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

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

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

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

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

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

[5]Lead contact

*Corresponding Author: sujatha.jagannathan@cuanschutz.edu

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

Keywords: NMD, PTC, translation, termination, splicing

SUMMARY

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

INTRODUCTION

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

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

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

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

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

RESULTS

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

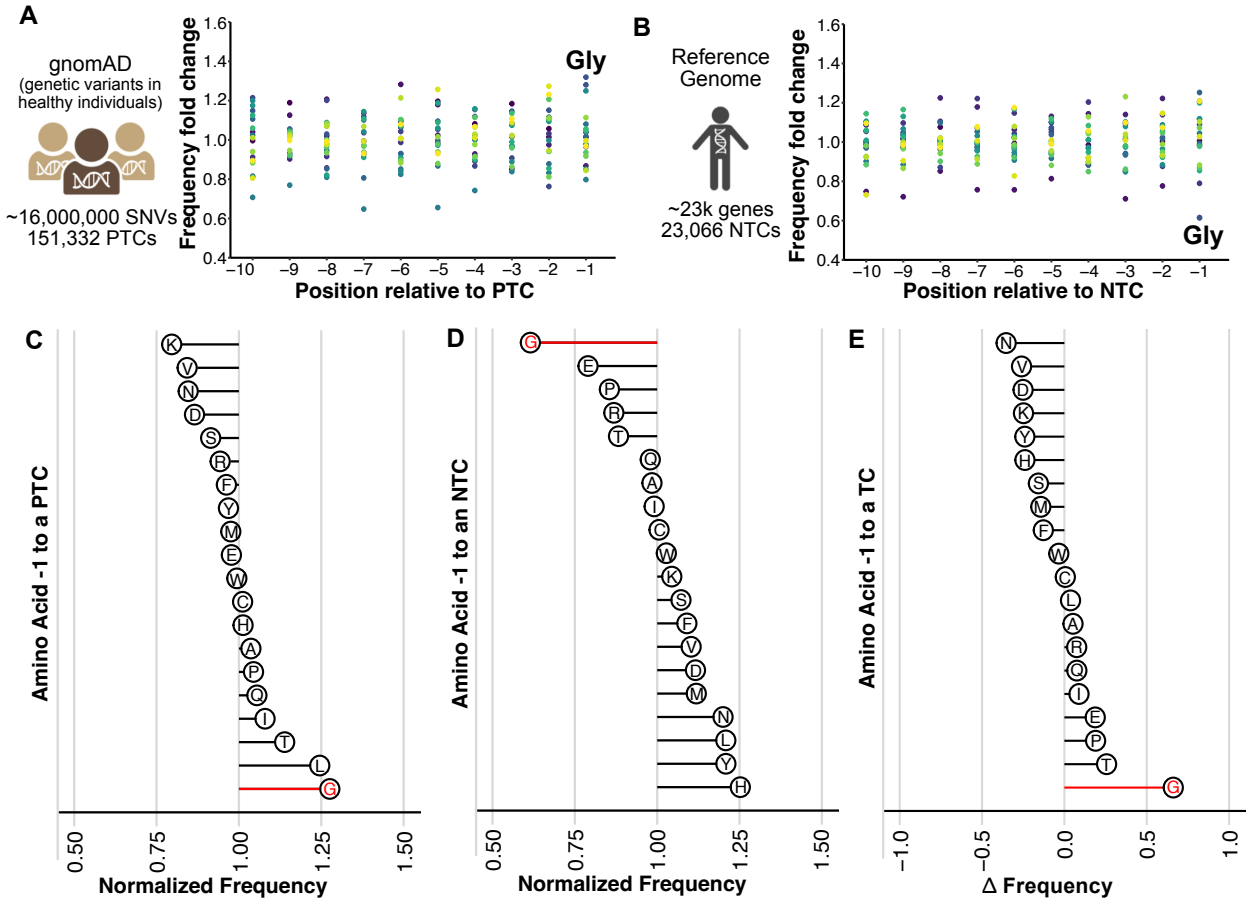


Figure 1. Gly is enriched preceding a PTC and depleted preceding an NTC. (**A-B**) The enrichment of amino acids in the 10 positions before a PTC (A) or the NTC (B) (**C-D**) Amino acid enrichment -1 to a PTC (C) or an NTC (D). (**E**) The frequency delta (enrichment before a PTC-NTC) for each amino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

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

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

133 If Gly-PTC promotes readthrough, we would expect to see production of GFP protein. If Gly-
134 PTC promotes efficient protein degradation, then we would expect to see much less RFP protein
135 accumulation in both NMD**+** and NMD**–** contexts compared to the Ala-PTC context. If Gly-PTC
136 promotes efficient NMD, we would expect to see less RFP protein levels compared to Ala-PTC
137 for the NMD**+** contexts but no difference for the NMD**–** contexts (**Fig 2A**, right). We transfected
138 HEK293T cells with these reporters, treating them with either DMSO vehicle control or SMG1i, an
139 inhibitor of the SMG1 kinase activity, and thus NMD. Transfected cells were imaged for BFP,
140 RFP, and GFP. We also included a control construct that lacks the stop codon between the RFP
141 and GFP, and therefore robustly expresses all three fluorescent proteins.
142 As shown in Fig 2B, all reporters yielded robust BFP expression. While the no PTC control
143 construct produced GFP, there was no GFP protein detected with either the glycine or alanine
144 PTC contexts (**Fig 2B**), ruling out the first hypothesis that a Gly-PTC context promotes
145 readthrough. Notably, both EJC-independent NMD**+** reporters with Gly-PTC resulted in less
146 RFP protein compared to the Ala-PTC reporter, a difference which recovered upon treatment
147 with SMG1i (quantified in **Fig 2C**). In comparison, only a modest reduction in in RFP protein
148 was observed for the Gly-PTC NMD**–** reporters compared to the Ala-PTC reporter (**Fig 2B**),
149 suggesting that a C-end glycine degron was not sufficient for the protein depletion we observed
150 in the NMD+ reporters. No RFP was detected from any of the EJC-enhanced NMD reporters,
151 consistent with their high NMD activity (**Supp. Fig 1B-C**). These data suggest that Gly-PTC
152 inherently influences NMD efficiency through a hitherto unknown mechanism.

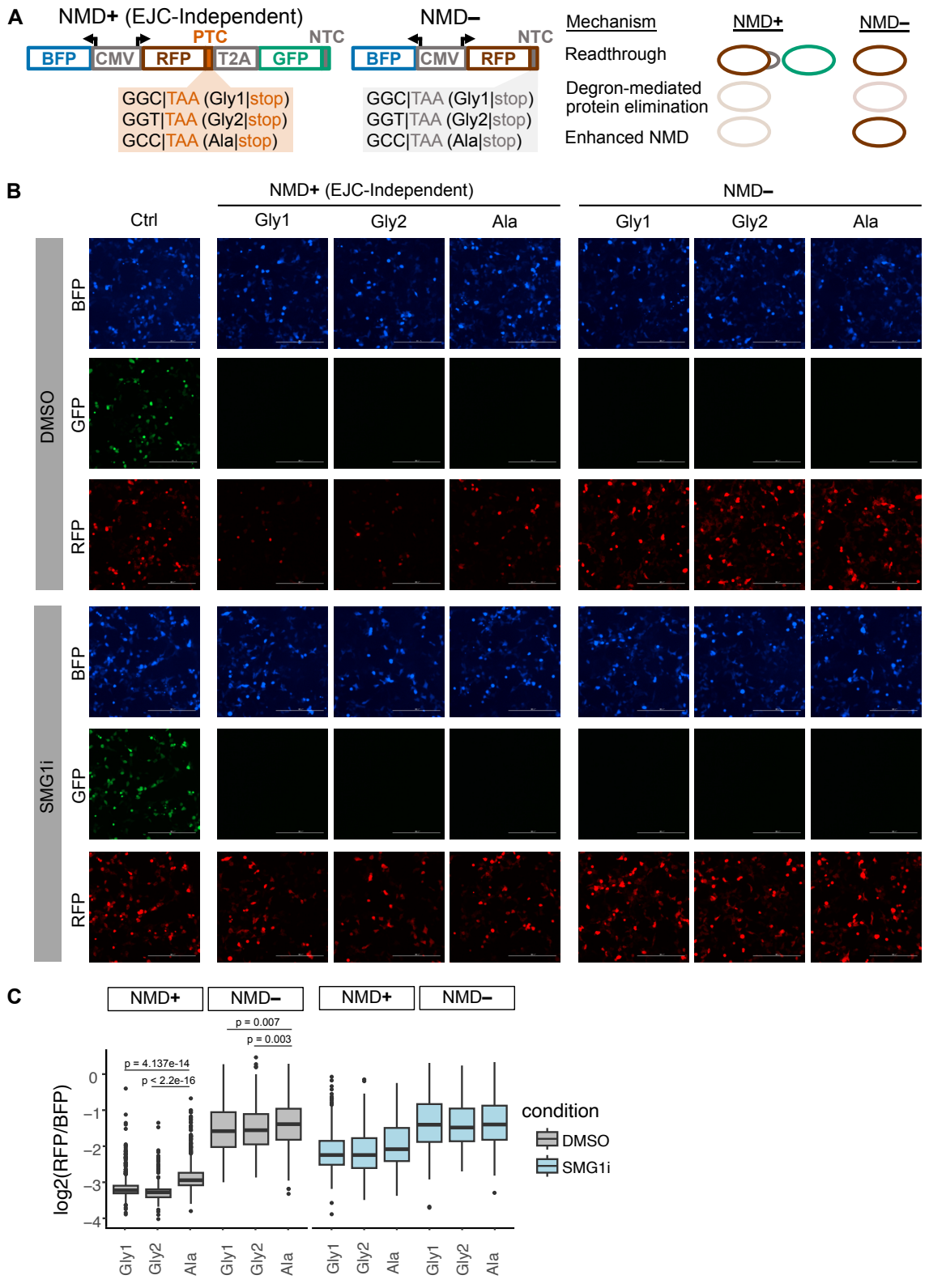


Figure 2. Gly-PTC contexts enhance NMD. (A) Schematic of NMD reporter system used. Left: EJC-independent NMD+ reporter, middle: NMD– reporter; right: expected outcomes. (B)

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

161

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

180 The NMD activity as measured by transcript fold changes for both the EJC-independent and
181 EJC-enhanced reporters had a bimodal distribution, corresponding to the stop codon identity of
182 the sequences in each population (**Fig. 3B**): sequences containing a stop codon were left-
183 shifted compared to the sequences containing the Trp control, indicating that the former were
184 targets of NMD. In the population of sequences undergoing NMD (i.e. those containing a stop
185 codon) there was a wide range of NMD activity for both the EJC-independent and the EJC-
186 enhanced reporters. EJC-enhanced reporter expression was reduced by 4.8 fold (log2FC: -
187 2.27), which is greater than the 3.3 fold decrease (log2FC: -1.71) for EJC-independent reporters
188 (**Fig. 3B**), consistent with EJC-enhanced targets undergoing more efficient NMD compared to

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

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

201 As expected, our modeling showed that all stop codons caused significantly stronger NMD
202 activity compared to the Trp control for both EJC-independent and EJC-enhanced reporters (**Fig.**
203 **3C**). Notably, the TGA stop codon increased NMD efficiency by ~20% compared to the TAG stop
204 codon for the EJC-independent reporters. However, this difference was not observed for the EJC-
205 enhanced reporters where all stop codons elicit similar levels of NMD activity (**Fig. 3C**). For
206 subsequent analyses, only transcripts expected to undergo NMD (i.e. transcripts with stop
207 codons) were analyzed. When determining the effect of the +4 nt identity on NMD, we found that
208 T, G, and C promoted weaker (~6% decrease) NMD activity compared to A for the EJC-
209 independent reporters. A similar effect was present only for G compared to A for the EJC-
210 enhanced reporters (**Fig. 3D**).

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

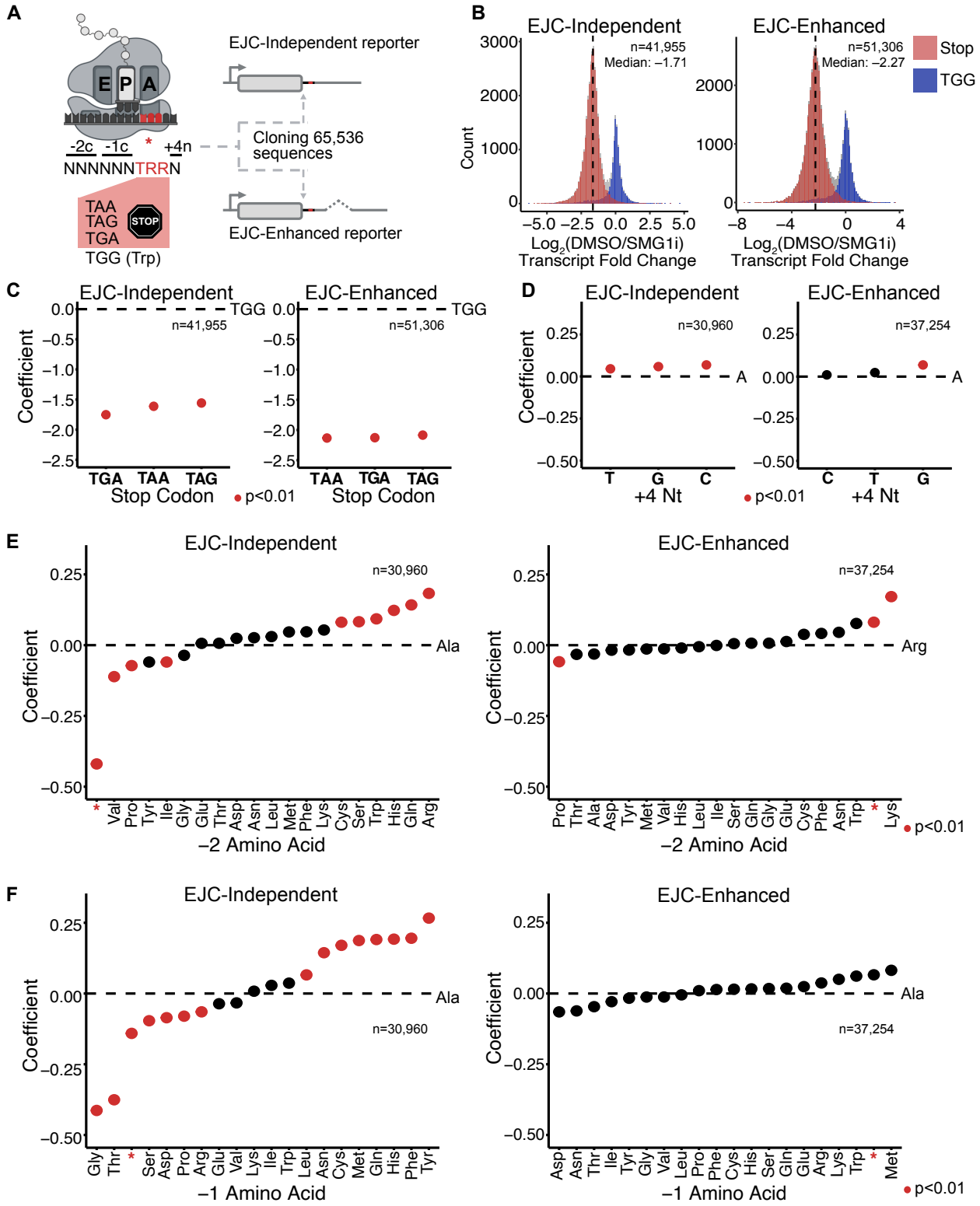


Figure 1. MPRA shows a wide range of PTC context-dependent NMD efficiencies. (A) Design of the MPRA. (B) Histogram of transcript fold change EJC-independent and EJC-enhanced NMD+ reporters comparing DMSO treated cells to SMG1i treated cells. (C-F) Dot

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

227

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

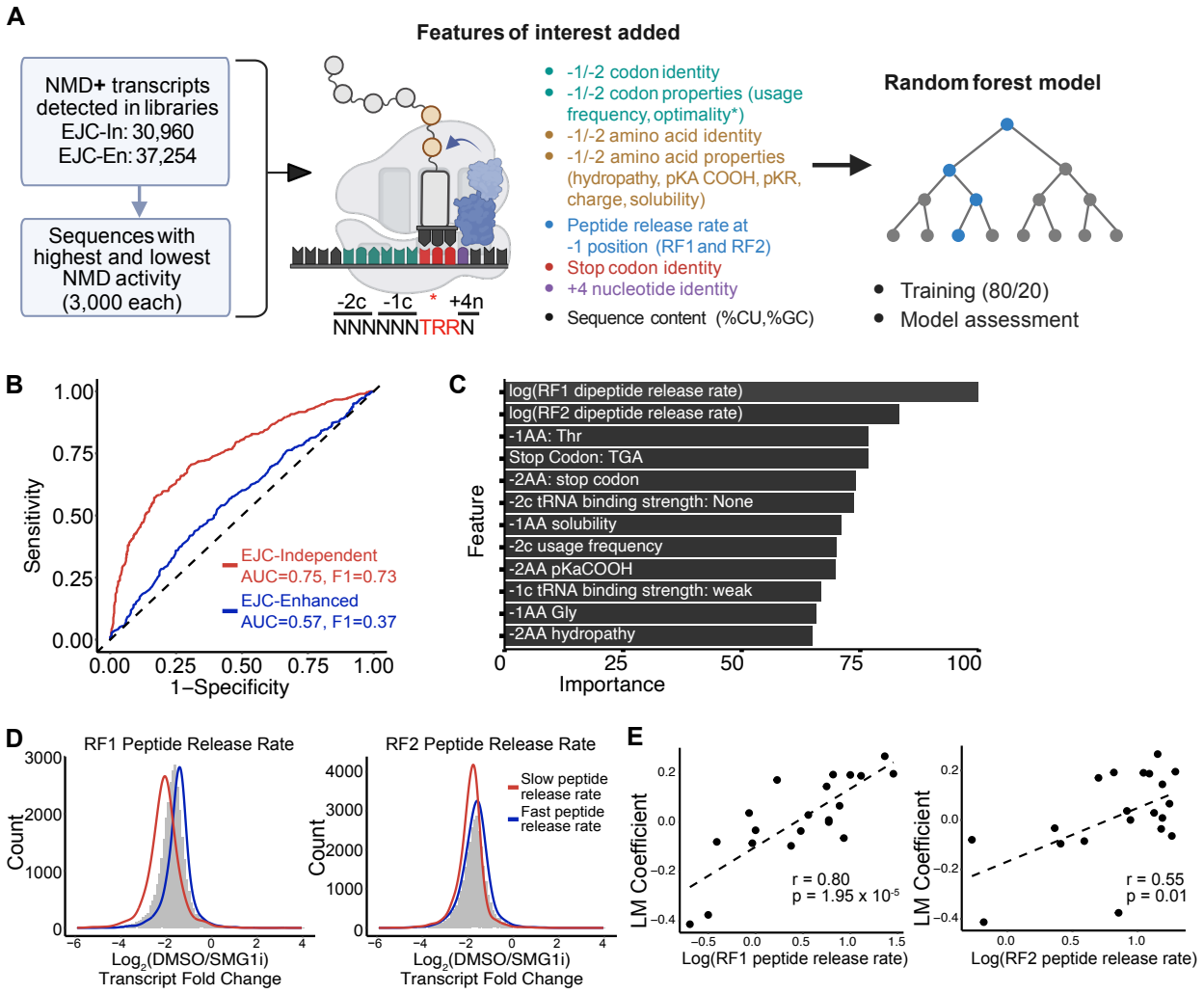
235 Next, we asked if there is codon-level specificity to the effect of amino acids on EJC-
236 independent NMD activity. We found that the effects of the codons in the -2 and -1 positions were
237 largely grouped by amino acid identity, with only a few that were significantly different (**Supp. Fig.**
238 **2C-D**, all codons compared to Lys-AAA). An exception to this trend was Leu at the -1 position
239 whose codon TTA seems to have the opposite effect on NMD compared to the rest of the Leu
240 codons. Reassuringly, all the effects at the amino acid level were replicated at the codon level
241 with the stop codons having the strongest NMD activity at the -2 position and Arg the weakest
242 (**Supp. Fig. 2C**), and Gly having the strongest NMD activity at the -1 position and Tyr the weakest
243 (**Supp. Fig. 2D**). Taken together these data suggest that the -2 and primarily the -1 amino acid
244 identity are key in determining the efficiency of EJC-independent NMD.

245

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

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

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



258

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

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

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

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

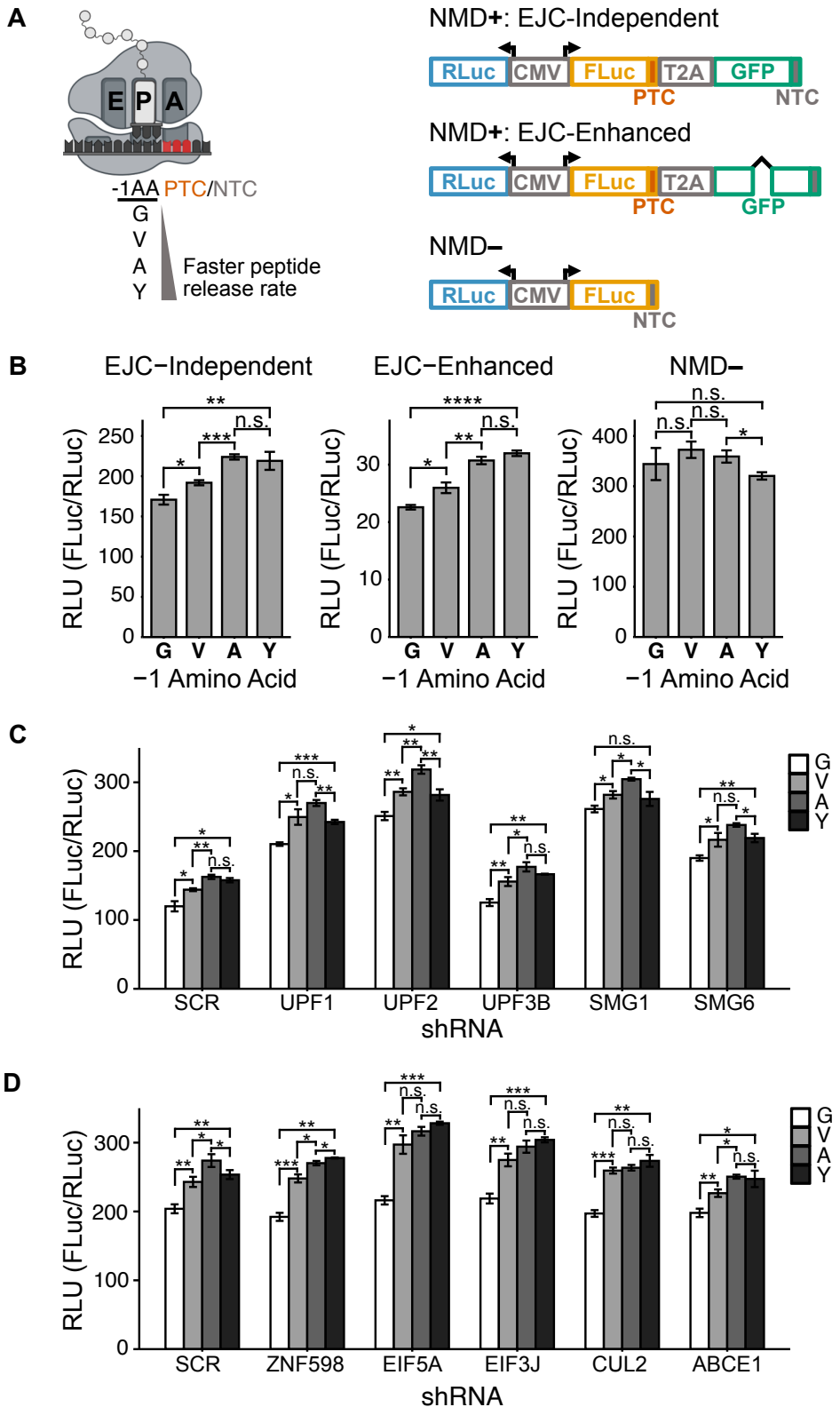
299 **Luminescence-based NMD reporters recapitulate -1 amino acid-dependent variation in**
300 **NMD activity in EJC-independent and EJC-enhanced contexts.** Results thus far suggest
301 that peptide release rate is a major influencer of NMD, but only in the context of EJC-
302 independent NMD and not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

308 To test whether the -1 amino acid effect was also present within this NMD reporter system
309 we varied the amino acid in the -1 position of either the PTC or the NTC, choosing Gly, Val, Ala,
310 and Tyr, which have a range of peptide release rates from slowest to fastest (**Fig. 5A**). As
311 expected, luciferase activity was ~10-fold lower for the EJC-enhanced reporters compared to
312 EJC-independent (**Fig. 5B**). Remarkably, the relative trend in -1 amino acid-dependent NMD
313 variability was the same for both the EJC-independent and EJC-enhanced reporters. The NMD
314 activity varied dependent on the -1 amino acid, with Gly having ~22% or ~29% more efficient
315 NMD compared to Tyr for the EJC-independent and EJC-enhanced reporters, respectively (**Fig**
316 **5B**). Compared to the EJC-independent and EJC-enhanced reporters, the NMD**–** reporters did
317 not vary with the change in the -1 amino acid (**Fig 5B**). Taken together, these data demonstrate
318 that the -1 amino acid effect can extend to both forms of NMD.

319

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

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


A
E P A
-1AA PTC/NTC
G
V
A
Y
Faster peptide release rate
NMD+: EJC-Independent
RLuc CMV FLuc T2A GFP
PTC NTC
NMD+: EJC-Enhanced
RLuc CMV FLuc T2A GFP
PTC
NMD−
RLuc CMV FLuc
NTC
B
EJC−Independent
EJC−Enhanced
NMD−
RLU (FLuc/RLuc)
G V A Y
−1 Amino Acid
C
RLU (FLuc/RLuc)
SCR UPF1 UPF2 UPF3B SMG1 SMG6
shRNA
G V A Y
D
RLU (FLuc/RLuc)
SCR ZNF598 EIF5A EIF3J CUL2 ABCE1
shRNA
G V A Y

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

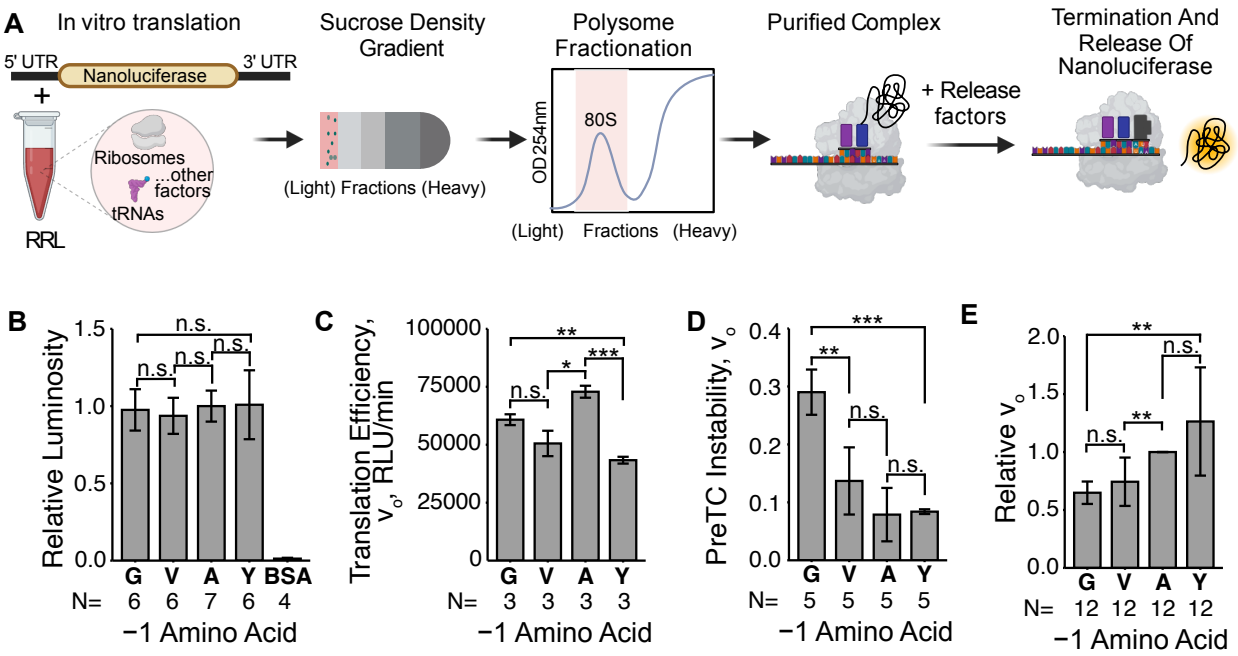
341

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

358

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

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



374

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

382

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

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

401

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

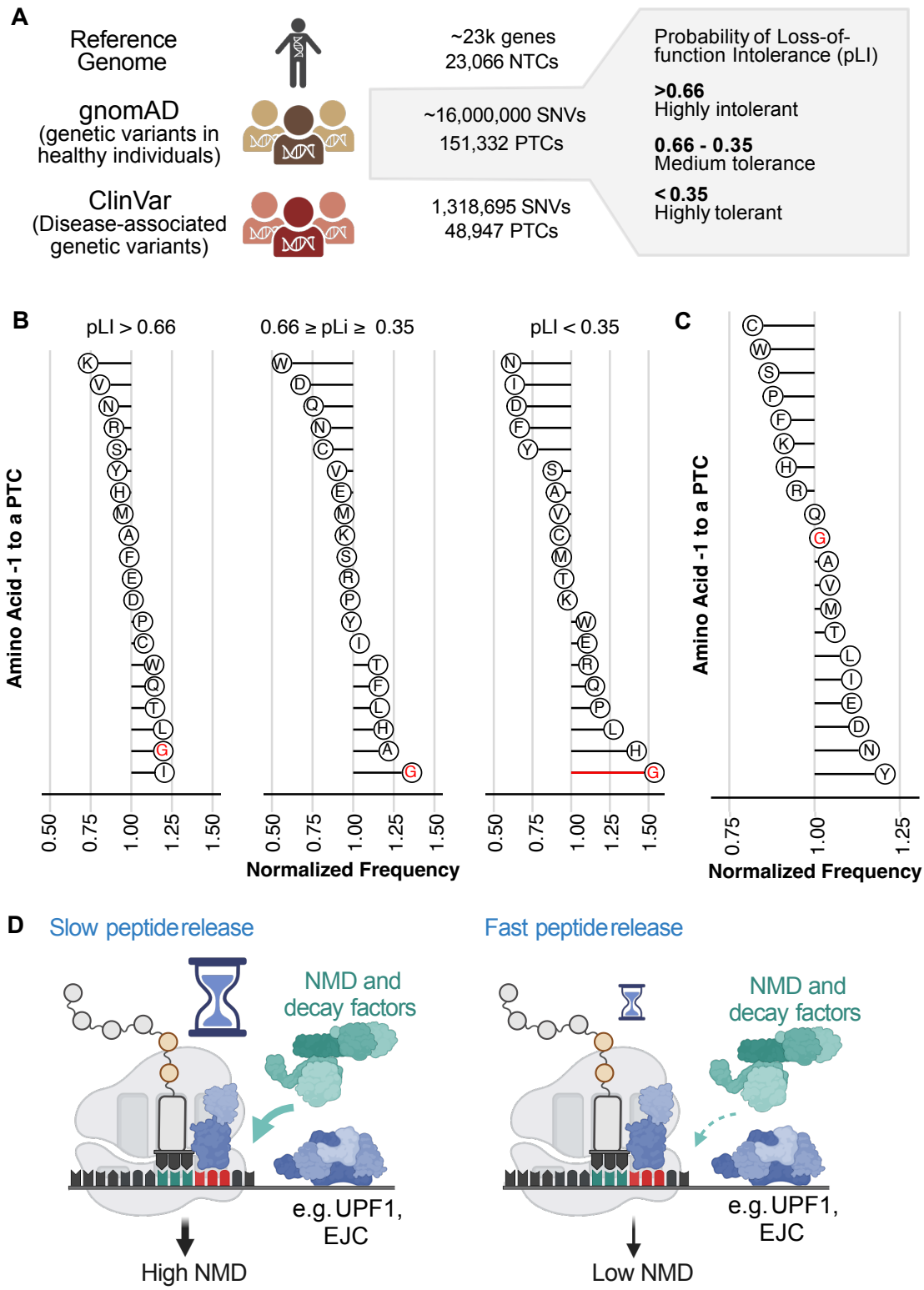


Figure 7. PTVs with Gly -1 to a PTC are tolerant to LoF and not enriched in genetic diseases. (A) Schematic for the analysis of LoF scores for rare PTVs in healthy individuals as well as PTVs in disease contexts. (B) The -1 amino acid enrichment of PTVs binned into 3

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

427 **DISCUSSION**

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

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

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

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

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

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

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

480

481

ACKNOWLEDGEMENTS

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

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

506 **STAR METHODS**

507

508 **RESOURCE AVAILABILITY**

509

510 **Lead contact**

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

513

514 **Materials availability**

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

517

518 **Data and code availability**

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

525

526 **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

527

528 **Cell lines and culture conditions**

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

534 **METHOD DETAILS**

535 **gnomAD and ClinVar analysis**

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

548 **Cloning**

549 The NMD reporter plasmids have been previously described (31). For individual
550 fluorescent and luminescent reporters containing Gly, Val, Ala, and Tyr amino acids before the
551 PTC, oligos containing these sequences (Integrated DNA Technologies) and matching the
552 EcoRI (NEB, R3101S) and XhoI (NEB, R0146S) restriction sites were synthesized. These
553 oligos were annealed and ligated using the Quick Ligation Kit (NEB, M2200L) into EJC-
554 independent and EJC-enhanced fluorescent and luminescent backbones digested with EcoRI
555 and XhoI. To make the NMD– reporters for each sequence, each EJC-independent reporter
556 was digested with EcoRI and MfeI (NEB, R3589S), the sticky ends were filled in using Klenow
557 Fragment (NEB, M0212S) and the blunt ends were ligated together using the Quick Ligation Kit
558 (NEB), removing the GFP 3′ UTR. The oligos synthesized for each sequence tested are as
559 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

560

561 **MPRA plasmid library**

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

573 The EJC-independent and EJC-enhanced fluorescent reporter backbones were
574 linearized using EcoRI (NEB) and XhoI (NEB) at room temperature overnight. Digested plasmid
575 DNA was gel purified using the NucleoSpin Gel and PCR Clean-up kit (Takara Bio,
576 740609.250). The oligo library was then cloned into the digested backbones using the Gibson
577 Assembly Master Mix (NEB, E2611L) using an insert:vector molar ratio of 7:1 for the EJC-
578 independent backbone and 3:1 for the EJC-enhanced backbone. The reactions were incubated
579 at 50°C for 1 h. The reactions were then purified using AMPure XP beads to remove excess
580 salts and a total of 200 ng of plasmid DNA was transformed into MegaX DH10B T1R
581 Electrocompetent Cells (ThermoFisher, C640003), using a Biorad GenePulser electroporator.
582 The transformed cells were recovered in the recovery medium provided with the
583 electrocompetent cells at 37°C for 1 h, then plated on 33 pre-warmed 15 cm Luria broth (LB)
584 agar-Carbenicillin (RPI, C46000-5.0) plates and incubated at 37°C overnight. The next day
585 colonies were collected by scraping plates, spun-down, and then midi-prepped (NucleoBond
586 Xtra Midi EF kit, 740420.50) to extract the plasmid DNA.

587

588 **Plasmid transfections**

589 Transfections for fluorescent and luminescent reporters were performed differently since
590 fluorescent proteins have longer maturation rates than luciferase proteins (56). For the
591 fluorescent reporters, cells were split 24 h after transfection equally into two new 6-wells. Forty-
592 eight hours after transfection, one set of cells was harvested for RNA and the other set for flow
593 cytometry. For the luminescent reporters, transfections were performed in 24-well plates at a
594 40% confluency with 500 ng of plasmid DNA using Lipofectamine 2000 (ThermoFisher
595 Scientific, 11668027), as per manufacturer's instructions. All cells were harvested for luciferase
596 assays.

597 **Fluorescence imaging**

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

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

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

613 **Targeted sequencing of MPRA libraries**

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

621 digested RNA was used to make cDNA using Superscript II Reverse Transcriptase in duplicate
622 20 µL reactions (ThermoFisher Scientific, 18064014).

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

638 **MPRA data analysis**

639 To process the MPRA sequencing libraries, the UMI-tools repository was used to add the
640 UMIs to the name of each read in the FASTA files (57). Following this, Cutadapt was used to trim
641 the 5′ and 3′ends of the reads in a sequence dependent manner, up to the 10 nt context of interest
642 (58); ~25% of reads were lost in this process. SeqPrep was used to merge sequencing reads 1
643 and 2, only accepting reads that perfectly align (https://github.com/jstjohn/SeqPrep); ~5% of reads
644 were lost in this process. Bowtie was used to map the reads to a reference file of all sequence
645 contexts (59); 1% of reads were lost in this process. Samtools was used to create a BAM file of
646 mapped reads, followed by deduplicating the BAM files based on UMIs and converting the BAM
647 files to BED files (60). Around 10% of reads were lost upon deduplication of UMIs. A custom
648 python script was used to remove reads on the negative strand and to get a bed file of reads in
649 the correct orientation.

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

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

657 **Luciferase assays**

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

666 **Lentiviral transductions**

667 Knockdowns were conducted using shRNAs. Plasmids containing shRNAs were first
668 packaged into lentivirus. To do this, HEK293T cells were transfected with 4 µg of the relevant
669 shRNA vector (Functional Genomics Core, CU Anschutz), 3 µg of psPAX2 (Addgene plasmid
670 #12260), and 1 µg of pMD2.G (Addgene plasmid #12259), in 80% confluent 6 cm plates. The
671 shRNAs used were for scrambled (SCH016), UPF1 (TRCN0000022254), UPF2
672 (TRCN0000151381), UPF3B (TRCN0000152769), SMG1 (TRCN0000194827**)**, SMG6
673 (TRCN0000040014), ZNF598 (TRCN0000073159), EIF5A (TRCN0000062552), EIF3J
674 (TRCN0000062013), CUL2 (TRCN0000006523), and ABCE1 (TRCN0000158417). Twenty-four
675 hours after transfection the media was replaced with 3 mL of fresh media. Forty-eight hours
676 after transfection media containing lentivirus was harvested and spun down at 1200 rpm for 5
677 minutes. Aliquots of the supernatant were frozen at -80°C for later use.

678 For the knockdown experiments with the luciferase reporters, cells were seeded at a
679 40% confluency in 6-wells and infected with 50 µL of virus. Twenty-four hours post transduction,
680 the cells were split into 24-wells at a 20% confluency. Forty-eight hours post transduction, the
681 luciferase reporters were transfected into cells as described above. Seventy-two hours post-
682 transduction the cells were harvested for luciferase assays. For knockdown validation
683 experiments, the same procedure was followed without the transfection of luciferase reporters.

684

685 **RNA extraction and RT-qPCR**

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

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

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

702

703 **Protein isolation and western blotting**

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

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

721 **Luminescent reporters for *in vitro* assays**

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

736 **Expression and purification of eRF1 and eRF3**

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

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

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

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

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

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

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

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

789 A peptide release assay with the preTC-NLuc (Termi-Luc) was performed as previously
790 described with some modifications (63). Peptide release was performed in a solution containing
791 1.5 pM preTC-Nluc, 45 mM HEPES-KOH pH 7.5, 1.4 mM Mg_2OAc, 56 mM KOAc pH 7.0, 1 mM
792 DTT, 177 μM spermidine, 1.5 % (w/w) sucrose, 0.8 mM $MgCl_2$, 0.2 mM GTP supplemented with
793 equimolar $MgCl_2$, and 0.5 % NanoGlo (Promega), in the absence or the presence of release
794 factors eRF1 and eRF3a at various concentrations. Luminescence was measured at 30°C
795 using a Tecan Infinite 200 Pro (Tecan, Männedorf, Switzerland) for 12 min. The translation
796 efficiency was calculated as the maximal derivative of the growing linear section of the
797 luminescence curve (v_0, RLU/min). For comparison of different preTCs, a single working
798 concentration of 3 nM of eRF1-eRF3A was chosen in the linear section. The instability of the
799 preTC-NLuc was evaluated in Termi-Luc assay in the absence of release factors.

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

805 **Translation of NLuc in RRL lysates**

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

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

811 **QUANTIFICATION AND STATISTICAL ANALYSIS**

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

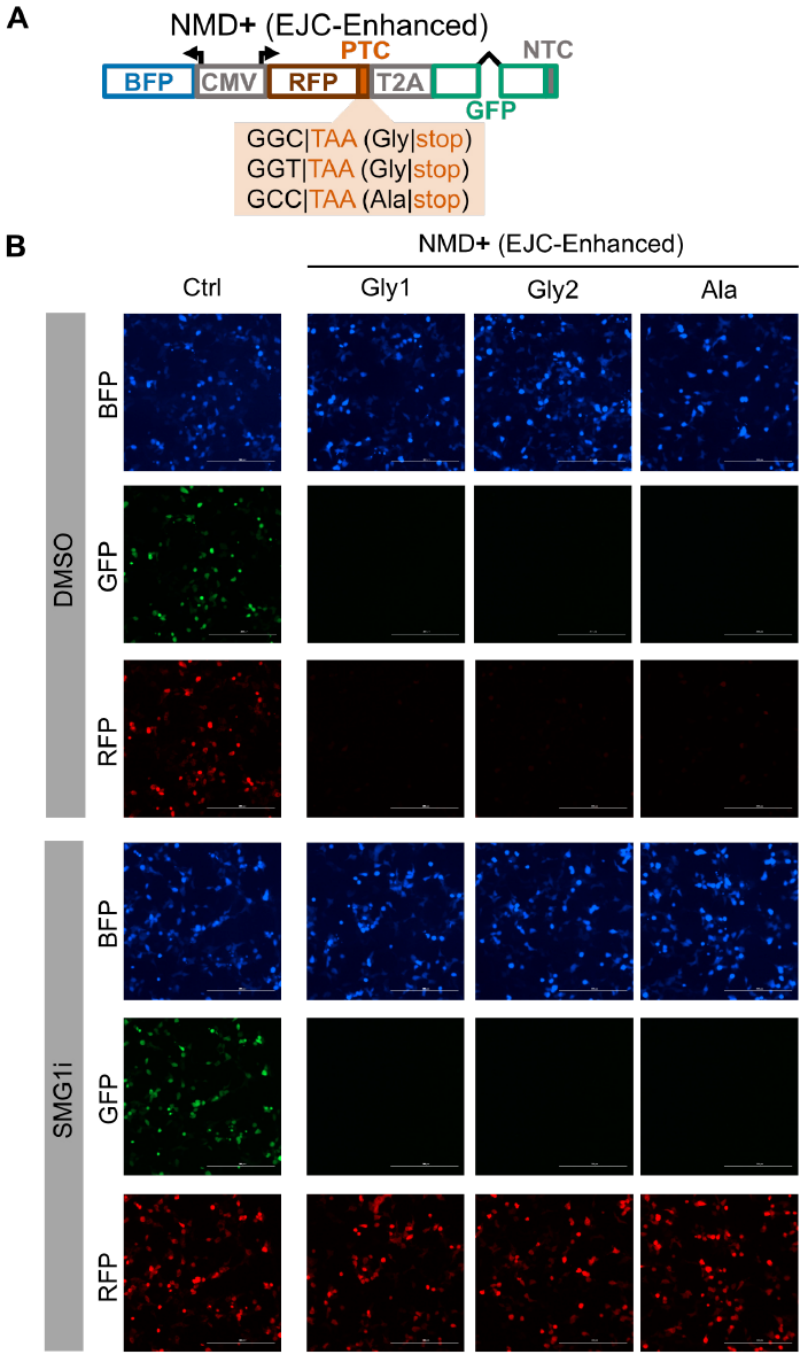
813 All experiments were performed with a sample size of at least n=3. For transient
814 transfections, each replicate represents an independent transfection. For stably integrated cells,
815 the same cell line was plated in 3 different wells, from which RNA and protein were harvested.
816 Statistical significance between various populations was calculated using a student's t-test in R
817 and p-values were two-sided (*$p<0.05$, **$p<0.005$, ***$p<0.0005$). Statistical details of specific
818 experiments can be found in the Results, Methods, and/or Figure Legends. Plots were
819 generated using R plotting functions and/or the ggplot2 package. The exact code and packages
820 used for these analyses can be found in this repository: https://github.com/jagannathan-
821 lab/2023-kolakada_et_al.
822

823 **SUPPLEMENTAL INFORMATION**

824

825 **SUPPLEMENTARY FIGURES**

826 **Figure S1**



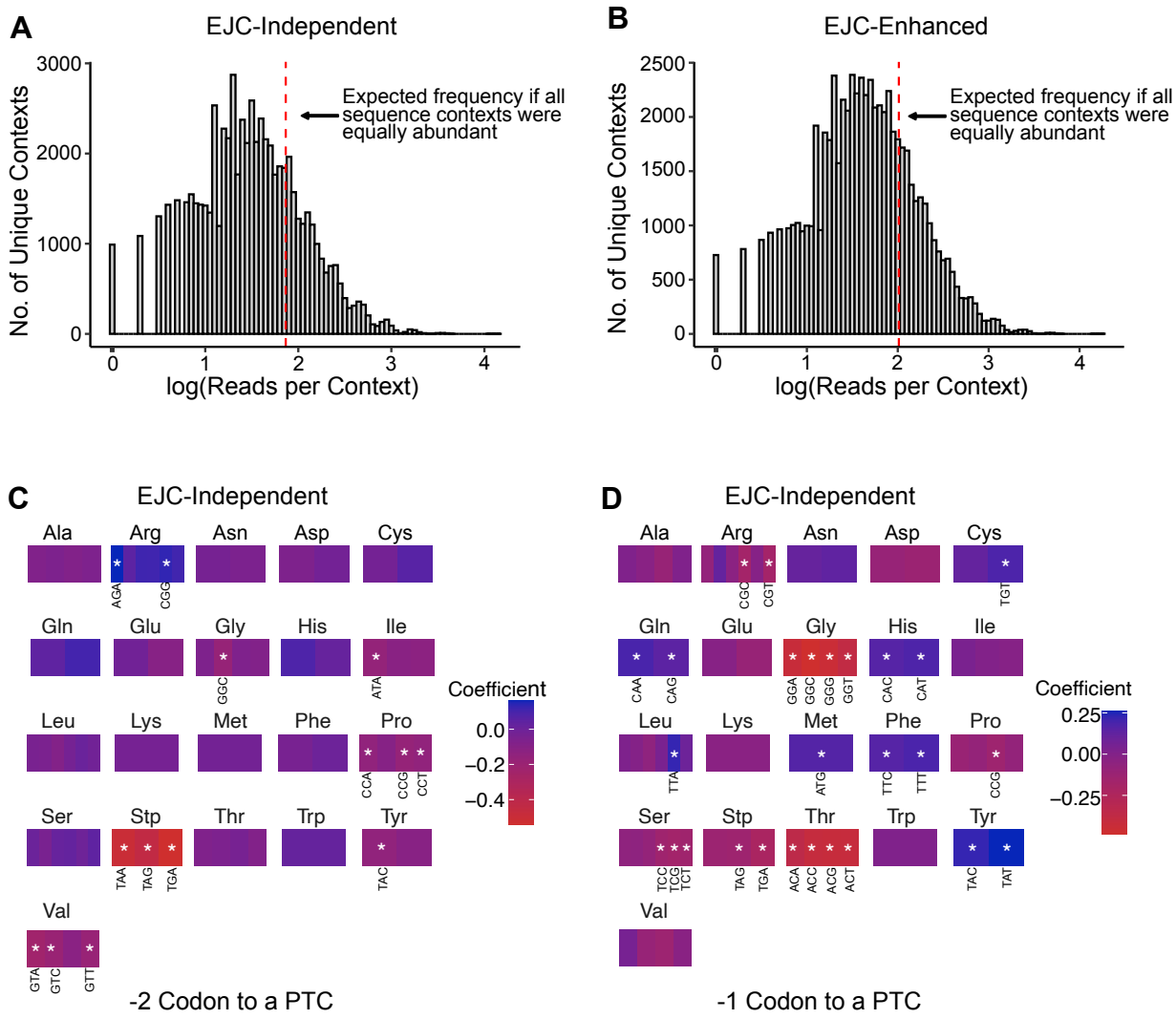
827

828 **Figure S1. EJC-enhanced reporters do not express RFP in any context.** (A) Schematic of

829 EJC-enhanced NMD reporter system used. (B) Fluorescence imaging of BFP, GFP, and RFP

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

834 **Figure S2.**



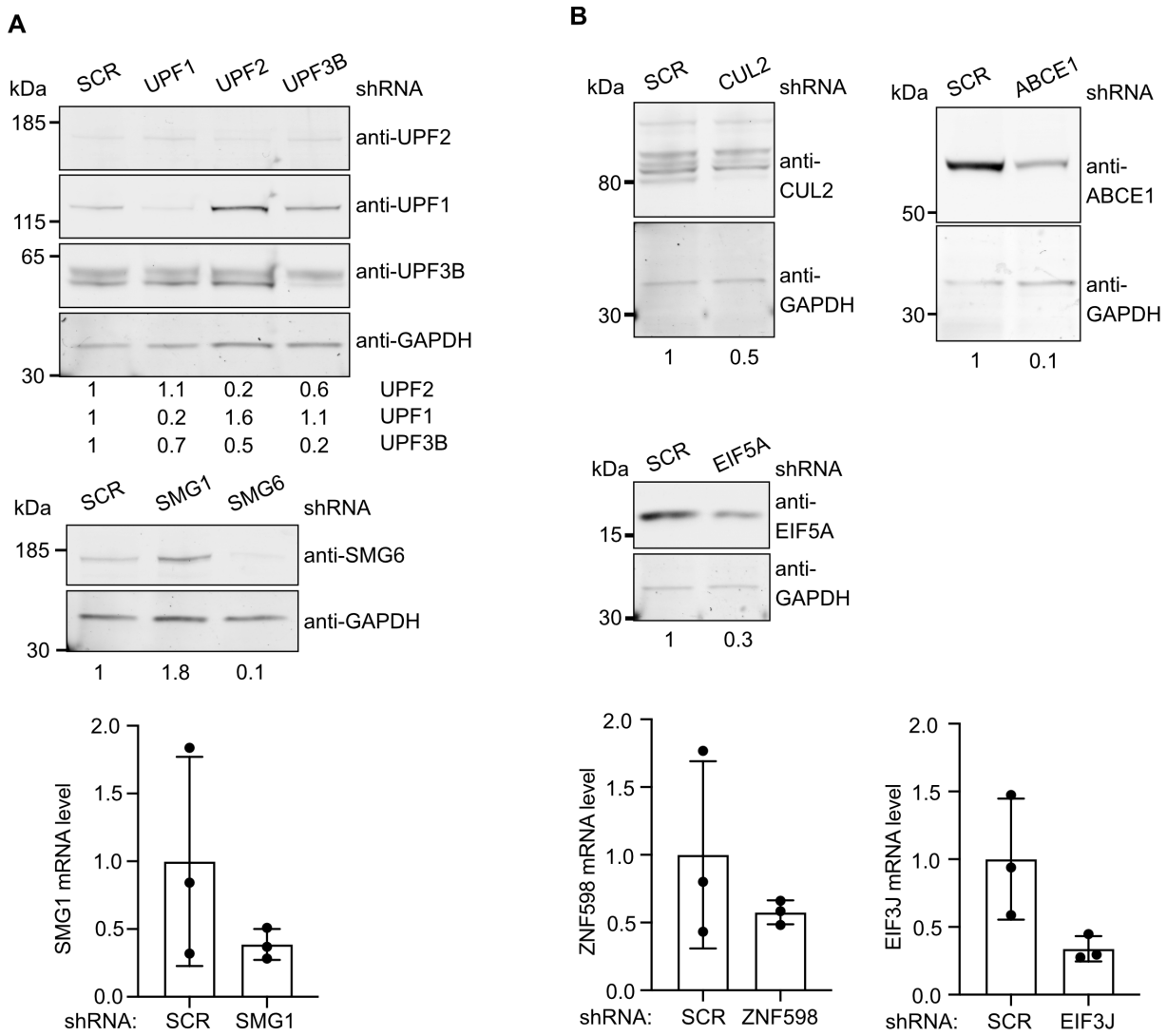
835

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

840

841

842 **Figure S3.**



843

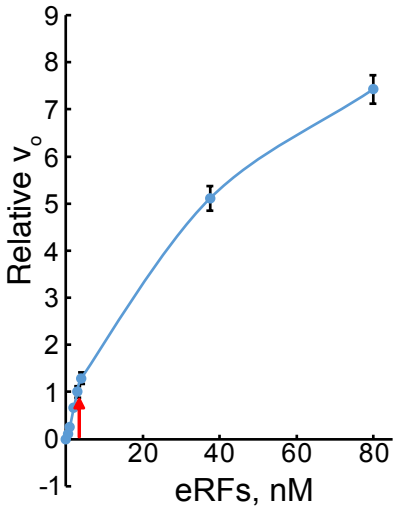
844 **Figure S3.** Knockdown validations for all shRNAs used. (**A**) Top: western blot validating UPF1,
845 UPF2, and UPF3B knockdowns; middle: western blot validating SMG6 knockdown; bottom:
846 qPCR validation of SMG1 knockdown. (**B**) Top left: western blot for CUL2 knockdown validation;
847 top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

851

852

853 **Figure S4.**

854



855

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

857 assay.

858

859

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

863 **SUPPLEMENTARY DATA**

864

865 **Supplementary Table 1.** DEseq2 results for the EJC-independent library of sequences
866 representing the fold change of DMSO over SMG1i treatment of cells.
867
868 **Supplementary Table 2.** DEseq2 results for the EJC-enhanced library of sequences
869 representing the fold change of DMSO over SMG1i treatment of cells.

870 **REFERENCES**

871

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

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

876 3. Nagy, E. and Maquat, L.E. (1998) A rule for termination-codon position within intron-
877 containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci*, **23**,
878 198-199.

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

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

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

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

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

892 9. Neu-Yilik, G., Amthor, B., Gehring, N.H., Bahri, S., Paidassi, H., Hentze, M.W. and
893 Kulozik, A.E. (2011) Mechanism of escape from nonsense-mediated mRNA decay of
894 human beta-globin transcripts with nonsense mutations in the first exon. *RNA*, **17**, 843-
895 854.

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

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

901 12. Jagannathan, S. and Bradley, R.K. (2016) Translational plasticity facilitates the
902 accumulation of nonsense genetic variants in the human population. *Genome Res*, **26**,
903 1639-1650.

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

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

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

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

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

918 18. Pierson, W.E., Hoffer, E.D., Keedy, H.E., Simms, C.L., Dunham, C.M. and Zaher, H.S.
919 (2016) Uniformity of Peptide Release Is Maintained by Methylation of Release Factors.
920 *Cell Rep*, **17**, 11-18.

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

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

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

930 22. Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S. and Jacobson, A.
931 (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated
932 mRNA decay. *Nature*, **432**, 112-118.

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

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

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

943 26. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K.,
944 Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B. *et al.* (2012) A systematic
945 survey of loss-of-function variants in human protein-coding genes. *Science*, **335**, 823-
946 828.

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

950 28. Koren, I., Timms, R.T., Kula, T., Xu, Q., Li, M.Z. and Elledge, S.J. (2018) The Eukaryotic
951 Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell*, **173**,
952 1622-1635 e1614.

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

956 30. Loughran, G., Li, X., O'Loughlin, S., Atkins, J.F. and Baranov, P.V. (2023) Monitoring
957 translation in all reading frames downstream of weak stop codons provides mechanistic
958 insights into the impact of nucleotide and cellular contexts. *Nucleic Acids Res*, **51**, 304-
959 314.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1026 53. Tate, W.P. and Mannering, S.A. (1996) Three, four or more: the translational stop signal
1027 at length. *Mol Microbiol*, **21**, 213-219.

1028 54. Imamachi, N., Salam, K.A., Suzuki, Y. and Akimitsu, N. (2017) A GC-rich sequence
1029 feature in the 3' UTR directs UPF1-dependent mRNA decay in mammalian cells.
1030 *Genome Res*, **27**, 407-418.

1031 55. Lejeune, F. and Maquat, L.E. (2005) Mechanistic links between nonsense-mediated
1032 mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol*, **17**, 309-
1033 315.

1034 56. Mazo-Vargas, A., Park, H., Aydin, M. and Buchler, N.E. (2014) Measuring fast gene
1035 dynamics in single cells with time-lapse luminescence microscopy. *Mol Biol Cell*, **25**,
1036 3699-3708.

1037 57. Smith, T., Heger, A. and Sudbery, I. (2017) UMI-tools: modeling sequencing errors in
1038 Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*, **27**, 491-
1039 499.

1040 58. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput
1041 sequencing reads. *2011*, **17**, 3.

1042 59. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat*
1043 *Methods*, **9**, 357-359.

1044 60. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
1045 Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence
1046 Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

1047 61. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using
1048 real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402-408.

1049 62. Frolova, L., Seit-Nebi, A. and Kisselev, L. (2002) Highly conserved NIKS tetrapeptide is
1050 functionally essential in eukaryotic translation termination factor eRF1. *RNA*, **8**, 129-136.

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

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

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

1059