and eRF3**

743 Recombinant eRF1 was expressed from the plasmid pET-SUMO-eRF1. To obtain pET-
744 SUMO-eRF1, the eRF1 coding sequence was amplified from the plasmid pET23b-eRF1 (Frolova
745 et al., 2002) using primers petSUMO_eRF1_F
746 (GAGAACAGATTGGTGGTATGGCGGACGACCCCAG) and petSUMO_eRF1_R

747 (CCGAATAAATACCTAAGCTCTAGTAGTCATCAAGGTCAAAAAATTCATCGTCTCCTCC).
748 eRF1 was expressed in E. coli BL21(DE3) cells. The lysate was prepared by ultrasonification of
749 the pelleted cells in a buffer composed of 20 mM Tris-HCl pH 7.5, 500 mM KCl, 10% glycerol,
750 0.1% Triton X-100, 0.5 mM PMSF, and 1 mM DTT. Following this, His-SUMO tag was cleaved by
751 His-tagged Ulp1 protease. Untagged eRF1 was further purified by anion-exchange
752 chromatography (HiTrap Q HP, Cytiva). Fractions enriched with eRF1 were collected, dialyzed in
753 storage buffer, frizzed in liquid nitrogen, and stored at −70 ºC. Recombinant eRF3A cloned into
754 baculovirus vector EMBacY from a MultiBac expression system was expressed in the insect cell
755 line Sf21. Following this, recombinant proteins were purified using Ni-NTA agarose and ion-
756 exchange chromatography, as described previously (Ivanov et al., 2016).

757 **Expression, purification, and determination of luminescence or recombinant**
758 **nanoluciferase**

759 To obtain recombinant NLuc with varying amino acids in the position -1 to the stop codon,
760 corresponding constructs of pNL-globine were created using the petSUMO vector (Invitrogen, cat.
761 K30001). Each vector was assembled via Gibson Assembly (NEB, cat. E2611L) using PCR
762 products of the petSUMO backbone (forward primer:
763 AGCTTAGGTATTTATTCGGCGCAAAGTG; reverse primer:
764 ACCACCAATCTGTTCTCTGTGAGC) and pNL-globine inserts (forward primer:
765 GAGAACAGATTGGTGGTATGGTCTTCACACTCGAAGATTTCGTTGG; reverse primer:
766 CCGAATAAATACCTAAGCTTACGCCAGAATGCGTTCGCA) amplified using Q5 High-Fidelity
767 DNA Polymerase (NEB, cat. M0491L). The resulting plasmids were expressed in *E. coli*
768 BL21(DE3) and His-SUMO-Nluc was purified using Ni-affinity chromatography. Recombinant
769 6xHis-ULP1 protease was added to purify 6xHisSUMO-Nluc and incubated for 2 hours. The
770 6xHisSUMO fragment and 6xHis-ULP1 were then removed with Ni-NTA agarose to obtain purified
771 Nluc without tags. 0.1 femtomole of recombinant Nluc was incubated in storage buffer, with the
772 addition of BSA up to 500 µg/mL and 0.5 % NanoGlo (Promega). Luminescence was measured
773 at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) in a 40 min time period.
774 The luminescence of NLuc was calculated as a maximum of relative luminescence units (RLU_{max}).

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

776 For the Termi-Luc assay, preTCs translating NLuc (preTC-NLuc) were purified using
777 previously published methods (Shuvalov et al., 2021; Susorov et al., 2020). 100% RRL lysate
778 was preincubated in a mixture containing 1 mM $CaCl_2$ and 3 U/µL Micrococcal nuclease
779 (Fermentas) at 30°C for 10 min, followed by the addition of EGTA to a final concentration of 4

780 mM. The lysate was then diluted to 70% (v/v) and supplemented with 20 mM HEPES-KOH (pH
781 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
782 (Promega), 0.5 mM spermidine, 0.45 μM aminoacylated total rabbit tRNA, 10 mM creatine
783 phosphate, 0.003 U/μL creatine kinase (Sigma), 2 mM DTT, and 0.2 U/μL Ribolock
784 (ThermoFisher) (70% RRL mix).

785 For preTC-NLuc assembly, 220 μL of 70% RRL mix was preincubated in the presence of
786 1.7 μM ERF1 G183A mutant (Alkalaeva et al., 2006) at 30° C for 10 min, followed by the addition
787 of 10.5 pmol of NLuc mRNA. The mixture was incubated at 30°C for 40 min. The KOAc
788 concentration was then adjusted to 300 mM and the mixture was layered on 5 ml of a 10–35%
789 linear sucrose gradient in a buffer containing 50 mM HEPES-KOH, pH 7.5, 7.5 mM $Mg(OAc)_2$,
790 300 mM KOAc, 2 mM DTT. The gradient was centrifuged in a SW55-Ti (Beckman Coulter) rotor
791 at 55 000 rpm (367 598 g_{max}) for 1 h. The gradient was fractionated in 15 fractions of 150 μL from
792 bottom to top and the remaining sucrose was collected separately. Fractions of the first peak,
793 enriched with the preTC-Nluc were analyzed by the Termi-Luc assay, merged, flash-frozen in
794 liquid nitrogen, and stored at −70°C.

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

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

811 **Translation of NLuc in RRL lysates**

812 For translation of NLuc, 19 µL of 70% RRL lysate prepared as described in Termi-Luc
813 section, was mixed with 0.19 µL of NanoGlo and 0.25 pmol of NLuc mRNA. Luminescence was
814 measured at 30°C using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) for 60 min.
815 The translation efficiency was calculated as a maximal derivative of the growing linear section of
816 the luminescence curve (v_0, RLU/min).

817

818 **QUANTIFICATION AND STATISTICAL ANALYSIS**

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

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

829

SUPPLEMENTAL INFORMATION

SUPPLEMENTARY FIGURES

Figure S1

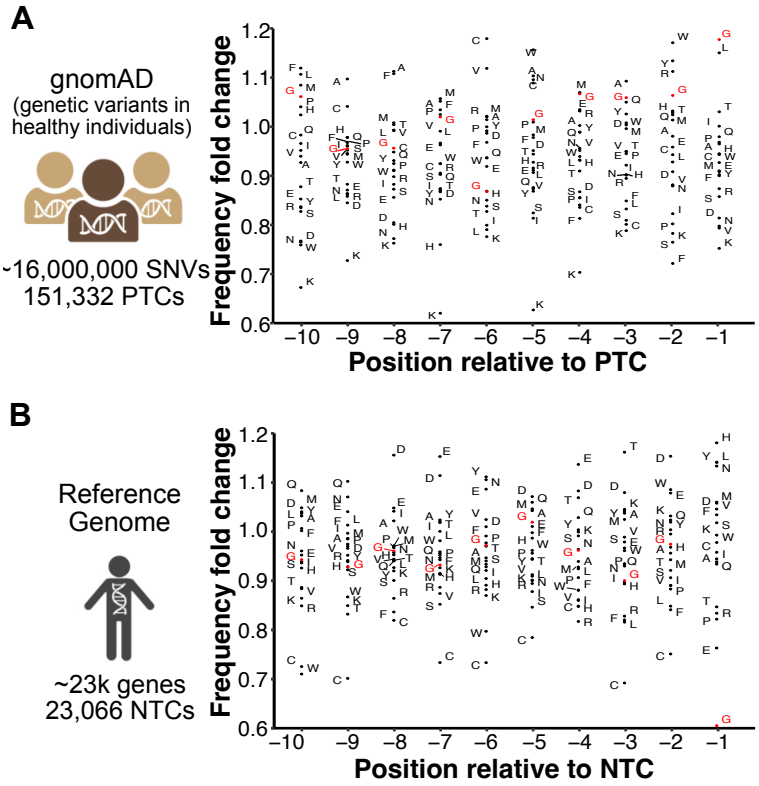
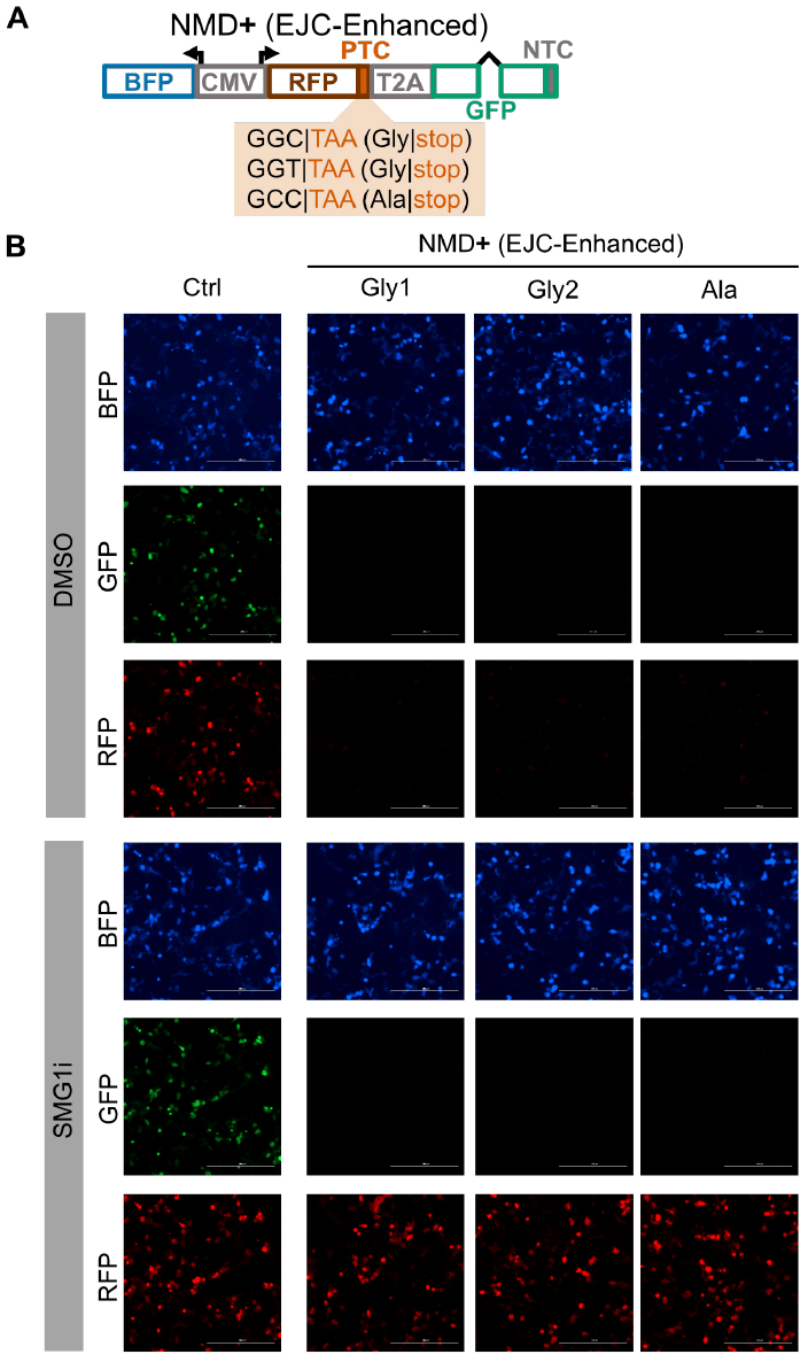


Figure S1. Amino acids enriched from the -10 to the -1 position before a termination codon. (**A**) Left: Schematic to indicate PTCs analysed in the gnomAD database, right: amino acid enrichment before the PTC. (**B**) Left: Schematic to indicate NTCs analyzed in the reference genome, right: amino acid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

843 **Figure S2**

844



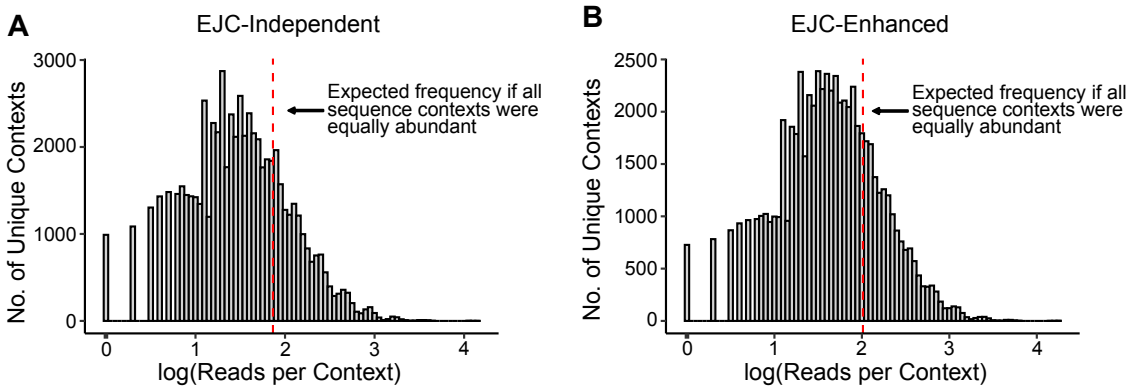
845

846 **Figure S2. EJC-enhanced reporters do not express RFP in any context.** (A) Schematic of
847 EJC-enhanced NMD reporter system used. (B) Fluorescence imaging of BFP, GFP, and RFP
848 from cells transfected with EJC-enhanced NMD reporters containing Gly-PTC or Ala-PTC
849 reporters, treated with a DMSO as a control or SMG1i. Ctrl refers to a construct that robustly

850 expresses all three fluorescent proteins. Gly1 and Gly2 are the GGC and GGT codons for Gly.
851 The codon used for Ala was GCC.

852

853 **Figure S3**

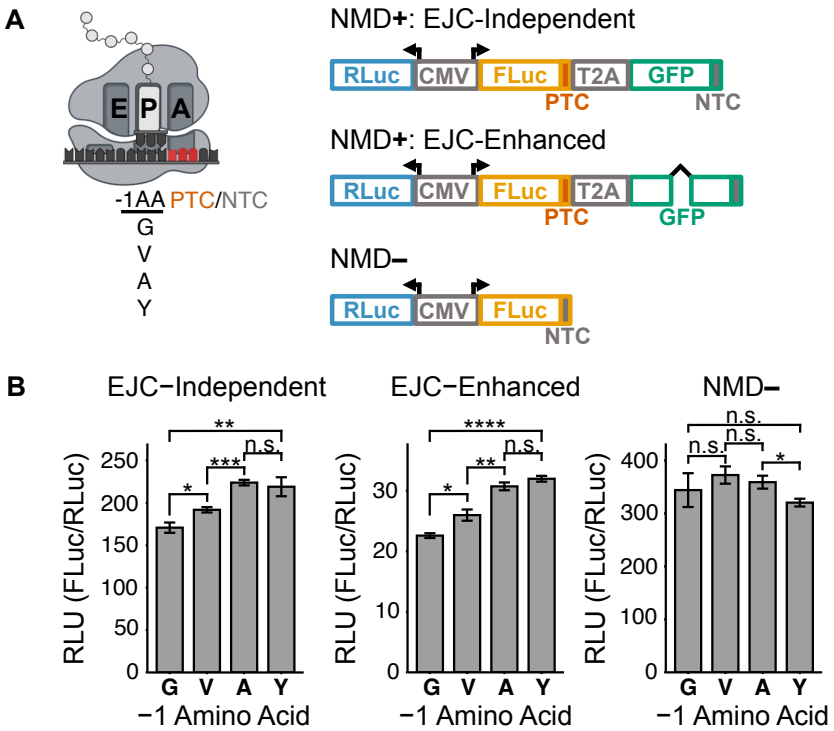


854

855 **Figure S3.** MPRA library representation. (**A**) Plasmid library representation for the EJC-
856 independent NMD**+** reporter library. (**B**) The same as (A) but for the EJC-enhanced NMD**+**
857 reporter library.

858

859 **Figure S4**



860

Figure S4. The codon identity at the -1 position of the PTC alters NMD activity of EJC-independent and EJC-enhanced luminescence-based reporters. (**A**) Schematic of luminescence based NMD reporters used (right) as well as the amino acids tested in the -1 position to the PTC/NTC (left). (**B**) Luciferase activity for each of the reporters in EJC-independent, EJC-enhanced, and NMD– contexts.

Figure S5

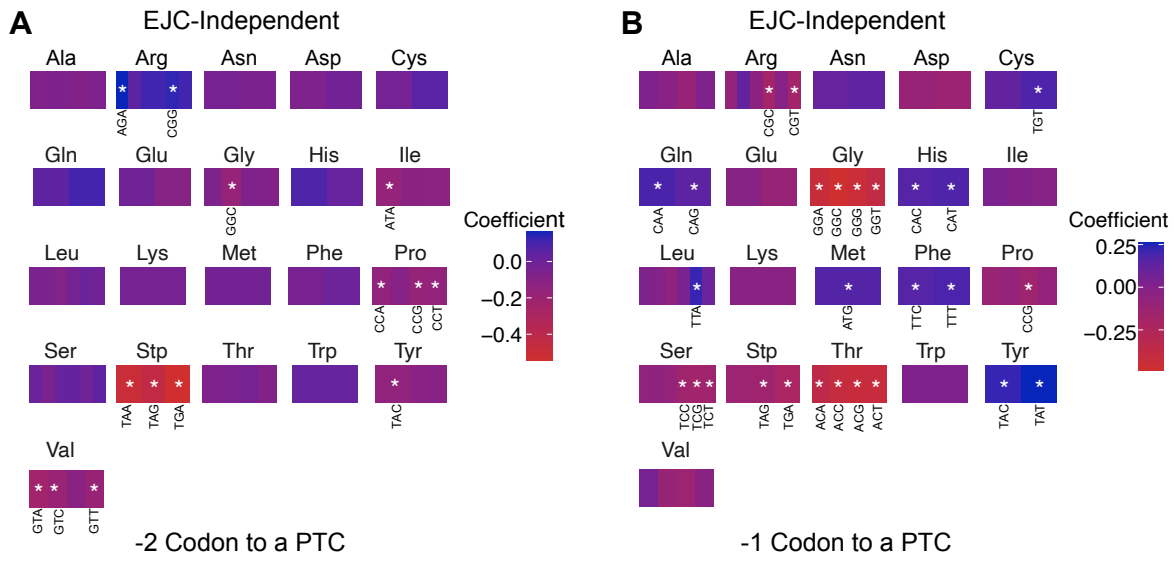
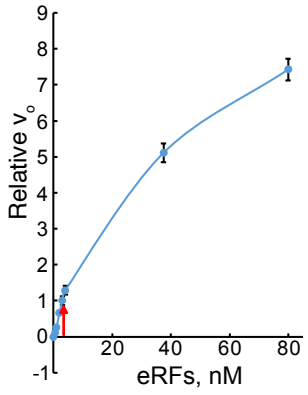


Figure S5. Codon based linear modeling of MPRA data. (**A**) Heat map depicting the linear modeling coefficients for the EJC-independent library for each codon -2 to a PTC. (**B**) The same as A but for the -1 codon.

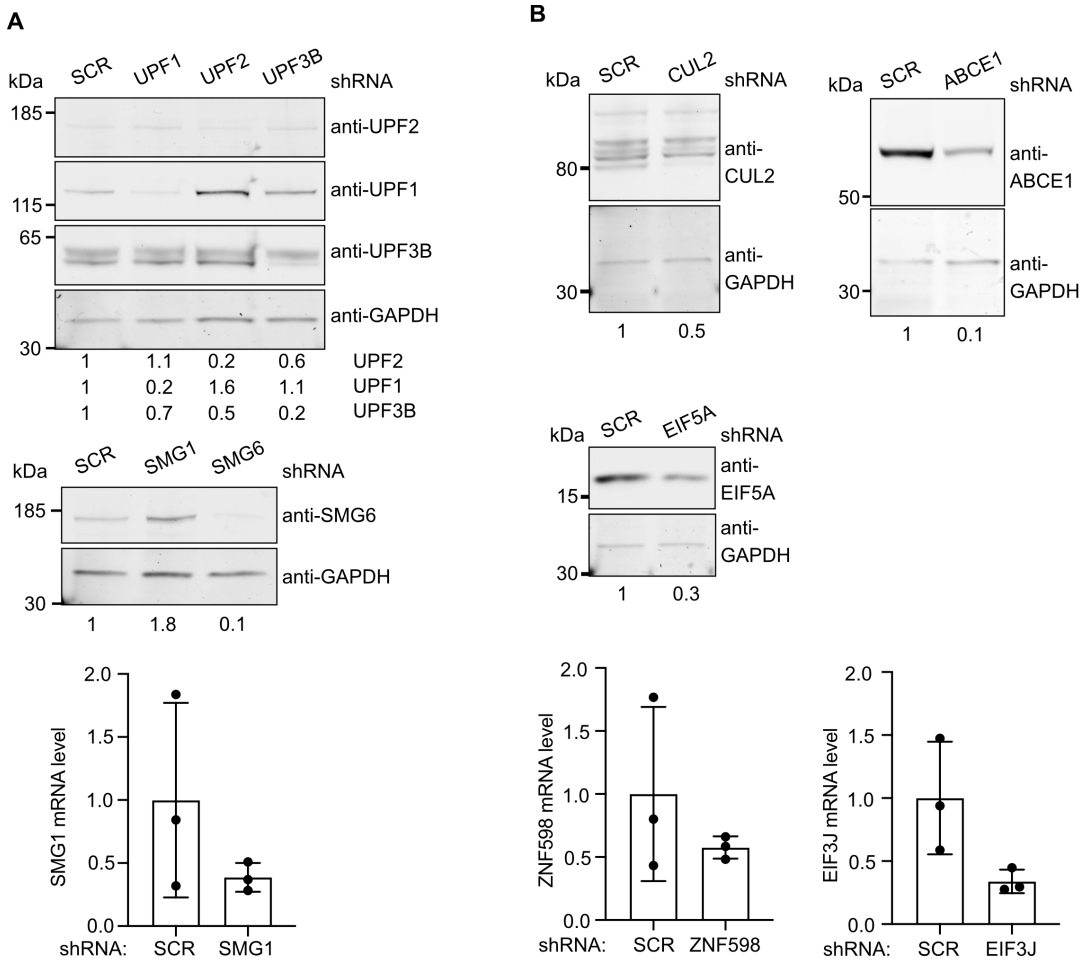
Figure S6.



875 **Figure S6. Determining the optimum concentration of release factors to use for the Termi-**
876 **Luc assay.**

877

878 **Figure S7.**



879

880 **Figure S7. Knockdown validations for all shRNAs used.** (**A**) Top: western blot validating
881 UPF1, UPF2, and UPF3B knockdowns; middle: western blot validating SMG6 knockdown;
882 bottom: qPCR validation of SMG1 knockdown. (**B**) Top left: western blot for CUL2 knockdown
883 validation; top right: western blot for ABCE1 knockdown validation; middle left: western blot for
884 EIF5A knockdown validation; bottom left: qPCR validation of ZNF598 knockdown; bottom right:
885 qPCR validation for EIF3J knockdown. GAPDH was the loading control for all western blots.
886 RPL27 was the housekeeping gene for all qPCRs. qPCRs were performed in triplicate.

887

888

SUPPLEMENTARY TABLES

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

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

892 **SUPPLEMENTARY DATA**

893

894 **Supplementary Table 1.** DEseq2 results for the EJC-independent library of sequences
895 representing the fold change of DMSO over SMG1i treatment of cells.
896
897 **Supplementary Table 2.** DEseq2 results for the EJC-enhanced library of sequences
898 representing the fold change of DMSO over SMG1i treatment of cells.

REFERENCES

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

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D. A., Ghosh, S., & Jacobson, A. (2004, Nov 4). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature, 432*(7013), 112-118. https://doi.org/10.1038/nature03060

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

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

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

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., & Izaurralde, E. (2007, Mar 21). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J, 26*(6), 1591-1601. https://doi.org/10.1038/sj.emboj.7601588

929 Bonetti, B., Fu, L., Moon, J., & Bedwell, D. M. (1995, Aug 18). The efficiency of translation
930 termination is determined by a synergistic interplay between upstream and downstream
931 sequences in Saccharomyces cerevisiae. *J Mol Biol, 251*(3), 334-345.
932 https://doi.org/10.1006/jmbi.1995.0438
933

934 Buhler, M., Steiner, S., Mohn, F., Paillusson, A., & Muhlemann, O. (2006, May). EJC-
935 independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR
936 length. *Nat Struct Mol Biol, 13*(5), 462-464. https://doi.org/10.1038/nsmb1081
937

938 Chu, V., Feng, Q., Lim, Y., & Shao, S. (2021, Dec 1). Selective destabilization of polypeptides
939 synthesized from NMD-targeted transcripts. *Mol Biol Cell, 32*(22), ar38.
940 https://doi.org/10.1091/mbc.E21-08-0382
941

942 Cridge, A. G., Crowe-McAuliffe, C., Mathew, S. F., & Tate, W. P. (2018, Feb 28). Eukaryotic
943 translational termination efficiency is influenced by the 3' nucleotides within the
944 ribosomal mRNA channel. *Nucleic Acids Res, 46*(4), 1927-1944.
945 https://doi.org/10.1093/nar/gkx1315
946

947 Egorova, T., Biziaev, N., Shuvalov, A., Sokolova, E., Mukba, S., Evmenov, K., Zotova, M.,
948 Kushchenko, A., Shuvalova, E., & Alkalaeva, E. (2021, Nov 8). eIF3j facilitates loading of
949 release factors into the ribosome. *Nucleic Acids Res, 49*(19), 11181-11196.
950 https://doi.org/10.1093/nar/gkab854
951

952 Frolova, L., Seit-Nebi, A., & Kisselev, L. (2002, Feb). Highly conserved NIKS tetrapeptide is
953 functionally essential in eukaryotic translation termination factor eRF1. *RNA, 8*(2), 129-
954 136. https://doi.org/10.1017/s1355838202013262
955

956 Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V.
957 M., Tatkov, S. I., & Kisselev, L. L. (1999, Aug). Mutations in the highly conserved GGQ
958 motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger
959 peptidyl-tRNA hydrolysis. *RNA, 5*(8), 1014-1020.
960 https://doi.org/10.1017/s135583829999043x
961

962 Gerbracht, J. V., Boehm, V., Britto-Borges, T., Kallabis, S., Wiederstein, J. L., Ciriello, S.,
963 Aschemeier, D. U., Kruger, M., Frese, C. K., Altmuller, J., Dieterich, C., & Gehring, N. H.
964 (2020, Sep 4). CASC3 promotes transcriptome-wide activation of nonsense-mediated
965 decay by the exon junction complex. *Nucleic Acids Res, 48*(15), 8626-8644.
966 https://doi.org/10.1093/nar/gkaa564
967

968 Grosjean, H., & Westhof, E. (2016, Sep 30). An integrated, structure- and energy-based view of
969 the genetic code. *Nucleic Acids Res, 44*(17), 8020-8040.
970 https://doi.org/10.1093/nar/gkw608
971

972 Hellen, C. U. T. (2018, Oct 1). Translation Termination and Ribosome Recycling in Eukaryotes.
973 *Cold Spring Harb Perspect Biol, 10*(10). https://doi.org/10.1101/cshperspect.a032656
974

975 Hogg, J. R., & Goff, S. P. (2010, Oct 29). Upf1 senses 3'UTR length to potentiate mRNA decay.
976 *Cell, 143*(3), 379-389. https://doi.org/10.1016/j.cell.2010.10.005
977

978 Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A.,
979 Schaffitzel, C., & Alkalaeva, E. (2016, Sep 19). PABP enhances release factor
980 recruitment and stop codon recognition during translation termination. *Nucleic Acids*
981 *Res, 44*(16), 7766-7776. https://doi.org/10.1093/nar/gkw635
982

983 Ivanov, P. V., Gehring, N. H., Kunz, J. B., Hentze, M. W., & Kulozik, A. E. (2008, Mar 5).
984 Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an
985 integrated model for mammalian NMD pathways. *EMBO J, 27*(5), 736-747.
986 https://doi.org/10.1038/emboj.2008.17
987

988 Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2012). Termination and post-termination events
989 in eukaryotic translation. *Adv Protein Chem Struct Biol, 86*, 45-93.
990 https://doi.org/10.1016/B978-0-12-386497-0.00002-5
991

992 Jagannathan, S., & Bradley, R. K. (2016, Dec). Translational plasticity facilitates the
993 accumulation of nonsense genetic variants in the human population. *Genome Res,*
994 *26*(12), 1639-1650. https://doi.org/10.1101/gr.205070.116
995

996 Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V., & Hegde, R. S.
997 (2018, Nov 1). ZNF598 Is a Quality Control Sensor of Collided Ribosomes. *Mol Cell,*
998 *72*(3), 469-481 e467. https://doi.org/10.1016/j.molcel.2018.08.037
999

1000 Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., Collins, R.
1001 L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson,
1002 M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki,
1003 J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed,
1004 C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z.,
1005 O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C.,
1006 Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly,
1007 S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne,
1008 C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A.,
1009 Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome
1010 Aggregation Database, C., Neale, B. M., Daly, M. J., & MacArthur, D. G. (2020, May).
1011 The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature,*
1012 *581*(7809), 434-443. https://doi.org/10.1038/s41586-020-2308-7

1013

1014 Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M.,
1015 Dreyfuss, G., & Ohno, S. (2006, Feb 1). Binding of a novel SMG-1-Upf1-eRF1-eRF3
1016 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and
1017 nonsense-mediated mRNA decay. *Genes Dev, 20*(3), 355-367.
1018 https://doi.org/10.1101/gad.1389006
1019

1020 Khandelia, P., Yap, K., & Makeyev, E. V. (2011, Aug 2). Streamlined platform for short hairpin
1021 RNA interference and transgenesis in cultured mammalian cells. *Proc Natl Acad Sci U S*
1022 *A, 108*(31), 12799-12804. https://doi.org/10.1073/pnas.1103532108
1023

1024 Kolakada, D., Campbell, A. E., Galvis, L. B., Li, Z., Lore, M., & Jagannathan, S. (2023). A
1025 system of reporters for comparative investigation of EJC-independent and EJC-
1026 enhanced nonsense-mediated mRNA decay. *bioRxiv*, 2023.2011.2014.567061.
1027 https://doi.org/10.1101/2023.11.14.567061
1028

1029 Kolakada, D., Campbell, A. E., Galvis, L. B., Li, Z., Lore, M., & Jagannathan, S. (2024, Feb 20).
1030 A system of reporters for comparative investigation of EJC-independent and EJC-
1031 enhanced nonsense-mediated mRNA decay. *Nucleic Acids Res*.
1032 https://doi.org/10.1093/nar/gkae121
1033

1034 Koren, I., Timms, R. T., Kula, T., Xu, Q., Li, M. Z., & Elledge, S. J. (2018, Jun 14). The
1035 Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons.
1036 *Cell, 173*(7), 1622-1635 e1614. https://doi.org/10.1016/j.cell.2018.04.028
1037

1038 Kurosaki, T., & Maquat, L. E. (2013, Feb 26). Rules that govern UPF1 binding to mRNA 3'
1039 UTRs. *Proc Natl Acad Sci U S A, 110*(9), 3357-3362.
1040 https://doi.org/10.1073/pnas.1219908110

1041

1042 Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J.,
1043 Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A.,
1044 McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., Kattman, B. L., & Maglott,
1045 D. R. (2018, Jan 4). ClinVar: improving access to variant interpretations and supporting
1046 evidence. *Nucleic Acids Res, 46*(D1), D1062-D1067.
1047 https://doi.org/10.1093/nar/gkx1153
1048

1049 Langmead, B., & Salzberg, S. L. (2012, Mar 4). Fast gapped-read alignment with Bowtie 2. *Nat*
1050 *Methods, 9*(4), 357-359. https://doi.org/10.1038/nmeth.1923
1051

1052 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1053 Durbin, R., & Genome Project Data Processing, S. (2009, Aug 15). The Sequence
1054 Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.
1055 https://doi.org/10.1093/bioinformatics/btp352
1056

1057 Lin, H. C., Yeh, C. W., Chen, Y. F., Lee, T. T., Hsieh, P. Y., Rusnac, D. V., Lin, S. Y., Elledge,
1058 S. J., Zheng, N., & Yen, H. S. (2018, May 17). C-Terminal End-Directed Protein
1059 Elimination by CRL2 Ubiquitin Ligases. *Mol Cell, 70*(4), 602-613 e603.
1060 https://doi.org/10.1016/j.molcel.2018.04.006
1061

1062 Lindeboom, R. G., Supek, F., & Lehner, B. (2016, Oct). The rules and impact of nonsense-
1063 mediated mRNA decay in human cancers. *Nat Genet, 48*(10), 1112-1118.
1064 https://doi.org/10.1038/ng.3664
1065

1066 Lindeboom, R. G. H., Vermeulen, M., Lehner, B., & Supek, F. (2019, Nov). The impact of
1067 nonsense-mediated mRNA decay on genetic disease, gene editing and cancer

immunotherapy. *Nat Genet, 51*(11), 1645-1651. https://doi.org/10.1038/s41588-019-0517-5

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

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

Loughran, G., Li, X., O'Loughlin, S., Atkins, J. F., & Baranov, P. V. (2023, Jan 11). Monitoring translation in all reading frames downstream of weak stop codons provides mechanistic insights into the impact of nucleotide and cellular contexts. *Nucleic Acids Res, 51*(1), 304-314. https://doi.org/10.1093/nar/gkac1180

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

Lykke-Andersen, J., Shu, M. D., & Steitz, J. A. (2001, Sep 7). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science, 293*(5536), 1836-1839. https://doi.org/10.1126/science.1062786

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J. K., Montgomery, S. B., Albers, C. A., Zhang, Z. D., Conrad, D. F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M. A., Banks, E., Hu, M., Handsaker, R.

1096 E., Rosenfeld, J. A., Fromer, M., Jin, M., Mu, X. J., Khurana, E., Ye, K., Kay, M.,
1097 Saunders, G. I., Suner, M. M., Hunt, T., Barnes, I. H., Amid, C., Carvalho-Silva, D. R.,
1098 Bignell, A. H., Snow, C., Yngvadottir, B., Bumpstead, S., Cooper, D. N., Xue, Y.,
1099 Romero, I. G., Genomes Project, C., Wang, J., Li, Y., Gibbs, R. A., McCarroll, S. A.,
1100 Dermitzakis, E. T., Pritchard, J. K., Barrett, J. C., Harrow, J., Hurles, M. E., Gerstein, M.
1101 B., & Tyler-Smith, C. (2012, Feb 17). A systematic survey of loss-of-function variants in
1102 human protein-coding genes. *Science, 335*(6070), 823-828.
1103 https://doi.org/10.1126/science.1215040
1104

1105 Mangkalaphiban, K., He, F., Ganesan, R., Wu, C., Baker, R., & Jacobson, A. (2021, Apr).
1106 Transcriptome-wide investigation of stop codon readthrough in Saccharomyces
1107 cerevisiae. *PLoS Genet, 17*(4), e1009538. https://doi.org/10.1371/journal.pgen.1009538
1108

1109 Martin, M. (2011, 2011-05-02). Cutadapt removes adapter sequences from high-throughput
1110 sequencing reads [next generation sequencing; small RNA; microRNA; adapter
1111 removal]. *2011, 17*(1), 3. https://doi.org/10.14806/ej.17.1.200
1112

1113 Mazo-Vargas, A., Park, H., Aydin, M., & Buchler, N. E. (2014, Nov 5). Measuring fast gene
1114 dynamics in single cells with time-lapse luminescence microscopy. *Mol Biol Cell, 25*(22),
1115 3699-3708. https://doi.org/10.1091/mbc.E14-07-1187
1116

1117 McCaughan, K. K., Brown, C. M., Dalphin, M. E., Berry, M. J., & Tate, W. P. (1995, Jun 6).
1118 Translational termination efficiency in mammals is influenced by the base following the
1119 stop codon. *Proc Natl Acad Sci U S A, 92*(12), 5431-5435.
1120 https://doi.org/10.1073/pnas.92.12.5431
1121

1122 Miller, J. N., & Pearce, D. A. (2014, Oct-Dec). Nonsense-mediated decay in genetic disease:
1123 friend or foe? *Mutat Res Rev Mutat Res, 762*, 52-64.
1124 https://doi.org/10.1016/j.mrrev.2014.05.001

Mort, M., Ivanov, D., Cooper, D. N., & Chuzhanova, N. A. (2008, 2008 Aug). A meta-analysis of nonsense mutations causing human genetic disease. *Human mutation, 29*(8). https://doi.org/10.1002/humu.20763

Nagy, E., & Maquat, L. E. (1998, Jun). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci, 23*(6), 198-199. https://doi.org/10.1016/s0968-0004(98)01208-0

Neu-Yilik, G., Amthor, B., Gehring, N. H., Bahri, S., Paidassi, H., Hentze, M. W., & Kulozik, A. E. (2011, May). Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. *RNA, 17*(5), 843-854. https://doi.org/10.1261/rna.2401811

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

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

Sato, H., & Singer, R. H. (2021, Dec 10). Cellular variability of nonsense-mediated mRNA decay. *Nat Commun, 12*(1), 7203. https://doi.org/10.1038/s41467-021-27423-0

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

1152 mediated decay pathway. *J Biol Chem, 281*(46), 35336-35346.
1153 https://doi.org/10.1074/jbc.M601460200
1154

1155 Schuller, A. P., Wu, C. C., Dever, T. E., Buskirk, A. R., & Green, R. (2017, Apr 20). eIF5A
1156 Functions Globally in Translation Elongation and Termination. *Mol Cell, 66*(2), 194-205
1157 e195. https://doi.org/10.1016/j.molcel.2017.03.003
1158

1159 Shuvalov, A., Shuvalova, E., Biziaev, N., Sokolova, E., Evmenov, K., Pustogarov, N.,
1160 Arnautova, A., Matrosova, V., Egorova, T., & Alkalaeva, E. (2021, Nov 12). Nsp1 of
1161 SARS-CoV-2 stimulates host translation termination. *RNA Biol, 18*(sup2), 804-817.
1162 https://doi.org/10.1080/15476286.2021.1999103
1163

1164 Singh, G., Rebbapragada, I., & Lykke-Andersen, J. (2008, Apr 29). A competition between
1165 stimulators and antagonists of Upf complex recruitment governs human nonsense-
1166 mediated mRNA decay. *PLoS Biol, 6*(4), e111.
1167 https://doi.org/10.1371/journal.pbio.0060111
1168

1169 Smith, T., Heger, A., & Sudbery, I. (2017, Mar). UMI-tools: modeling sequencing errors in
1170 Unique Molecular Identifiers to improve quantification accuracy. *Genome Res, 27*(3),
1171 491-499. https://doi.org/10.1101/gr.209601.116
1172

1173 Susorov, D., Egri, S., & Korostelev, A. A. (2020, Dec). Termi-Luc: a versatile assay to monitor
1174 full-protein release from ribosomes. *RNA, 26*(12), 2044-2050.
1175 https://doi.org/10.1261/rna.076588.120
1176

1177 Tate, W. P., & Mannering, S. A. (1996, Jul). Three, four or more: the translational stop signal at
1178 length. *Mol Microbiol, 21*(2), 213-219. https://doi.org/10.1046/j.1365-
1179 2958.1996.6391352.x

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

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

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

Zhang, J., & Maquat, L. E. (1997, Feb 17). Evidence that translation reinitiation abrogates nonsense-mediated mRNA decay in mammalian cells. *EMBO J, 16*(4), 826-833. 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