Translation of upstream open reading frames does not trigger nonsense-mediated decay in Arabidopsis (99/120 characters)

Author Affiliations: Hsin-Yen Larry Wu[1] and Polly Yingshan Hsu[1, #]

[1] Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824 USA

[#] Corresponding author: pollyhsu@msu.edu, phone: 517-353-5284

ORCIDs:

0000-0001-9407-338X (H.L.W.)

0000-0001-7071-5798 (P.Y.H.)

Keywords: translation, upstream open reading frames, nonsense-mediated decay, mRNA half-life, ribosome profiling, Arabidopsis

Figures 1-6

Supplementary Figures 1-7

Supplementary Tables 1-4

ABSTRACT (199/200 Words)

Upstream ORFs (uORFs) are widespread cis-regulatory elements in the 5' untranslated regions of eukaryotic genes. In addition to repressing main ORF translation, uORF translation in animals also reduces mRNA stability through nonsense-mediated decay (NMD). In contrast, the roles of uORFs in plants are less understood. One limitation of most prior studies is the inclusion of both translated and untranslated uORFs for analysis. Here, we identified actively translated uORFs (TuORFs) in Arabidopsis through ribosome profiling and examined their roles in gene expression by integrating multiple genome-wide datasets. Similar to animal systems, TuORFs are associated with 38%, 14%, and 43% reductions in translation efficiency, mRNA stability, and protein levels, respectively. Surprisingly, TuORF-containing mRNAs in plants have 51% higher transcript levels, which is opposite to the observations in animals. We present multiple lines of evidence showing that uORF translation generally does not trigger NMD in Arabidopsis. First, TuORF-containing mRNAs are not upregulated in various NMD mutants. Second, TuORF-containing mRNAs and known NMD targets have distinct expression patterns, and they are translationally repressed via different mechanisms. Finally, TuORF-containing mRNAs and NMD targets are degraded through separate pathways. Together, our results highlight fundamental differences in gene regulation mediated by TuORFs in plants and animals.

INTRODUCTION

Upstream open reading frames (uORFs) are cis-acting translational repressors that regulate many eukaryotic protein-coding transcripts. Mutations or modifications of uORFs are associated with human diseases and plant traits (Calvo et al., 2009; Xu et al., 2017; Zhang et al., 2018; Lee et al., 2021). Considering AUG as a translation start codon, 30%–70% of genes contain at least one potential uORF in moss, Arabidopsis, fly, mouse, and human (Kim et al., 2007; Johnstone et al., 2016; Zhang et al., 2021). The genes with predicted uORFs are enriched for signal transduction proteins, transcription factors, and membrane proteins in animals (Zhang et al., 2021) and transcription factors and protein kinases in plants (Kim et al., 2007), underlining the potential importance of uORFs in modulating the expression of vital cellular regulators.

Many studies in eukaryotes have reported that uORF translation reduces main ORF (mORF) translation (Chew et al., 2016; Johnstone et al., 2016; Liu et al., 2013; Wu et al., 2019; Lin et al., 2019), and it is estimated to reduce protein production by 30%–80% (Calvo et al., 2009; Chew et al., 2016). According to the scanning model of mRNA translation, the 43S pre-initiation complex, which contains the small ribosomal subunit, several translation initiation factors, and initiator tRNA, binds the 5’ cap and scans the 5’ untranslated region (UTR) to search for an optimal initiation sequence (Hinnebusch et al., 2016). When a uORF is translated, the ribosome may stall during elongation or at termination, or it may dissociate from the mRNA after termination, thus reducing the translation of the downstream mORF (Dever et al., 2020). The mORFs downstream of uORFs may get translated through leaky scanning or reinitiation (Kozak, 1999).

uORF translation is also expected to trigger mRNA degradation, presumably through nonsense-mediated decay (NMD). NMD is an evolutionarily conserved regulatory mechanism to eliminate aberrant mRNAs, such as those that contain premature termination codons and unusually long 3’ UTRs (He and Jacobson, 2015; Kurosaki et al., 2019). Two pathways have

70 been established in mammals: the Exon Junction Complex (EJC)-dependent and long 3' UTR
71 pathways. Normally, EJC and two NMD factors (UP-FRAMESHIFT SUPPRESSOR 2 and 3;
72 UPF2 and UPF3) deposited during splicing of mRNAs will be evicted by the ribosome in the first
73 round of translation. In the presence of a premature termination codon, EJC-UPF2-UPF3
74 complexes downstream of the termination codon remain attached to the mRNA. This allows
75 UPF1, which is part of the Surveillance Complex formed at translation termination, to interact
76 with UPF2 and UPF3 in proximity. Interaction of UPF1, UPF2, and UPF3 activates the
77 SUPPRESSOR WITH MORPHOLOGICAL EFFECT ON GENITALIA1 (SMG1) kinase, which
78 phosphorylates UPF1. This establishes the first phase of NMD, substrate recognition, and
79 activates downstream signaling and eventually the second phase of NMD, substrate
80 degradation, which involves SMG5, SMG6, and SMG7. Phosphorylated UPF1 also inhibits
81 translation initiation (Isken et al., 2008). Alternatively, unusually long 3' UTRs, which result in
82 poor translation termination, are believed to promote NMD through phosphorylated UPF1
83 independent of EJC (Kurosaki et al., 2014).
84 The plant NMD pathway is best characterized in *Arabidopsis thaliana*. Arabidopsis
85 possesses homologous NMD factors involved in the first phase, including UPF1, UPF2, and
86 UPF3, and the second phase, SMG7 (Raxwal and Riha, 2016). However, Arabidopsis is
87 missing the homologs of animal SMG1, SMG5, and SMG6 (Lloyd and Davies, 2013). The
88 absence of key NMD components in plants suggests significant variation in the NMD
89 mechanisms between plants and animals. It remains an open question how the differences in
90 NMD components affect substrate recognition and degradation in plants.
91 In Arabidopsis, putative NMD targets have been identified based on transcripts that
92 overly accumulate in NMD-defective mutants (Kurihara et al., 2009; Kalyna et al., 2012;
93 Drechsel et al., 2013; Degtiar et al., 2015). However, there is a concern that upregulation of
94 some transcripts in NMD mutants might be caused by secondary effects of NMD (Raxwal et al.,
95 2020). For example, NMD mutants affect numerous aspects of plant physiology, and some

96 alleles cause severe growth retardation and lethality due to constitutive immune responses
97 (Jeong et al., 2011; Rayson et al., 2012; Riehs-Kearnan et al., 2012; Filichkin et al., 2015;
98 Merchante et al., 2015; Chiam et al., 2019). Moreover, UPF1 has been reported to have multiple
99 functions in plants, such as alternative splicing regulation and translational control, in addition to
100 NMD (Yoine et al., 2006; Raxwal et al., 2020). To identify *bona fide* NMD targets, *upf1* and
101 *smg7*, which are defective in the first and second phases of NMD, respectively, were introduced
102 into *pad4*, in which the immune response is inactivated (Riehs-Kearnan et al., 2012; Raxwal et
103 al., 2020). The 333 transcripts upregulated in both *upf1 pad4* and *smg7 pad4* were defined as
104 high-confidence NMD targets. It was shown these NMD targets are translationally regulated by
105 UPF1 and that their turnover requires decapping and 5' to 3' decay (Raxwal et al., 2020).
106 As uORF translation could result in a transcript with an unusually long 3' UTR, uORF-
107 containing mRNAs are predicted to be targeted by NMD (Kertész et al., 2006; Kurosaki et al.,
108 2019). Consistent with this expectation, genome-wide studies in zebrafish, mouse, and human
109 have found that transcripts with translated uORFs have lower mRNA levels globally (Chew et
110 al., 2016; Johnstone et al., 2016). Moreover, transcripts containing translated uORFs are
111 upregulated in knockdown of NMD factors in mouse embryonic stem cells and human cell lines
112 (Hurt et al., 2013; Baird et al., 2018), supporting the idea that these transcripts are NMD targets.
113 In plants, whether uORF translation triggers NMD is still under debate. Two genome-
114 wide studies suggest that uORF-containing mRNAs are targets of NMD, as transcripts
115 possessing predicted uORFs are enriched in the genes upregulated in Arabidopsis *upf1* and
116 *upf3* mutants (Kurihara et al., 2009; Drechsel et al., 2013). In contrast, three studies suggested
117 that NMD only controls a small fraction of uORF-containing mRNAs in plants, such as those
118 containing uORFs with > 35 codons (~2% of uORF-containing genes), or those responsive to
119 certain environmental stimuli, such as polyamine or blue light (Nyikó et al., 2009; Uchiyama-
120 Kadokura et al., 2014; Kurihara et al., 2018). Moreover, evidence from several gene-specific or
121 genome-wide studies suggests that uORFs do not trigger NMD. First, the mRNA levels of

122 several uORF-containing transcripts, including *NIP5;1* and *AtMHX*, are not increased in NMD
123 mutants (Saul et al., 2009; Tanaka et al., 2016). Second, a recent genome-wide study reported
124 that genes with increasing numbers of uORFs did not show lower transcript abundances in
125 Arabidopsis and tomato (Li and Liu, 2020). It is unknown whether some uORF-containing
126 mRNAs escape NMD or whether uORF-containing mRNAs are generally not targets of NMD in
127 plants. It has been concluded that the mRNA features studied so far is insufficient to infer NMD
128 targetability in plants (Ohtani and Wachter, 2019). Overall, it remains unclear how translated
129 uORFs affect mRNA stability and what the role of NMD in this process is in plants.
130 One limitation of the previous studies is that predicted uORFs, which include both
131 translated and untranslated uORFs, were commonly included in the analyses. Although many
132 uORFs have been predicted based on nucleotide sequences (Zhang et al., 2021), determining
133 which uORFs are actively translated still requires experimental evidence. Recently, deep
134 sequencing of ribosome footprints – known as ribosome profiling or Ribo-seq – was exploited to
135 study genome-wide mRNA translation (Ingolia et al., 2009; Brar and Weissman, 2015). Several
136 Ribo-seq studies in Arabidopsis have deployed different strategies to identify translated uORFs
137 (TuORFs) (Liu et al., 2013; Juntawong et al., 2014; Hsu et al., 2016; Hu et al., 2016; Bazin et
138 al., 2017; Kurihara et al., 2020). Among different parameters used, 3-nucleotide (3-nt)
139 periodicity, which corresponds to translating ribosomes deciphering 3 nts per codon, is
140 considered to be the most stringent criterion to identify high-confidence translation events
141 (Ingolia et al., 2009; Calviello et al., 2016; Xu et al., 2018; Harnett et al., 2021). Our previous
142 work and a recent study using RiboTaper (Calviello et al., 2016), which computes the statistical
143 significance of 3-nt periodicity, uncovered 187–1378 AUG-initiated TuORFs in Arabidopsis (Hsu
144 et al., 2016; Kurihara et al., 2020).
145 Here we further improved the Ribo-seq methodology and identified 2093 AUG-initiated
146 TuORFs in Arabidopsis seedlings. We systematically examined the assumed roles of TuORFs
147 in mORF translation and mRNA decay by integrating multiple genome-wide datasets. As

148 predicted, we found that TuORF mRNAs are associated with lower translation efficiency, shorter
149 mRNA half-lives, and lower protein abundance. Surprisingly, we found that, in contrast to the
150 animal paradigm, plant TuORF mRNAs have higher steady-state mRNA levels, and they are not
151 regulated by the NMD pathway. Examination of numerous parameters revealed that uORFs and
152 NMD regulate gene expression through separate mechanisms. Taken together, our results
153 reveal unexpected features and regulation of uORFs in Arabidopsis and suggest that plants and
154 animals have evolved different uORF-dependent mechanisms for gene regulation.
155

156 **RESULTS**

157 **Translated uORFs identified via an improved Ribo-seq method**

158 Previously we optimized Ribo-seq to reach high 3-nt periodicity, and we exploited this
159 unique feature of mRNA translation to identify translated ORFs in Arabidopsis (Hsu et al., 2016).
160 Although the data revealed many unannotated translation events, the number of TuORFs
161 uncovered was low, presumably due to insufficient Ribo-seq coverage within uORFs. We
162 proposed that increasing Ribo-seq coverage could improve the efficiency of translated ORF
163 identification, especially for relatively short ORFs, such as uORFs. Here we optimized the
164 coverage of Ribo-seq in 7-day-old Arabidopsis seedlings by minimizing the potential loss of
165 ribosome footprints during the purification process (see MATERIALS AND METHODS for
166 details).

167 In addition to high correlations among biological replicates (**Figure S1A**), our new Ribo-
168 seq data displayed characteristics expected for high-quality data, including strong 3-nt
169 periodicity (92% in-frame reads), high enrichment for coding sequences (CDSs), and
170 characteristic ribosome footprint lengths (**Figure 1A-C**). The improved coverage is evident when
171 examining the number of unique ribosome footprints detected globally. At a sequencing depth of
172 20 million reads from randomly selected reads, the number of unique ribosome footprints in our
173 new data was 4.01-fold higher than that in our previous dataset (**Figure 1D**). The coverage

174 improvement is also evident in individual transcript profiles within both uORFs and mORFs. For
175 example, a TuORF in *REPRESSOR OF GA1* (*RGA1*) was identified in our current data but
176 missed in our previous study despite the mRNA levels being similar (**Figure S1B**).
177 The high-coverage data increased the number of TuORFs uncovered. In total, 2093
178 TuORFs were identified in 1768 genes based on strong 3-nt periodicity (**Figure 1E-G** and **Table**
179 **S1A**). This result is in strong contrast to the relatively few TuORFs identified in Arabidopsis
180 seedling shoots (93 TuORFs) and roots (138 TuORFs) from our previous data obtained using
181 plants grown under similar conditions and using the same statistical analysis and annotations
182 (**Figure 1E** and **Table S1B-C**). The new data also revealed translation from a higher number of
183 Conserved Peptide uORFs (CPuORFs) and annotated coding sequences (mORFs of protein-
184 coding genes) (**Figure 1E** and **Table S1D-F**). Taken together, these results showed that our
185 enhanced-coverage Ribo-seq data dramatically improved the identification of TuORFs.
186 Previous analyses of predicted uORFs found that transcription factors and protein
187 kinases are overrepresented among uORF-containing genes (Von Arnim et al., 2014; Kim et al.,
188 2007). Consistent with these analyses, functions related to transcription factors and protein
189 kinases/phosphatases were identified as being enriched in our Gene Ontology (GO) analysis of
190 TuORF-containing genes (**Table S2**). We found that many TuORF-containing genes encode
191 extensively studied regulators of plant growth and environmental responses. For example, this
192 list included well-known transcription factors, such as those involved in hormone signaling
193 (RGA1 and AUXIN RESPONSE FACTOR2/3/8), root development (SCARECROW), the
194 circadian clock (TIMING OF CAB EXPRESSION1), and light signaling (PHYTOCHROME
195 INTERACTING FACTOR3/4/7) (Silverstone et al., 1998; Strayer et al., 2000; Di Laurenzio et al.,
196 1996; Ellis et al., 2005; Nagpal et al., 2005; Ni et al., 1998; Huq and Quail, 2002; Leivar et al.,
197 2008). The list also included well-characterized protein kinases, such as Phytochromes B and
198 E, ribosomal-protein S6 kinases (S6K1/2), and SNF1-related protein kinase 2 (SNRK2.1/2.5)
199 (**Figure S1B, S2A-N**) (Sharrock and Quail, 1989; Devlin et al., 1998; Zhang et al., 1994;

200 Mizoguchi et al., 1995; Park et al., 1993; Boudsocq et al., 2004). These results support the
201 notion that TuORFs control important cellular regulators in plants.
202
203 **TuORF-containing genes have higher mRNA abundance but are associated with lower**
204 **TE, shorter mRNA half-lives, and lower protein abundance**
205 To understand how uORF translation affects gene expression, we compared mRNA
206 abundance and translation efficiency of TuORF-containing genes (hereafter, TuORF genes),
207 genes containing sequence-predicted uORFs but undetected by RiboTaper (hereafter, UuORF
208 genes), and genes containing an annotated 5' UTR but no potential uORFs (hereafter, no-uORF
209 genes). Note that some UuORFs may be translated but missed in our analysis because their 3-
210 nt periodicity is insufficient to pass the statistical test: these include extremely short uORFs,
211 those translated at a low frequency, and those overlapping with other uORFs, thus disrupting
212 the 3-nt periodicity. To study the global effects of uORFs on mRNA stability and protein
213 abundance, we compared genes with TuORFs, UuORFs, or no-uORFs, using available mRNA
214 half-life data determined by metabolic labeling and our previous quantitative proteomics data,
215 both also from Arabidopsis seedlings (Szabo et al., 2020; Song et al., 2018) (**Figure 2A-D**). The
216 growth conditions and source of the data are summarized in **Table S3**.
217 Studies in vertebrates have shown that uORF translation is associated with lower
218 steady-state mRNA abundance, consistent with the expectation that TuORF mRNAs are
219 regulated by NMD (Chew et al., 2016; Johnstone et al., 2016). Surprisingly, we found that
220 TuORF genes in Arabidopsis have higher steady-state mRNA levels than that of no-uORFs
221 genes (**Figure 2A**), despite having shorter mRNA half-lives as expected (**Figure 2C**). The
222 median mRNA level of TuORF genes was 51% higher than that of no-uORF genes. The higher
223 steady-state mRNA levels and shorter mRNA half-lives suggest that TuORF genes are highly
224 transcribed. To examine whether the high transcript levels of TuORF genes are conserved in
225 other plant species, we reanalyzed our previous tomato Ribo-seq data from seedling roots (Wu

226 et al., 2019) and found the TuORF genes in tomato also have higher steady-state mRNA levels
227 (**Figure 2E**). Given that Arabidopsis and tomato diverged more than 100 million years ago
228 (Clarke et al., 2011), our results suggest that TuORF-associated higher mRNA abundance
229 could be prevalent in plants. To investigate whether our observation is specific to seedlings or
230 common to different developmental stages/tissues/genetic backgrounds in Arabidopsis and
231 tomato, we examined the mRNA levels of the uORF-containing genes in 128 independent RNA-
232 seq studies collected by the EMBO-EBI expression atlas (Papatheodorou et al., 2020).
233 Consistent with our Arabidopsis seedling and Tomato seedling root data, we found the TuORF
234 genes are expressed at higher levels in diverse tissues, developmental stages, and ecotypes in
235 Arabidopsis and tomato (**Figure 2G and S3**), supporting the hypothesis that TuORF genes are
236 preferentially expressed at high levels in these plants under a wide range of sampling
237 conditions.

238 One trivial explanation for the higher mRNA abundance of TuORF genes is that TuORFs
239 are more likely to be discovered in highly abundant transcripts. To test this possibility, we first
240 examined genes containing CPuORFs that were identified independently from our approach by
241 evolutionary conservation (Hayden and Jorgensen, 2007; Jorgensen and Dorantes-Acosta,
242 2012; Takahashi et al., 2012; Vaughn et al., 2012; Van Der Horst et al., 2019). We found that
243 the median mRNA abundance of CPuORF genes was 2.1-fold higher than that of no-uORF
244 genes (**Figure S4**), consistent with our observations that genes with TuORFs are expressed at
245 higher levels. Next, we examined the distributions of TuORF and UuORF mRNA levels (**Figure**
246 **2H**). We found many UuORF genes were highly expressed as well. In fact, the distributions of
247 TuORF and UuORF mRNA levels in seedlings completely overlapped (**Figure 2H**), indicating
248 that mRNA abundance is not a major determinant of whether a uORF is detected as translated.
249 Finally, the fact that TuORFs are associated with lower TE, shorter mRNA half-lives, and less
250 protein accumulation than UuORF genes (**Figure 2B-D, F**), indicates that TuORFs are potent
251 inhibitors for gene expression. Together, our results support the hypothesis that TuORFs are

252 selectively translated to repress the translation and mRNA stability. Our observation of the high
253 mRNA abundance of plant TuORF genes also reveals that plants and animals implement
254 TuORFs differently in gene regulation.

255 Interestingly, even though globally TuORF genes in Arabidopsis start with 51% more
256 mRNAs than no-uORF genes, they produce 41%–43% less protein, on average (**Figure 2A, 2D,**
257 **and S5**), likely due to the combined effects of a 38% reduction in translation efficiency and 14%
258 reduction in mRNA half-life (**Figure 2B, 2C**). These observations suggest that global repression
259 of TuORF gene protein synthesis is determined by multilayer post-transcriptional regulation.
260 Notably, for the four parameters examined here, UuORF genes consistently have the levels
261 between those of TuORF genes and no-uORF genes: 4% increase in mRNA levels and 11%,
262 6%, and 21% reductions in TE, mRNA half-life, and protein abundance, respectively (**Figure**
263 **2A-D**). These results indicate that UuORFs have smaller effects on gene expression than
264 TuORFs and highlight the importance of distinguishing TuORFs and UuORFs in downstream
265 analysis.

266

267 **TuORF genes are larger and likely evolutionarily older than other genes**

268 We next investigated whether there are general structural differences among TuORF,
269 UuORF, and no-uORF mRNAs. Considering the directionality of ribosome scanning, we
270 hypothesized TuORFs might be associated with longer 5' UTRs. Indeed, we found that the
271 median 5' UTR of TuORF mRNAs is 2.5-fold of that of no-uORF mRNAs (**Figure 3A**), which is
272 consistent with a previous report that uORFs in longer 5' UTRs are more likely to be translated
273 (Liu et al., 2013). However, unexpectedly, we found that TuORF mRNAs also have longer CDSs
274 and 3'UTRs, as well as longer transcript lengths, than UuORF mRNAs (**Figure 3B-D**). No-uORF
275 mRNAs were the shortest in all categories examined (**Figure 3A-D**). These results support the
276 hypothesis that TuORFs are more likely to be found in longer 5' UTRs and that TuORF mRNAs
277 tend to encode larger proteins (**Figure 3B**). The enrichment of TuORF genes for transcription

278 factors, protein kinases, and related functions in GO term analysis (**Table S2**) also implies
279 TuORF genes tend to encode proteins with conserved domains. It is known that evolutionarily
280 older genes are generally longer, expressed at higher levels, and have protein sequences that
281 are under stronger purifying selection (Wolfa et al., 2009; Cui et al., 2015). Our observations
282 that TuORFs have higher expression levels, longer transcript lengths, and are enriched for
283 transcription factors and protein kinases imply that older genes are more likely to be regulated
284 through mRNA translation and degradation mediated by TuORFs.
285
286 **TuORF mRNAs are not upregulated in the NMD mutants**
287 The established paradigm from vertebrates is that uORF translation triggers NMD (Chew
288 et al., 2016; Johnstone et al., 2016; Hurt et al., 2013). Despite having higher steady state mRNA
289 levels (**Figure 2A**), TuORF genes in Arabidopsis also have shorter mRNA half-lives (**Figure**
290 **2C**). Therefore, we investigated whether TuORF-containing transcripts in Arabidopsis are
291 subject to NMD.
292 If TuORFs trigger NMD, one would expect TuORF mRNAs to be among the known NMD
293 targets and their mRNA levels to be increased in NMD-deficient mutants. We first examined
294 whether TuORF mRNAs were reported in a high-confidence NMD target list (Raxwal et al.,
295 2020). Among the 333 NMD targets that were upregulated in both *upf1 pad4* and *smg7 pad4*
296 adult leaves (Raxwal et al., 2020), 248 genes with an annotated 5' UTR were expressed in our
297 data (**Table S4A**). However, only 49 of 1748 TuORF genes were identified in the high-
298 confidence NMD target list (**Table S4B**). The 49 genes represent 2.8% of TuORF-containing
299 genes and 20% of the NMD targets. These results suggest that TuORFs generally do not trigger
300 NMD; rather, a notable fraction (20%) of NMD targets are regulated by both TuORFs and NMD.
301 One caveat is that the poor overlap between the two gene lists could be due to differences in
302 growth conditions, developmental stages, and sample handling in different labs (**Table S3**).

303 To further investigate if TuORF mRNAs are NMD targets, we examined whether they are
304 upregulated in NMD-deficient mutants by analyzing four published RNA-seq datasets for NMD
305 mutants (Drechsel et al., 2013; Merchante et al., 2015; Raxwal et al., 2020; Gloggnitzer et al.,
306 2014). Although these datasets were collected from different NMD mutants and at various
307 developmental stages (**Table S3**), the high-confidence NMD targets were consistently and
308 significantly upregulated in all the NMD mutants examined compared with the control plants
309 (i.e., 3.05-, 1.54-, 4.98-, and 2.65-fold upregulation for NMD targets without TuORFs; 3.04-,
310 2.23-, 5.04-, 2.66-fold upregulation for NMD targets with TuORFs) (**Figure 4A-D**). In contrast,
311 the mRNA levels of TuORF mRNAs showed no or little change in the NMD mutants (i.e., 1.01-,
312 0.98-, 1.10- and 1.10-fold difference). Thus, the NMD targets, but not TuORF mRNAs, are
313 dramatically increased in the NMD mutants, further supporting the idea that uORF translation in
314 Arabidopsis generally does not trigger NMD.
315
316 **Lengths of the TuORF and 3' UTR alone are insufficient to predict NMD targetability**
317 A previous study suggested that uORFs that encode a peptide longer than 35 amino
318 acids are likely to trigger NMD in Arabidopsis (Nyikó et al., 2009). To test whether TuORF
319 length is associated with NMD targetability, we analyzed TuORF lengths in the NMD targets and
320 non-NMD targets. We found that the NMD target TuORF genes tend to have longer uORF
321 peptides than non-NMD target TuORF genes (median length 28 amino acids vs. 20.5 amino
322 acids, $p = 9.874e-07$, Wilcoxon signed-rank test). However, the distributions of the uORF
323 peptide lengths in NMD targets and non-NMD targets largely overlapped (**Figure S6**), indicating
324 that TuORF length alone is insufficient to predict whether a TuORF triggers NMD of a transcript.
325 We also examined whether the high-confidence NMD targets have longer 3' UTRs, a
326 feature that triggers NMD. Interestingly, the NMD targets do not have significantly longer 3'
327 UTRs, no matter whether they contain TuORFs or not (**Figure S7**). Taken together, our results

328 are consistent with a recent suggestion that no definite RNA features studied to date are
329 sufficient to predict NMD targets in plants (Ohtani and Wachter, 2019).
330
331 **TuORF mRNAs and NMD targets have distinct expression patterns**
332 To further characterize the roles of NMD and TuORFs in gene expression regulation, we
333 compared the expression patterns of the NMD targets and TuORF mRNAs (**Figure 5**).
334 Consistent with the expectation that NMD targets are being actively degraded, we found that
335 NMD targets have lower mRNA abundances than other genes; this is in strong contrast to
336 TuORF mRNAs, which have higher mRNA abundances than other genes (**Figure 5A**). While
337 both NMD targets and TuORF genes are associated with lower TE, shorter mRNA half-lives,
338 and lower protein abundance than other genes (**Figure 5B-D**), the NMD targets are associated
339 with even lower protein levels than TuORF genes (**Figure 5D**).
340 Interestingly, while both NMD targets and TuORF mRNAs have lower translation
341 efficiency (**Figure 5B**), TuORF-containing NMD targets have significantly lower translation
342 efficiency than either NMD targets without TuORFs or TuORF-containing non-NMD target
343 genes (**Figure 5B**). This additive repression of translation efficiency for TuORF-containing NMD
344 targets suggests that NMD and TuORFs suppress translation through separate mechanisms.
345 Our observation is consistent with the expectation that TuORFs prevent the scanning pre-
346 initiation complex from reaching the mORF while NMD suppresses translation through UPF1
347 and/or other translational repressors (Isken et al., 2008; Raxwal et al., 2020; Zinshteyn et al.,
348 2021). Together, the differences in the expression patterns between TuORF mRNAs and NMD
349 targets further support the idea that TuORFs are not an mRNA feature that triggers NMD in
350 Arabidopsis.
351
352 **TuORF mRNAs and NMD targets are degraded through different mRNA decay**
353 **mechanisms**

354 We next investigated the mechanisms underlying the decay of TuORF mRNAs and NMD
355 targets. mRNA decay can occur in the 5' to 3' or 3' to 5' direction, which is carried out by
356 different machinery (Sieburth and Vincent, 2018). A genome-wide survey of mRNA decay rates
357 in wild type and mutants defective in either 5' to 3' decay or 3' to 5' decay, or both, has been
358 conducted in Arabidopsis (Sorenson et al., 2018). In the survey, the *vcs-7* mutant, a null mutant
359 of *VARICOSE*, which encodes a scaffold protein for the mRNA decapping complex, was used to
360 examine the 5' to 3' mRNA decay; the *sov* mutant, defective in *SUPPRESSOR OF VARICOSE*,
361 which encodes a ribonuclease II, was used to examine 3' to 5' decay; and the *vcs/sov* double
362 mutant was used to examine decay affected in both directions (Deyholos et al., 2003; Goeres et
363 al., 2007; Xu et al., 2006; Zhang et al., 2010). We exploited this published dataset (Sorenson et
364 al., 2018) to study the genetic requirements for the mRNA decay of TuORF mRNAs, NMD
365 targets, TuORF-containing NMD targets, and other genes (**Figure 6**).
366 Similar to mRNAs that are non-NMD targets and lack TuORFs (**Figure 6D, 6H**), TuORF
367 mRNAs had dramatically reduced decay rates in the *vcs* mutant and the *vcs/sov* double mutant
368 (**Figure 6A, 6E**), indicating that decapping and 5' to 3' decay are critical for degrading these
369 transcripts. In contrast, the mRNA decay of NMD targets without TuORFs was only significantly
370 lower in the *vcs/sov* double mutant, suggesting that NMD targets are efficiently degraded by
371 both the 5' to 3' and 3' to 5' pathways (**Figure 6B, 6F**). Although there are fewer NMD targets
372 containing TuORFs, which resulted in a small sample size and statistical insignificance (**Figure**
373 **6C, 6G**), the decay rates in this group of genes in the four genetic backgrounds showed trends
374 similar to those of NMD targets without TuORFs (**Figure 6B, 6F**).
375 Together, our results suggest that TuORF mRNAs and NMD targets have distinct
376 genetic requirements for mRNA decay (see model in **Figure 6I**). Although other mRNAs are
377 also decayed through the 5' to 3' pathway (**Figure 6D, 6H**), the TuORF mRNAs appear to
378 undergo accelerated decay as their mRNA half-lives are shorter (**Figure 2C, 6I**).
379

380 **DISCUSSION**

381 In this study, we identified Arabidopsis TuORFs using improved Ribo-seq data and
382 systematically examined their presumed roles in gene expression. While our results support the
383 expected roles of uORFs in repressing mORF translation and mRNA stability, our study reveals
384 distinct features and regulation of uORFs in plants compared with animals.

385 Our observations that plant TuORF genes have higher expression levels (**Figure 2**),
386 encode larger proteins (**Figure 3**), and are enriched for evolutionarily conserved domains
387 (**Table S2**) raised an interesting question as to whether older genes are more likely to be
388 controlled at the protein level through mRNA degradation and translation via uORFs. Higher
389 mRNA abundance may allow TuORF mRNAs to be highly translated under certain conditions
390 that require a high level of specific proteins. For example, it was suggested that derepression of
391 downstream translation is a general mechanism of uORF-mediated expression under stress
392 (Andreev et al., 2018). Thus, TuORFs could enable fast and dynamic regulation of mORF
393 translation. Future work monitoring the translation of TuORFs and their mORFs under different
394 conditions will help decipher the regulation and physiological roles of specific TuORFs.

395 In this report, we present five pieces of evidence that Arabidopsis TuORFs generally do
396 not trigger NMD. First, fewer than 3% of TuORF genes were identified in the high-confidence
397 NMD target list (**Tabls S4B**). Second, NMD targets, but not TuORF mRNAs, are upregulated in
398 various NMD mutants (**Figure 4**). Third, TuORF mRNAs have significantly higher transcript
399 levels, while the NMD targets have lower transcript levels compared with other genes (**Figure**
400 **5A**). Fourth, NMD and TuORFs additively repress translation (**Figure 5B**). Finally, while NMD
401 targets are efficiently degraded through both 5' to 3' and 3' to 5' decay, TuORF mRNAs rely on
402 5' to 3' decay like most other genes (**Figure 6**). It is unlikely that a large number of TuORF
403 transcripts escape NMD altogether; rather, our results suggest TuORFs generally do not trigger
404 NMD in Arabidopsis.

405 The fact that Arabidopsis is missing key NMD factors found in animals, including SMG1,

406 SMG5, and SMG6, suggests significant variations in the NMD mechanisms between plants and
407 animals. In particular, SMG1 is the kinase responsible for UPF1 phosphorylation required for
408 substrate recognition in animal NMD (He and Jacobson, 2015; Kurosaki et al., 2019). The lack
409 of SMG1 and use of an alternative kinase for Arabidopsis UPF1 phosphorylation (Kerényi et al.,
410 2013) may lead to a difference in substrate recognition: while TuORF mRNAs are recognized as
411 NMD targets in animals, plant TuORF mRNAs are not.

412 Despite the difference in NMD components in plants and animals, our findings that
413 Arabidopsis NMD targets are degraded through both 5' to 3' decay and 3' to 5' decay is
414 consistent with the two-directional decay of NMD targets in animals. Note that plant SOV is the
415 homolog of DIS3L2, the exonuclease responsible for 3' to 5' decay of NMD targets in animals
416 (Chang et al., 2013; Lubas et al., 2013). Two-directional decay could be advantageous to
417 quickly remove aberrant mRNAs from the cytoplasm to prevent further translation and facilitate
418 the recycling of NMD machinery.

419 Genes controlled by both NMD and TuORFs might encode critical regulators, and their
420 protein levels require precise regulation. These genes included key enzymes in gibberellin
421 hormone and polyamine biosynthesis and many transcription factors (**Figure 1F-G, and Table**
422 **S4B**). For example, one of the transcription factors, *SUPPRESSOR OF ACAULIS 51* (*SAC51*),
423 was discovered in a genetic screen for suppressors of the dwarf phenotype of the *acaulis5*
424 mutant (Imai et al., 2006). When the uORF in *SAC51* was mutated, it increased the transcript
425 level of *SAC51*, which encodes a bHLH transcription factor that was proposed to activate the
426 genes required for stem elongation.

427 It is well established that there is an important interplay between translation efficiency
428 and mRNA stability. Two mechanisms were proposed to explain this phenomenon: the "stalled
429 ribosome-triggered decay model," which predicts that ribosome stalling or slowing promotes
430 mRNA degradation, and the "translation factor-protected model," which predicts that active
431 translation promotes mRNA stability because the translation initiation factors protect the mRNA

432 from the decapping complex (Presnyak et al., 2015; Radhakrishnan et al., 2016; Beelman and
433 Parker, 1994; LaGrandeur and Parker, 1999; Schwartz and Parker, 2000; Chan et al., 2018).
434 Consistent with the first model, several CPuORF-containing mRNAs in Arabidopsis were
435 reported to have strong ribosome stalling patterns, and the endonucleolytic cleavages upstream
436 of the stalled ribosomes were observed (Hou et al., 2016; Yu et al., 2016). While most of the
437 TuORFs we identified did not display clear stalling patterns (e.g., **Figure 1F-G, S1B, S2A-N**), it
438 remains possible that ribosomes have lower translation speeds, on average, during uORF
439 translation. Because ribosomes will quickly approach termination after initiation due to the short
440 length of uORFs, and because both initiation and termination are major time-consuming steps in
441 translation (Shah et al., 2013; Lawson et al., 2021), the average translation speed per codon for
442 uORFs is likely to be slower than that of longer ORFs. Alternatively, if cells mainly monitor the
443 translation efficiency of the mORF, then the poorer mORF translation caused by uORFs could
444 explain the faster degradation of TuORF mRNAs predicted by this model. In contrast, since
445 ribosomes actively engage with TuORFs, the “translation factor-protected” model does not fit
446 TuORF mRNAs, especially because their degradation relies on decapping and 5’ to 3’ decay in
447 plants. Further investigation into TuORF-mediated mRNA degradation will help elucidate the
448 role of uORFs in regulating mRNA stability.
449

450 **MATERIALS AND METHODS**

451 **Plant growth conditions and lysate preparation**

452 Arabidopsis Col-0 seeds were surface sterilized with 70% ethanol for 5 min, followed by 33%
453 bleach and 0.03% Tween 20 for 10 min, then rinsed with sterile water 5 times. The seeds were
454 imbibed at 4°C in the dark for 2 days, then grown hydroponically in sterile liquid media (2.15 g/L
455 Murashige and Skoog salt, 1% sucrose, 0.5 g/L MES, pH 5.7) while shaking at 85 rpm under a
456 16-h light (75–80 μmol $m^{-2} \cdot s^{-1}$ from cool white fluorescent bulbs) and 8-h dark cycle at 22°C for
457 7 days. At Zeitgeber time 4 (4 h after lights on), DMSO corresponding to 0.1% of the media
458 volume was added to the media (these were mock samples of our large-scale experiment). After
459 20 and 60 min, three biological replicates (~300 plants per sample) were harvested at each time
460 point and immediately flash frozen in liquid nitrogen.

461

462 Plant lysates were prepared as previously described (Hsu et al., 2016). Briefly, 0.1 g of ground
463 tissue power was resuspended in 400 μL of lysis buffer (100 mM Tris-HCl [pH 8], 40 mM KCl,
464 20 mM MgCl2, 2% [v/v] polyoxyethylene [10] tridecyl ether [Sigma, P2393], 1% [w/v] sodium
465 deoxycholate [Sigma, D6750], 1 mM dithiothreitol, 100 μg/mL cycloheximide [Sigma, C4859],
466 100 μg/mL chloramphenicol [Sigma R4408], and 10 units/mL DNase I [Epicenter, D9905K]).
467 The lysates were spun at 3,000 g for 3 min, and the supernatant was transferred to a new tube
468 and subsequently centrifuged at 20,000 g for 10 min. The supernatant was transferred to a new
469 tube and the RNA concentration was determined with 10x dilutions using the Qubit RNA HS
470 assay (Thermo Fisher Scientific; Q32852). Aliquots of 100 μL and 200 μL of the lysates were
471 made, and they were flash frozen in liquid nitrogen and stored at -80°C until further processing.

472

473 **Ribo-seq library construction**

474 Ribosome footprints were obtained using 200 μL of the lysates described above, and
475 sequencing libraries were constructed according to our previous method (Hsu et al., 2016) with

476 the following modifications: briefly, after RNase I digestion (50 units nuclease per 40 µg of RNA;
477 the nuclease was included in TruSeq Mammalian Ribo Profile Kit, Illumina, RPHMR12126) and
478 passing through a size exclusion column (Illustra MicroSpin S-400 HR Columns; GE Healthcare;
479 27-5140-01), RNA > 17 nt was isolated with the RNA Clean & Concentrator-25 kit (Zymo
480 Research, R1017) and separated on 15% urea-TBE gels (Invitrogen, EC68852BOX). Gel slices
481 between 28 and 30 nt were isolated, and the RNAs were purified as previously described. Next,
482 rRNA depletion was performed using RiboZero Plant Leaf kit (Illumina, MRZPL1224) in one
483 quarter of the recommended reaction volume. Ribo-seq libraries were then constructed using
484 the TruSeq Mammalian Ribo Profile Kit (illumina, RPHMR12126) as previously described with
485 nine cycles of PCR amplification. Libraries with equal molarity were pooled and sequenced on a
486 Hi-Seq 4000 sequencer using single-end 50-bp sequencing.
487
488 **RNA-seq library construction**
489 Total RNA greater than 200 nt was purified from 100 µL of the lysates described above using
490 the RNA Clean & Concentrator-25 kit (Zymo Research, R1017) as previously described (Hsu et
491 al., 2016). RNA integrity was evaluated using a Bioanalyzer (Agilent) RNA pico chip, and RNA
492 integrity numbers (RINs) ranging from 7.2 to 7.7 were observed among the samples. A total of 4
493 µg of RNA per sample was subjected to rRNA depletion using the RiboZero Plant Leaf kit
494 (Illumina, MRZPL1224) following the manufacturer's recommendations. Then, 100 ng of rRNA-
495 depleted RNA was fragmented to around 200-nt long based on the RIN reported by the
496 Bioanalyzer, and strand-specific sequencing libraries were made using the NEBNext Ultra II
497 Directional RNA Library Prep Kit (New England Biolabs, E7760S) with eight cycles of
498 amplification. Libraries of equal molarity were pooled and sequenced on a Hi-Seq 4000 using
499 paired-end 100-bp sequencing.
500
501 **Data pre-processing and analysis**

502 Data pre-processing and analysis were performed similarly to that previously described (Hsu et
503 al., 2016), except that the Araport11 annotation was used in this study. Briefly, for Ribo-seq
504 libraries, the adaptor (AGATCGGAAGAGCACACGTCT) was clipped with fastx_clipper (FASTX
505 toolkit v0.0.14) (http://hannonlab.cshl.edu/fastx_tool- kit/). For both RNA-seq and Ribo-seq, we
506 used Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012) to remove
507 rRNA/tRNA/snRNA/snoRNA sequences. Both the RNA-seq and Ribo-seq reads were mapped
508 to the transcriptome with STAR aligner (Dobin et al., 2013) (RNA-seq parameters: --
509 outFilterMismatchNmax 2 --outFilterMultimapNmax 20 --outFilterType BySJout --
510 alignSJoverhangMin 8 --alignSJDBoverhangMin 2; all parameters used for Ribo-seq were the
511 same except --outFilterMismatchNmax 1). Combining both 20-min and 60-min samples, the
512 total reads mapped to the genome for Ribo-seq and RNA-seq were 298.01 million and 180.67
513 million, respectively. We then used the bam files from both RNA-seq and Ribo-seq for
514 RiboTaper (Calviello et al., 2016) to identify translated ORFs. The Ribo-seq metaplot was
515 created using the create_metaplots.bash function in RiboTaper (v1.3.1a). The distribution of
516 Ribo-seq reads in different genome features was calculated using Ribo-seQC. The Ribo-seq
517 read lengths and offsets for RiboTaper were 24, 25, 26, 27, 28 and 8, 9, 10, 11, 12,
518 respectively. The discovered uORFs and mORFs were extracted from the RiboTaper output
519 ORF_max_filt file. To calculate TE, we first used STAR to map the RNA-seq and Ribo-seq
520 reads to the CDS of annotated coding genes. The resulting bam files were used to quantify the
521 transcripts per million (TPM) of each gene with RSEM (v1.3.1) (Li and Dewey, 2011). Then,
522 translation efficiency was calculated by dividing the Ribo-seq TPM by the RNA-seq TPM. The
523 root and shoot data from Hsu et al., 2016 (Hsu et al., 2016) were reanalyzed using the
524 Araport11 annotation. The P_sites_all files from RiboTaper were processed using the following
525 code: cut -f 1,3,6 P_sites_all | sort | uniq -c | sed -r 's/^(*[^]+) +/\1\t/' > output.txt to combine the
526 read counts at each P-site. Data visualization and statistical analysis were performed in R

527 (v4.0.3) (R Core Team (2013), 2017). In particular, the gene levels from the Ribo-seq and RNA-
528 seq reads were visualized with RiboPlotR (Wu and Hsu, 2021).
529
530 The mRNA half-life and mRNA-decay rate data were downloaded from Szabo et al. and
531 Sorenson et al., respectively (Szabo et al., 2020; Sorenson et al., 2018). The quantitative
532 proteomics data for Arabidopsis shoots and roots were downloaded from Song et al. (Song et
533 al., 2018). The Arabidopsis gene expression levels (TPM) used for investigating mRNA levels
534 for uORF genes in different Arabidopsis organs, growth stages, and ecotypes were directly
535 downloaded from the EMBO-EBI expression atlas (E-MTAB-7978: Arabidopsis tissue atlas, E-
536 GEOD-53197: 17 Arabidopsis thaliana accessions, E-MTAB-4812: tomato root, leaf, flower [two
537 stages], and fruit [six stages], and E-MTAB-4813: three longitudinal sections of six stages during
538 tomato fruit development) (Papatheodorou et al., 2020). Our previous Ribo-seq and RNA-seq
539 data for Arabidopsis roots and shoots were described in (Hsu et al., 2016); our previous Ribo-
540 seq and RNA-seq data for tomato roots were described in (Wu et al., 2019). The published
541 RNA-seq data for the NMD mutants *upf1* (*lba1*) *upf3-1* and its wild-type control (GEO:
542 GSE41432), *smg7-1 pad4* and its *pad4* control (GEO: GSE55884), *upf1-3 pad4* and its *pad4*
543 control (GEO: GSE146340), and *upf2-10* and its wild-type control (SRA: SRP056795) were
544 downloaded, and the mRNA levels were quantified by Kalisto (Bray et al., 2016) and the TPM
545 values were further analyzed to compare the steady-state mRNA levels for different categories
546 of genes.
547
548 **DATA ACCESS**
549 All raw and processed sequencing data generated in this study have been submitted to the
550 NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession
551 number GSE183264.
552

553 **COMPETING INTEREST STATEMENT**

554 The authors declare that there is no conflict of interest.

555

556 **ACKNOWLEDGEMENTS**

557 We thank Robert Last and Melissa Lehti-Shiu for helpful suggestions on the manuscript. This
558 work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported
559 by an NIH S10 OD018174 Instrumentation Grant. This work was supported by a National
560 Science Foundation grant (2051885) and a Michigan State University startup grant to PYH.

561

562 **AUTHOR CONTRIBUTIONS**

563 HLW and PYH designed the research, PYH performed the sequencing experiments, HLW
564 analyzed the sequencing data, HLW and PYH interpreted the data, and HLW and PYH wrote
565 the paper.

566

567 **REFERENCES**

568 **Andreev, D.E., Arnold, M., Kiniry, S.J., Loughran, G., Michel, A.M., Rachinskii, D., and**
569 **Baranov, P. V.** (2018). TASEP modelling provides a parsimonious explanation for the
570 ability of a single uORF to derepress translation during the integrated stress response.
571 eLife **7**: e32563.

572 **Von Arnim, A.G., Jia, Q., and Vaughn, J.N.** (2014). Regulation of plant translation by
573 upstream open reading frames. Plant Science **214**: 1–12.

574 **Baird, T.D., Cheng, K.C.C., Chen, Y.C., Buehler, E., Martin, S.E., Inglese, J., and Hogg, J.R.**
575 (2018). ICE1 promotes the link between splicing and nonsense-mediated mRNA decay.
576 eLife **7**.

577 **Bazin, J., Baerenfaller, K., Gosai, S.J., Gregory, B.D., Crespi, M., and Bailey-Serres, J.**
578 (2017). Global analysis of ribosome-associated noncoding RNAs unveils new modes of
579 translational regulation. Proceedings of the National Academy of Sciences of the United
580 States of America **114**: E10018–E10027.

581 **Beelman, C.A. and Parker, R.** (1994). Differential effects of translational inhibition in cis and in

582 trans on the decay of the unstable yeast MFA2 mRNA. Journal of Biological Chemistry
583 **269**: 9687–9692.
584 **Boudsocq, M., Barbier-Brygoo, H., and Laurière, C.** (2004). Identification of nine sucrose
585 nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in
586 Arabidopsis thaliana. Journal of Biological Chemistry **279**: 41758–41766.
587 **Brar, G.A. and Weissman, J.S.** (2015). Ribosome profiling reveals the what, when, where and
588 how of protein synthesis. Nature Reviews Molecular Cell Biology **16**: 651–664.
589 **Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L.** (2016). Near-optimal probabilistic RNA-
590 seq quantification. Nature biotechnology **34**: 525–527.
591 **Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler,**
592 **M., Obermayer, B., and Ohler, U.** (2016). Detecting actively translated open reading
593 frames in ribosome profiling data. Nature Methods **13**: 165–170.
594 **Calvo, S.E., Pagliarini, D.J., and Mootha, V.K.** (2009). Upstream open reading frames cause
595 widespread reduction of protein expression and are polymorphic among humans.
596 Proceedings of the National Academy of Sciences of the United States of America **106**:
597 7507–7512.
598 **Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K.** (2018). Non-invasive
599 measurement of mRNA decay reveals translation initiation as the major determinant of
600 mRNA stability. eLife **7**: e32536.
601 **Chang, H.M., Triboulet, R., Thornton, J.E., and Gregory, R.I.** (2013). A role for the Perlman
602 syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature **497**: 244–248.
603 **Chew, G.L., Pauli, A., and Schier, A.F.** (2016). Conservation of uORF repressiveness and
604 sequence features in mouse, human and zebrafish. Nature Communications **7**: 11663.
605 **Chiam, N.-C., Fujimura, T., Sano, R., Akiyoshi, N., Hiroyama, R., Watanabe, Y., Motose, H.,**
606 **Demura, T., and Ohtani, M.** (2019). Nonsense-Mediated mRNA Decay Deficiency Affects
607 the Auxin Response and Shoot Regeneration in Arabidopsis. Plant and Cell Physiology **60**:
608 2000–2014.
609 **Clarke, J.T., Warnock, R.C.M., and Donoghue, P.C.J.** (2011). Establishing a time-scale for
610 plant evolution. The New Phytologist **192**: 266–301.
611 **Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D.** (2015). Young genes out of
612 the male: An insight from evolutionary age analysis of the pollen transcriptome. Molecular
613 Plant **8**: 935–945.
614 **Degtiar, E., Fridman, A., Gottlieb, D., Vexler, K., Berezin, I., Farhi, R., Golani, L., and**
615 **Shaul, O.** (2015). The feedback control of UPF3 is crucial for RNA surveillance in plants.

616 Nucleic Acids Research **43**: 4219–4235.
617 **Dever, T.E., Ivanov, I.P., and Sachs, M.S.** (2020). Conserved Upstream Open Reading Frame
618 Nascent Peptides That Control Translation. Annual Review of Genetics **54**: 237–264.
619 **Devlin, P.F., Patel, S.R., and Whitelam, G.C.** (1998). Phytochrome E Influences Internode
620 Elongation and Flowering Time in Arabidopsis. The Plant Cell **10**: 1479–1487.
621 **Deyholos, M.K., Cavaness, G.F., Hall, B., King, E., Punwani, J., Van Norman, J., and**
622 **Sieburth, L.E.** (2003). Varicose, a WD-domain protein, is required for leaf blade.
623 Development **130**: 6577–6588.
624 **Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,**
625 **Chaisson, M., and Gingeras, T.R.** (2013). STAR: ultrafast universal RNA-seq aligner.
626 Bioinformatics **29**: 15–21.
627 **Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Rätsch, G., and**
628 **Wachter, A.** (2013). Nonsense-mediated decay of alternative precursor mRNA splicing
629 variants is a major determinant of the Arabidopsis steady state transcriptome. The Plant
630 Cell **25**: 3726–3742.
631 **Ellis, C.M., Nagpal, P., Young, J.C., Hagen, G., Guilfoyle, T.J., and Reed, J.W.** (2005).
632 AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence
633 and floral organ abscission in Arabidopsis thaliana. Development **132**: 4563–4574.
634 **Filichkin, S.A., Cumbie, J.S., Dharmawardhana, P., Jaiswal, P., Chang, J.H., Palusa, S.G.,**
635 **Reddy, A.S.N., Megraw, M., and Mockler, T.C.** (2015). Environmental stresses modulate
636 abundance and timing of alternatively spliced circadian transcripts in Arabidopsis.
637 Molecular Plant **8**: 207–227.
638 **Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., Bautor,**
639 **J., Dekrout, B., Jonak, C., Jiménez-Gómez, J.M., Parker, J.E., and Riha, K.** (2014).
640 Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant
641 antibacterial defense. Cell Host and Microbe **16**: 376–390.
642 **Goeres, D.C., Van Norman, J.M., Zhang, W., Fauver, N.A., Spencer, M. Lou, and Sieburth,**
643 **L.E.** (2007). Components of the Arabidopsis mRNA decapping complex are required for
644 early seedling development. The Plant Cell **19**: 1549–1564.
645 **Harnett, D., Meerdink, E., Calviello, L., Sydow, D., and Ohler, U.** (2021). Genome-Wide
646 Analysis of Actively Translated Open Reading Frames Using RiboTaper/ORFquant.
647 Methods in Molecular Biology **2252**: 331–346.
648 **Hayden, C.A. and Jorgensen, R.A.** (2007). Identification of novel conserved peptide uORF
649 homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups

650 and preferential association with transcription factor-encoding genes. BMC Biology **5**: 32.
651 **He, F. and Jacobson, A.** (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective
652 Transcripts Is only Part of the Story. Annual Review of Genetics **49**: 339–366.
653 **Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N.** (2016). Translational control by 5′-
654 untranslated regions of eukaryotic mRNAs. Science **352**: 1413–1416.
655 **Van Der Horst, S., Snel, B., Hanson, J., and Smeekens, S.** (2019). Novel pipeline identifies
656 new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid
657 sequences in the 5′ leader of mRNAs in Arabidopsis thaliana. RNA **25**: 292–304.
658 **Hou, C.-Y., Lee, W.-C., Chou, H.-C., Chen, A.-P., Chou, S.-J., and Chen, H.-M.** (2016).
659 Global analysis of truncated RNA ends reveals new insights into ribosome stalling in
660 plants. The Plant Cell **28**: 2398–2416.
661 **Hsu, P.Y., Calviello, L., Wu, H.-Y.L., Li, F.-W., Rothfels, C.J., Ohler, U., and Benfey, P.N.**
662 (2016). Super-resolution ribosome profiling reveals unannotated translation events in
663 Arabidopsis. Proceedings of the National Academy of Sciences of the United States of
664 America **113**: E7126–E7135.
665 **Hu, Q., Merchante, C., Stepanova, A.N., Alonso, J.M., and Heber, S.** (2016). Genome-Wide
666 Search for Translated Upstream Open Reading Frames in Arabidopsis Thaliana. IEEE
667 Transactions on Nanobioscience **15**: 150–159.
668 **Huq, E. and Quail, P.H.** (2002). PIF4, a phytochrome-interacting bHLH factor, functions as a
669 negative regulator of phytochrome B signaling in Arabidopsis. EMBO Journal **21**: 2441–
670 2450.
671 **Hurt, J.A., Robertson, A.D., and Burge, C.B.** (2013). Global analyses of UPF1 binding and
672 function reveal expanded scope of nonsense-mediated mRNA decay. Genome Research
673 **23**: 1636–1650.
674 **Imai, A., Hanzawa, Y., Komura, M., Yamamoto, K.T., Komeda, Y., and Takahashi, T.** (2006).
675 The dwarf phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an
676 upstream ORF of a bHLH gene. Development **133**: 3575–3585.
677 **Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S.** (2009). Genome-
678 wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.
679 Science **324**: 218–223.
680 **Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., and Maquat, L.E.** (2008).
681 Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA
682 Decay. Cell **133**: 314–327.
683 **Jeong, H.J., Kim, Y.J., Kim, S.H., Kim, Y.H., Lee, I.J., Kim, Y.K., and Shin, J.S.** (2011).

684 Nonsense-mediated mRNA decay factors, UPF1 and UPF3, contribute to plant defense.
685 Plant and Cell Physiology **52**: 2147–2156.
686 **Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J.** (2016). Upstream ORFs are prevalent
687 translational repressors in vertebrates. The EMBO Journal **35**: 706–723.
688 **Jorgensen, R.A. and Dorantes-Acosta, A.E.** (2012). Conserved peptide upstream open
689 reading frames are associated with regulatory genes in angiosperms. Frontiers in Plant
690 Science **3**: 191.
691 **Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J.** (2014). Translational dynamics
692 revealed by genome-wide profiling of ribosome footprints in Arabidopsis. Proceedings of
693 the National Academy of Sciences of the United States of America **111**: E203–E212.
694 **Kalyna, M. et al.** (2012). Alternative splicing and nonsense-mediated decay modulate
695 expression of important regulatory genes in Arabidopsis. Nucleic Acids Research **40**:
696 2454–2469.
697 **Kerényi, F., Wawer, I., Sikorski, P.J., Kufel, J., and Silhavy, D.** (2013). Phosphorylation of
698 the N- and C-terminal UPF1 domains plays a critical role in plant nonsense-mediated
699 mRNA decay. Plant Journal **76**: 836–848.
700 **Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E., and Silhavy, D.** (2006).
701 Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-
702 mediated decay in plants. Nucleic Acids Research **34**: 6147–6157.
703 **Kim, B.H., Cai, X., Vaughn, J.N., and Von Arnim, A.G.** (2007). On the functions of the h
704 subunit of eukaryotic initiation factor 3 in late stages of translation initiation. Genome
705 Biology **8**: R60.
706 **Kozak, M.** (1999). Initiation of translation in prokaryotes and eukaryotes. Gene **234**: 187–208.
707 **Kurihara, Y. et al.** (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs
708 by NMD in Arabidopsis. Proceedings of the National Academy of Sciences of the United
709 States of America **106**: 2453–2458.
710 **Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., and Matsui, M.** (2018).
711 Transcripts from downstream alternative transcription start sites evade uORF-mediated
712 inhibition of gene expression in Arabidopsis. Proceedings of the National Academy of
713 Sciences of the United States of America **115**: 7831–7836.
714 **Kurihara, Y., Makita, Y., Shimohira, H., Fujita, T., Iwasaki, S., and Matsui, M.** (2020).
715 Translational landscape of protein-coding and non-protein-coding RNAs upon light
716 exposure in Arabidopsis. Plant and Cell Physiology **61**: 536–545.
717 **Kurosaki, T., Li, W., Hoque, M., Popp, M.W.L., Ermolenko, D.N., Tian, B., and Maquat, L.E.**

718 (2014). A Post-Translational regulatory switch on UPF1 controls targeted mRNA
719 degradation. Genes and Development **28**: 1900–1916.
720 **Kurosaki, T., Popp, M.W., and Maquat, L.E.** (2019). Quality and quantity control of gene
721 expression by nonsense-mediated mRNA decay. Nature Reviews Molecular Cell Biology
722 2019 20:7 **20**: 406–420.
723 **LaGrandeur, T. and Parker, R.** (1999). The cis acting sequences responsible for the differential
724 decay of the unstable MFA2 and stable PGK1 transcipts in yeast include the context of the
725 translational start codon. RNA **5**: 420–433.
726 **Langmead, B. and Salzberg, S.L.** (2012). Fast gapped-read alignment with Bowtie 2. Nature
727 Methods **9**: 357–359.
728 **Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G.,**
729 **Hahn, M.G., Feldmann, K.A., and Benfey, P.N.** (1996). The SCARECROW gene
730 regulates an asymmetric cell division that is essential for generating the radial organization
731 of the Arabidopsis root. Cell **86**: 423–433.
732 **Lawson, M.R., Lessen, L.N., Wang, J., Prabhakar, A., Corsepius, N.C., Green, R., and**
733 **Puglisi, J.D.** (2021). Mechanisms that ensure speed and fidelity in eukaryotic translation
734 termination. Science **373**: 876–882.
735 **Lee, D.S.M., Park, J., Kromer, A., Baras, A., Rader, D.J., Ritchie, M.D., Ghanem, L.R., and**
736 **Barash, Y.** (2021). Disrupting upstream translation in mRNAs is associated with human
737 disease. Nature Communications 2021 12:1 **12**: 1–14.
738 **Leivar, P., Monte, E., Al-Sady, B., Carle, C., Storer, A., Alonso, J.M., Ecker, J.R., and Quail,**
739 **P.H.** (2008). The Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and
740 PIF4, regulates responses to prolonged red light by modulating phyB levels. The Plant Cell
741 **20**: 337–352.
742 **Li, B. and Dewey, C.N.** (2011). RSEM: accurate transcript quantification from RNA-Seq data
743 with or without a reference genome. BMC Bioinformatics **12**: 323.
744 **Li, Y.R. and Liu, M.J.** (2020). Prevalence of alternative AUG and non-AUG translation initiators
745 and their regulatory effects across plants. Genome Research **30**: 1418–1433.
746 **Lin, Y., May, G.E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., and Joel**
747 **McManus, C.** (2019). Impacts of uORF codon identity and position on translation
748 regulation. Nucleic Acids Research **47**: 9358–9367.
749 **Liu, M.J., Wu, S.H., Wu, J.F., Lin, W.D., Wu, Y.C., Tsai, T.Y., Tsai, H.L., and Wu, S.H.** (2013).
750 Translational landscape of photomorphogenic Arabidopsis. Plant Cell **25**: 3699–3710.
751 **Lloyd, J.P.B. and Davies, B.** (2013). SMG1 is an ancient nonsense-mediated mRNA decay

752 effector. The Plant journal : for cell and molecular biology **76**: 800–810.
753 **Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H., and Dziembowski, A.**
754 (2013). Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation
755 pathway of human cytoplasmic mRNA. The EMBO journal **32**: 1855–1868.
756 **Merchante, C., Brumos, J., Yun, J., Hu, Q., Spencer, K.R., Enríquez, P., Binder, B.M.,**
757 **Heber, S., Stepanova, A.N., and Alonso, J.M.** (2015). Gene-specific translation
758 regulation mediated by the hormone-signaling molecule EIN2. Cell **163**: 684–697.
759 **Mizoguchi, T., Hayashida, N., Yamaguchi-Shinozaki, K., Kamada, H., and Shinozaki, K.**
760 (1995). Two genes that encode ribosomal-protein S6 kinase homologs are induced by cold
761 or salinity stress in Arabidopsis thaliana. FEBS Letters **358**: 199–204.
762 **Nagpal, P., Ellis, C.M., Weber, H., Ploense, S.E., Barkawi, L.S., Guilfoyle, T.J., Hagen, G.,**
763 **Alonso, J.M., Cohen, J.D., Farmer, E.E., Ecker, J.R., and Reed, J.W.** (2005). Auxin
764 response factors ARF6 and ARF8 promote jasmonic acid production and flower
765 maturation. Development **132**: 4107–4118.
766 **Ni, M., Tepperman, J.M., and Quail, P.H.** (1998). PIF3, a phytochrome-interacting factor
767 necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix
768 protein. Cell **95**: 657–667.
769 **Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D.** (2009). Plant upstream
770 ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant
771 Molecular Biology **71**: 367–378.
772 **Ohtani, M. and Wachter, A.** (2019). NMD-Based Gene Regulation - A Strategy for Fitness
773 Enhancement in Plants? Plant and Cell Physiology **60**: 1953–1960.
774 **Papatheodorou, I. et al.** (2020). Expression Atlas update: From tissues to single cells. Nucleic
775 Acids Research **48**: D77–D83.
776 **Park, Y.S., Hong, S.W., Oh, S.A., Kwak, J.M., Lee, H.H., and Nam, H.G.** (1993). Two putative
777 protein kinases from Arabidopsis thaliana contain highly acidic domains. Plant Molecular
778 Biology **22**: 615–624.
779 **Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S.,**
780 **Weinberg, D., Baker, K.E., Graveley, B.R., and Coller, J.** (2015). Codon optimality is a
781 major determinant of mRNA stability. Cell **160**: 1111–1124.
782 **R Core Team (2013)** (2017). R: A language and environment for statistical computing. R
783 Foundation for Statistical Computing, Vienna, Austria.
784 **Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Coller, J.** (2016).
785 The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring

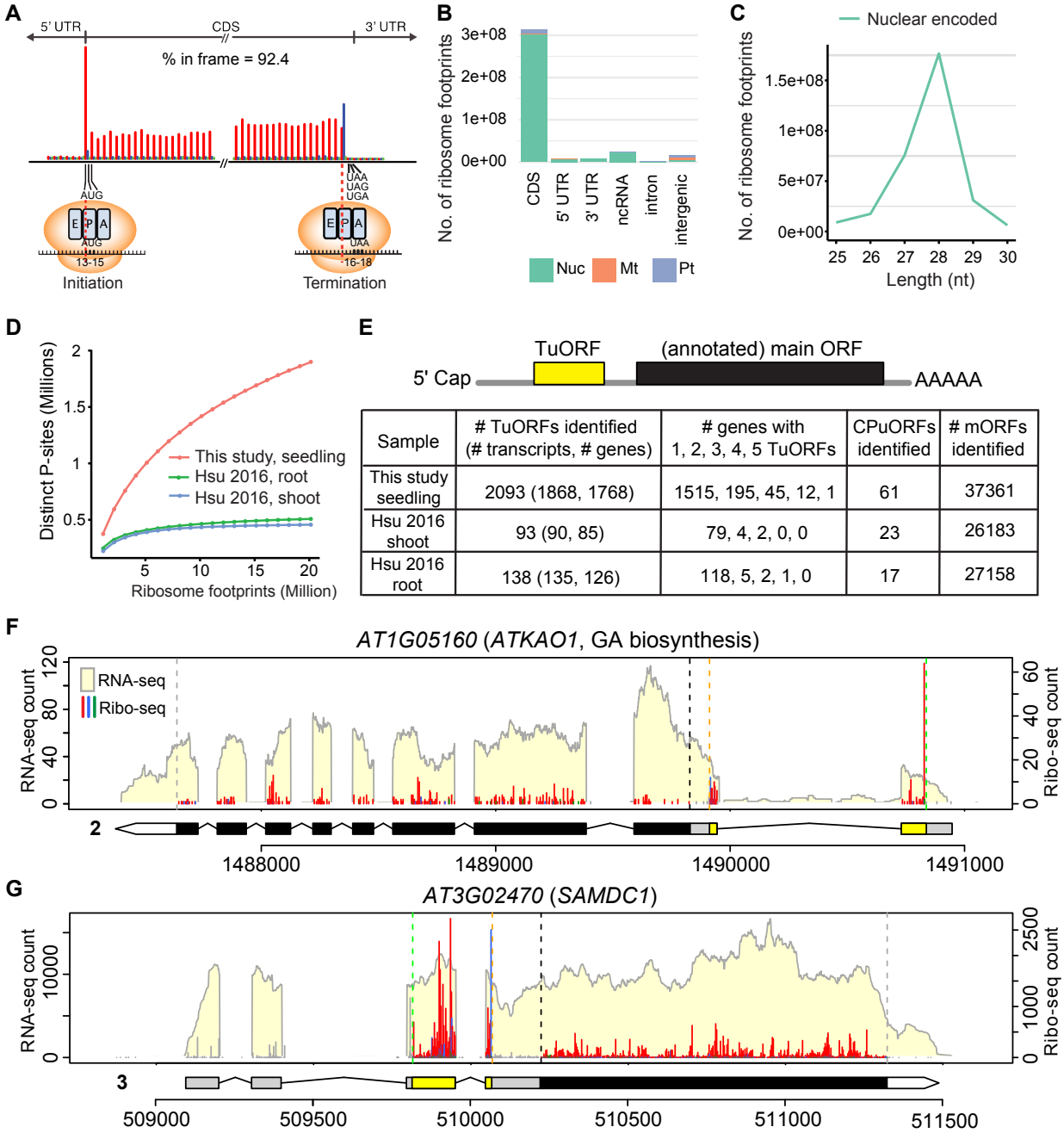
786 Codon Optimality. Cell **167**: 122-132.e9.
787 **Raxwal, V.K. and Riha, K.** (2016). Nonsense mediated RNA decay and evolutionary
788 capacitance. Biochimica et Biophysica Acta **1859**: 1538–1543.
789 **Raxwal, V.K., Simpson, C.G., Gloggnitzer, J., Entinze, J.C., Guo, W., Zhang, R., Brown,**
790 **J.W.S., and Riha, K.** (2020). Nonsense-Mediated RNA Decay Factor UPF1 Is Critical for
791 Posttranscriptional and Translational Gene Regulation in Arabidopsis. The Plant Cell **32**:
792 2725–2741.
793 **Rayson, S., Arciga-Reyes, L., Wootton, L., de Torres Zabala, M., Truman, W., Graham, N.,**
794 **Grant, M., and Davies, B.** (2012). A role for nonsense-mediated mrna decay in plants:
795 Pathogen responses are induced in arabidopsis thaliana nmd mutants. PLoS ONE **7**:
796 e31917.
797 **Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C., and Riha, K.** (2012). Aberrant
798 growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is
799 caused by autoimmune-like response. Nucleic Acids Research **40**: 5615–5624.
800 **Saul, H. et al.** (2009). The upstream open reading frame of the Arabidopsis AtMHX gene has a
801 strong impact on transcript accumulation through the nonsense-mediated mRNA decay
802 pathway. The Plant Journal **60**: 1031–1042.
803 **Schwartz, D.C. and Parker, R.** (2000). mRNA Decapping in Yeast Requires Dissociation of the
804 Cap Binding Protein, Eukaryotic Translation Initiation Factor 4E. Molecular and Cellular
805 Biology **20**: 7933–7942.
806 **Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B.** (2013). Rate-Limiting Steps in
807 Yeast Protein Translation. Cell **153**: 1589–1601.
808 **Sharrock, R.A. and Quail, P.H.** (1989). Novel phytochrome sequences in Arabidopsis thaliana:
809 structure, evolution, and differential expression of a plant regulatory photoreceptor family.
810 Genes & Development **3**: 1745–1757.
811 **Sieburth, L.E. and Vincent, J.N.** (2018). Beyond transcription factors: Roles of mrna decay in
812 regulating gene expression in plants [version 1; referees: 3 approved]. F1000Research **7**:
813 1940.
814 **Silverstone, A.L., Ciampaglio, C.N., and Sun, T.P.** (1998). The Arabidopsis RGA gene
815 encodes a transcriptional regulator repressing the gibberellin signal transduction pathway.
816 The Plant Cell **10**: 155–169.
817 **Song, G., Hsu, P.Y., and Walley, J.W.** (2018). Assessment and Refinement of Sample
818 Preparation Methods for Deep and Quantitative Plant Proteome Profiling. Proteomics **18**:
819 e1800220.

820 **Sorenson, R.S., Deshotel, M.J., Johnson, K., Adler, F.R., and Sieburth, L.E.** (2018).
821 Arabidopsis mRNA decay landscape arises from specialized RNA decay substrates,
822 decapping-mediated feedback, and redundancy. Proceedings of the National Academy of
823 Sciences of the United States of America **115**: E1485–E1494.
824 **Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps,**
825 **J.A., and Kay, S.A.** (2000). Cloning of the Arabidopsis clock gene TOC1, an
826 autoregulatory response regulator homolog. Science **289**: 768–771.
827 **Szabo, E.X., Reichert, P., Lehniger, M.K., Ohmer, M., de Francisco Amorim, M., Gowik, U.,**
828 **Schmitz-Linneweber, C., and Laubinger, S.** (2020). Metabolic labeling of RNAs uncovers
829 hidden features and dynamics of the Arabidopsis transcriptome. The Plant Cell **32**: 871–
830 887.
831 **Takahashi, H., Takahashi, A., Naito, S., and Onouchi, H.** (2012). BAIUCAS: A novel BLAST-
832 based algorithm for the identification of upstream open reading frames with conserved
833 amino acid sequences and its application to the Arabidopsis thaliana genome.
834 Bioinformatics **28**: 2231–2241.
835 **Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., Chiba, Y., Hirai,**
836 **M.Y., Akiyama, T., Onouchi, H., Naito, S., and Fujiwara, T.** (2016). The minimum open
837 reading frame, AUG-Stop, induces boron-dependent ribosome stalling and mRNA
838 degradation. The Plant Cell **28**: 2830–2849.
839 **Uchiyama-Kadokura, N., Murakami, K., Takemoto, M., Koyanagi, N., Murota, K., Naito, S.,**
840 **and Onouchi, H.** (2014). Polyamine-responsive ribosomal arrest at the stop codon of an
841 upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA
842 decay in Arabidopsis thaliana. Plant and Cell Physiology **55**: 1556–1567.
843 **Vaughn, J.N., Ellingson, S.R., Mignone, F., and Arnim, A. von** (2012). Known and novel
844 post-transcriptional regulatory sequences are conserved across plant families. RNA **18**:
845 368–84.
846 **Wolfa, Y.I., Novichkovb, P.S., Kareva, G.P., Koonina, E. V., and Lipmana, D.J.** (2009). The
847 universal distribution of evolutionary rates of genes and distinct characteristics of
848 eukaryotic genes of different apparent ages. Proceedings of the National Academy of
849 Sciences of the United States of America **106**: 7273–7280.
850 **Wu, H.-Y.L. and Hsu, P.Y.** (2021). RiboPlotR: a visualization tool for periodic Ribo-seq reads.
851 Plant Methods **17**: 1–7.
852 **Wu, H.-Y.L., Song, G., Walley, J.W., and Hsu, P.Y.** (2019). Translational landscape in tomato
853 revealed by transcriptome assembly and ribosome profiling. bioRxiv **181**: 367–380.

854 **Xu, G., Yuan, M., Ai, C., Liu, L., Zhuang, E., Karapetyan, S., Wang, S., and Dong, X.** (2017).
855 uORF-mediated translation allows engineered plant disease resistance without fitness
856 costs. Nature **545**: 491–494.
857 **Xu, J., Yang, J.Y., Niu, Q.W., and Chua, N.H.** (2006). Arabidopsis DCP2, DCP1, and
858 VARICOSE form a decapping complex required for postembryonic development. The Plant
859 Cell **18**: 3386–3398.
860 **Xu, Z., Hu, L., Shi, B., Geng, S., Xu, L., Wang, D., and Lu, Z.J.** (2018). Ribosome elongating
861 footprints denoised by wavelet transform comprehensively characterize dynamic cellular
862 translation events. Nucleic Acids Research **46**: e109.
863 **Yoine, M., Nishii, T., and Nakamura, K.** (2006). Arabidopsis UPF1 RNA helicase for
864 nonsense-mediated mRNA decay is involved in seed size control and is essential for
865 growth. Plant and Cell Physiology **47**: 572–580.
866 **Yu, X., Willmann, M.R., Anderson, S.J., and Gregory, B.D.** (2016). Genome-wide mapping of
867 uncapped and cleaved transcripts reveals a role for the nuclear mrna cap-binding complex
868 in cotranslational rna decay in Arabidopsis. The Plant Cell **28**: 2385–2397.
869 **Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C.** (2018). Genome
870 editing of upstream open reading frames enables translational control in plants. Nature
871 Biotechnology **36**: 894–898.
872 **Zhang, H., Wang, Y., Wu, X., Tang, X., Wu, C., and Lu, J.** (2021). Determinants of genome-
873 wide distribution and evolution of uORFs in eukaryotes. Nature Communications **12**: 1–17.
874 **Zhang, S.H., Lawton, M.A., Hunter, T., and Lamb, C.J.** (1994). atpk1, a novel ribosomal
875 protein kinase gene from Arabidopsis. I. Isolation, characterization, and expression.
876 Journal of Biological Chemistry **269**: 17586–17592.
877 **Zhang, W., Murphy, C., and Sieburth, L.E.** (2010). Conserved RNaseII domain protein
878 functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant
879 phenotypes. Proceedings of the National Academy of Sciences of the United States of
880 America **107**: 15981–15985.
881 **Zinshteyn, B., Sinha, N.K., Enam, S.U., Koleske, B., and Green, R.** (2021). Translational
882 repression of NMD targets by GIGYF2 and EIF4E2. PLoS Genetics **17**: e1009813.
883

884 **FIGURE and FIGURE LEGENDS**

Figure 1



Sample	# TuORFs identified (# transcripts, # genes)	# genes with 1, 2, 3, 4, 5 TuORFs	CPuORFs identified	# mORFs identified
This study seedling	2093 (1868, 1768)	1515, 195, 45, 12, 1	61	37361
Hsu 2016 shoot	93 (90, 85)	79, 4, 2, 0, 0	23	26183
Hsu 2016 root	138 (135, 126)	118, 5, 2, 1, 0	17	27158

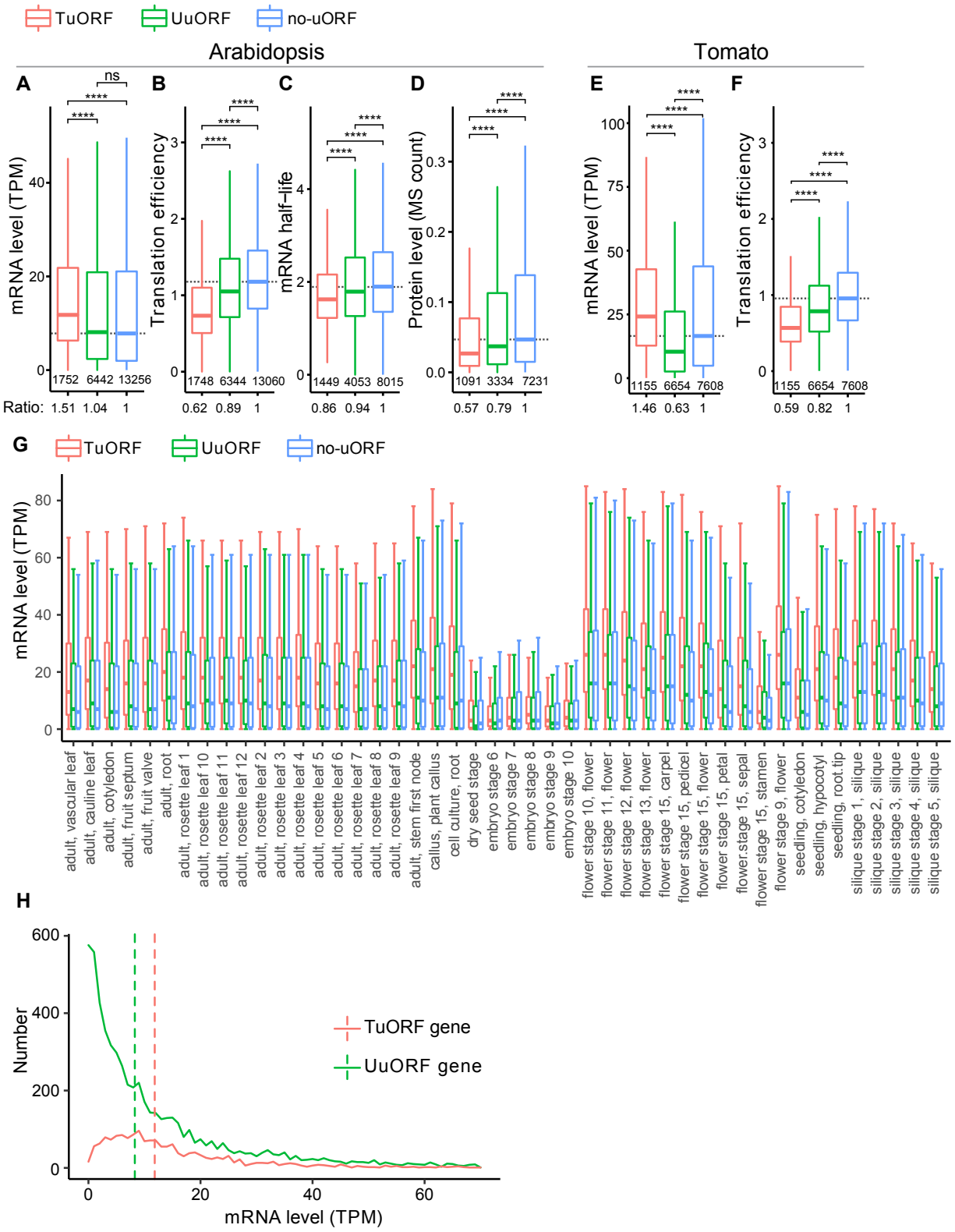
885

886 **Figure 1. Enhanced Ribo-seq coverage improves uORF identification**

887 (A) Metagene analysis of 28-nt ribosome footprints mapped to regions near the start and
888 stop codons of annotated ORFs. The reads are presented with their first nt at the P-site,
889 which is the 13th nt for 28-nt footprints. The reads are colored in red, blue, and green to

890 indicate they are in the first (expected), second, and third reading frames, respectively.
891 The majority of footprints were mapped to the CDS in the expected reading frame
892 (92.4% in frame).
893 (B) Genomic features mapped based on ribosome footprints. Reads that mapped to nuclear
894 (Nuc)-, mitochondrial (Mt)-, and plastid (Pt)-encoded genes are shown.
895 (C) Length distribution of ribosome footprints. Reads that mapped to nuclear-encoded genes
896 are presented.
897 (D) Distinct P-sites detected in 1 to 20 million randomly selected ribosome footprints from
898 our current and previous datasets.
899 (E) Numbers of TuORFs, CPuORFs, and mORFs identified in our current and previous
900 datasets.
901 (F-G) RNA-seq and Ribo-seq profiles of *ATKAO1 and SAMDC1* from our current data. RNA-
902 seq coverage is shown with a light-yellow background. Ribo-seq reads are presented
903 with their first nt at the P-site, and they are colored in red, blue, and green to indicate
904 they are in the first (expected), second, and third reading frames, respectively. Within the
905 gene models, yellow boxes represent the TuORFs, black boxes represent the annotated
906 mORFs, and gray and white regions indicate 5' UTRs and 3' UTRs, respectively. The
907 specific isoform being considered is indicated to the left of the gene model. Within the
908 profiles, green and orange vertical dashed lines represent translation start and stop,
909 respectively, for the TuORF. Black and gray vertical dashed lines represent translation
910 start and stop, respectively, for the annotated mORF.
911
912
913
914
915
916

Figure 2


TuORF
UuORF
no-uORF
Arabidopsis
Tomato
A
B
C
D
E
F
G
H
mRNA level (TPM)
Translation efficiency
mRNA half-life
Protein level (MS count)
ns
****
1752 6442 13256
1748 6344 13060
1449 4053 8015
1091 3334 7231
1155 6654 7608
1155 6654 7608
Ratio: 1.51 1.04 1
0.62 0.89 1
0.86 0.94 1
0.57 0.79 1
1.46 0.63 1
0.59 0.82 1
adult, vascular leaf
adult, cauline leaf
adult, cotyledon
adult, fruit septum
adult, fruit valve
adult, root
adult, rosette leaf 1
adult, rosette leaf 10
adult, rosette leaf 11
adult, rosette leaf 12
adult, rosette leaf 2
adult, rosette leaf 3
adult, rosette leaf 4
adult, rosette leaf 5
adult, rosette leaf 6
adult, rosette leaf 7
adult, rosette leaf 8
adult, rosette leaf 9
adult, stem first node
callus, plant callus
cell culture, root
dry seed stage
embryo stage 6
embryo stage 7
embryo stage 8
embryo stage 9
embryo stage 10
flower stage 10, flower
flower stage 11, flower
flower stage 12, flower
flower stage 13, flower
flower stage 15, carpel
flower stage 15, pedicel
flower stage 15, flower
flower stage 15, petal
flower.stage 15, sepal
flower stage 15, stamen
flower stage 9, flower
seedling, cotyledon
seedling, hypocotyl
seedling, root.tip
silique stage 1, silique
silique stage 2, silique
silique stage 3, silique
silique stage 4, silique
silique stage 5, silique
Number
TuORF gene
UuORF gene

919 **Figure 2. TuORF-containing genes have higher mRNA abundance but are associated with**
920 **lower TE, shorter mRNA half-lives, and lower protein abundance**

921 (A-D) The mRNA levels, translation efficiencies, mRNA half-lives, and protein levels of TuORF,
922 UuORF, and no-uORF genes in Arabidopsis seedlings. The mRNA half-life data, which
923 were measured with metabolic labeling in Arabidopsis seedlings, were extracted from
924 (Szabo et al., 2020), and the protein abundances as determined by quantitative
925 proteomics in Arabidopsis seedling shoots were extracted from (Song et al., 2018). Within
926 the boxplots, the dashed lines mark the median level of the no-uORF genes; the number
927 of genes in each category is listed below the lower whisker. The ratios underneath the
928 plots indicate the median of each group normalized to that of the no-uORF genes. The
929 statistical significance was determined by Wilcoxon signed-rank test. The adjusted p-value
930 was determined using the Benjamini and Yekutieli procedure to control for the false
931 discovery rate in multiple testing (* 0.05 > p > 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-
932 4, **** 1e-4 > p > 0). These statistical analyses and graphic layouts are the same for all
933 boxplots in the study.

934 (E-F) The mRNA levels and translation efficiencies of TuORF, UuORF, and no-uORF genes
935 from reanalysis of our previous tomato root data (Wu et al., 2019). The dashed lines mark
936 the median level of the no-uORF genes.

937 (G) The mRNA levels of TuORF, UuORF, and no-uORF genes in 46 Arabidopsis tissues and
938 developmental stages from independent studies. The TPM values from RNA-seq
939 quantification were downloaded from the EMBO-EBI expression atlas (Papatheodorou et
940 al., 2020).

941 (H) The distributions of the mRNA levels of TuORF and UuORF genes in Arabidopsis seedlings.
942 Vertical dashed lines indicate the median value of each group.

943
944

Figure 3

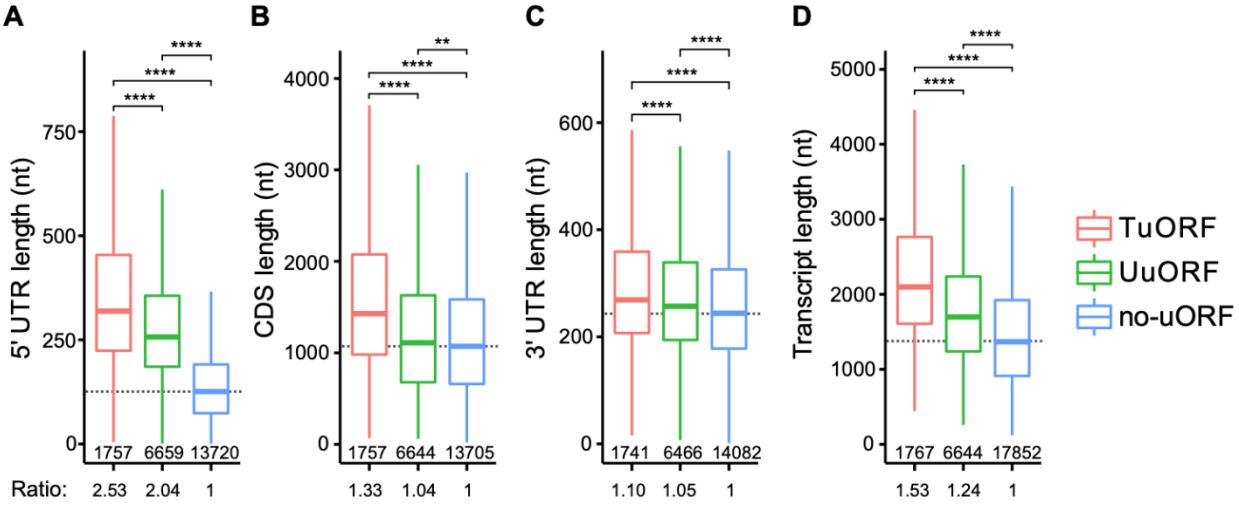


Figure 3. TuORF genes are larger than other genes

Lengths of the 5' UTRs (A), CDSs (B), 3' UTRs (C), and entire transcripts (D) of TuORF, UuORF, and no-uORF genes. Within the boxplots, the dashed lines mark the median level of the no-uORF genes; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of the no-uORF genes. The statistical significance was determined as described in the Figure 2 legend.

Figure 4

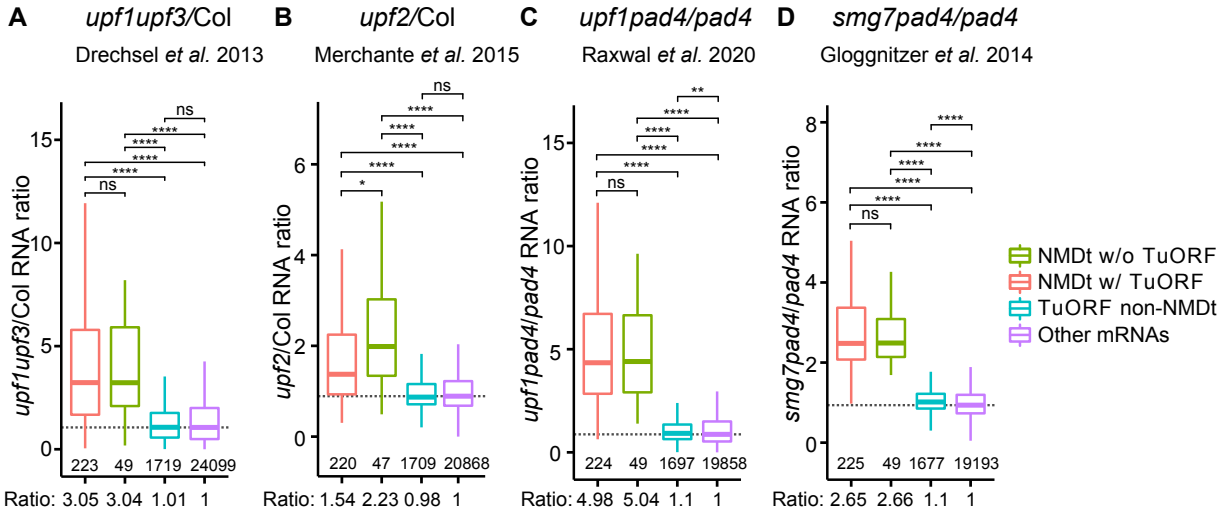


Figure 4: TuORF mRNAs are not upregulated in the NMD mutants

The ratios of mRNA levels between NMD mutants and their control plants from four RNA-seq datasets (Drechsel et al., 2013; Merchante et al., 2015; Raxwal et al., 2020; Gloggnitzer et al., 2014) were used to examine the effects of the NMD pathway on NMD targets (NMDt) and TuORF genes. (A) *upf1* (*lba1* allele) *upf3-1* double mutant compared with wild type (Col-0). (B) *upf2-10* compared with wild type (Col-0). (C) *upf1-3 pad4* compared with *pad4*. (D) *smg7-1 pad4* compared to *pad4*. See Table S3 for specific growth conditions for each dataset. Within the boxplots, the dashed lines mark the median level of other mRNAs; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of other mRNAs. The statistical significance was determined as described in the Figure 2 legend.

Figure 5

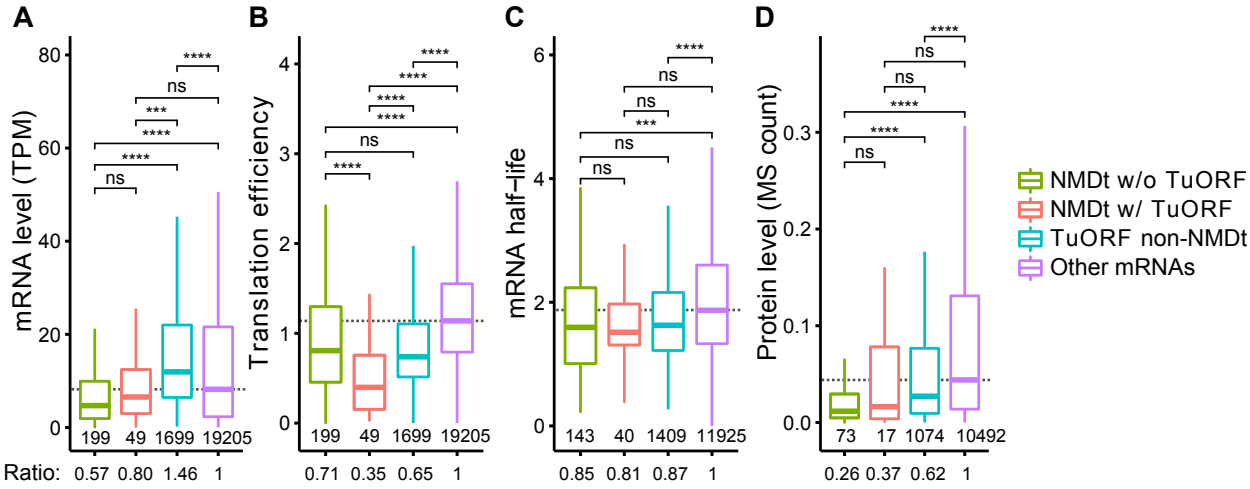


Figure 5. Expression patterns of NMD targets and TuORF genes.

(A-D) The mRNA levels (A), translation efficiencies (B), mRNA half-lives (C), and protein levels (D) of NMD targets (NMDt) without or with TuORFs compared with those of TuORF mRNAs and other mRNAs. Within the boxplots, the dashed lines mark the median level of other mRNAs; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of other mRNAs. The statistical significance was determined as described in the Figure 2 legend.

Figure 6

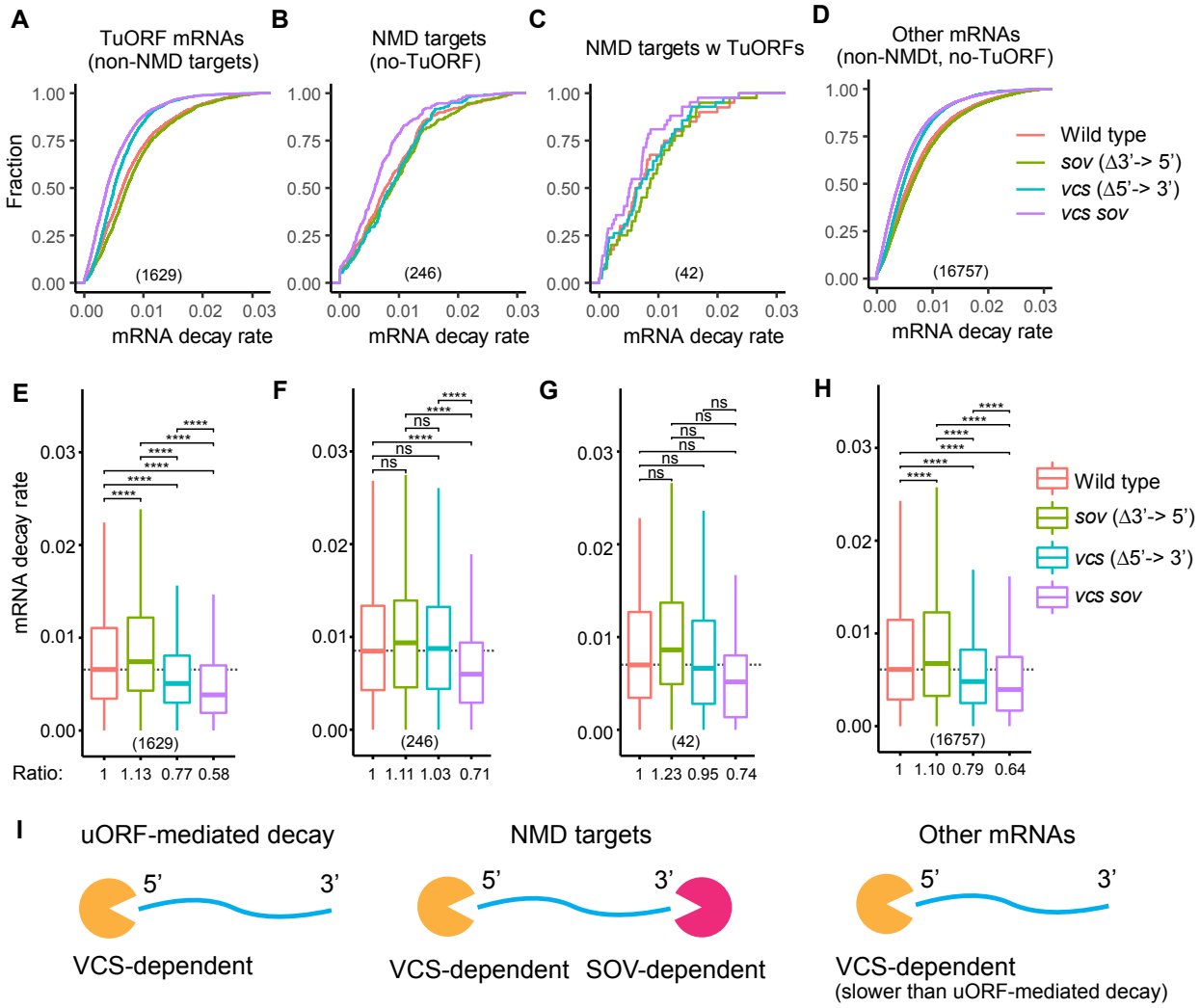


Figure 6. TuORF mRNAs and NMD targets are degraded through different mRNA decay mechanisms

Cumulative plots (A-D) and boxplots (E-H) showing the mRNA decay rates for NMD targets and TuORF genes in wild type and mutants defective in either 5' to 3' decay (*vcs-7*), 3' to 5' decay (*sov*), or both (*vcs sov).* The mRNA decay rates were extracted from (Sorenson et al., 2018). Because Col-0 is naturally an *sov* mutant, the wild type here is Col-0 carrying the functional *SOV* allele from Landsberg *erecta* driven by its native promoter (Zhang et al., 2010; Goeres et al., 2007). The *sov* mutant here is Col-0, and *vcs* is Col-0 with the *vsc-7* allele and the L*er SOV* transgene. The *vcs sov* double mutant is Col-0 carrying the *vcs-7* mutation (Sorenson et al.,

987 2018). Within the boxplots, the dashed lines mark the median level of wild type; the number of
988 genes in each category is listed below the lower whisker. The ratios underneath the plots
989 indicate the median of each group normalized to that of wild type. The statistical significance
990 was determined as described in the Figure 2 legend. (I) Model of mRNA decay for NMD targets,
991 TuORF mRNAs, and other mRNAs. Orange packman indicates 5'-to-3' decay machinery and
992 hot pink packman indicates 3'-to-5' decay machinery.
993

994 **SUPPLEMENTARY DATA**

995 **Figure S1.** Correlations among samples and comparisons between our previous and current
996 datasets

997 **Figure S2.** Examples of TuORFs in important regulatory genes.

998 **Figure S3.** mRNA levels of uORF-containing genes in different ecotypes/cultivars, growth
999 stages, tissues in Arabidopsis and tomato

1000 **Figure S4.** mRNA levels of CPuORF-containing genes compared to no-uORF genes.

1001 **Figure S5.** Protein abundance of the uORF-containing genes in Arabidopsis root.

1002 **Figure S6.** Distributions of uORF peptide lengths in NMD target and non-NMD target TuORF
1003 genes.

1004 **Figure S7.** Length of UTRs and CDS for NMD targets and TuORF genes

1005

1006 **Table S1**. Translated ORFs identified by RiboTaper (Excel)

1007 Table S1A-C. Translated uORFs identified in the current dataset, our previous Arabidopsis
1008 shoot and root datasets

1009 Table S1D-F. Translated mORFs identified in the current dataset, our previous Arabidopsis
1010 shoot and root datasets

1011 **Table S2.** GO-term analsyis of TuORF genes (Excel)

1012 **Table S3.** Summary of the large-scale datasets used in this study

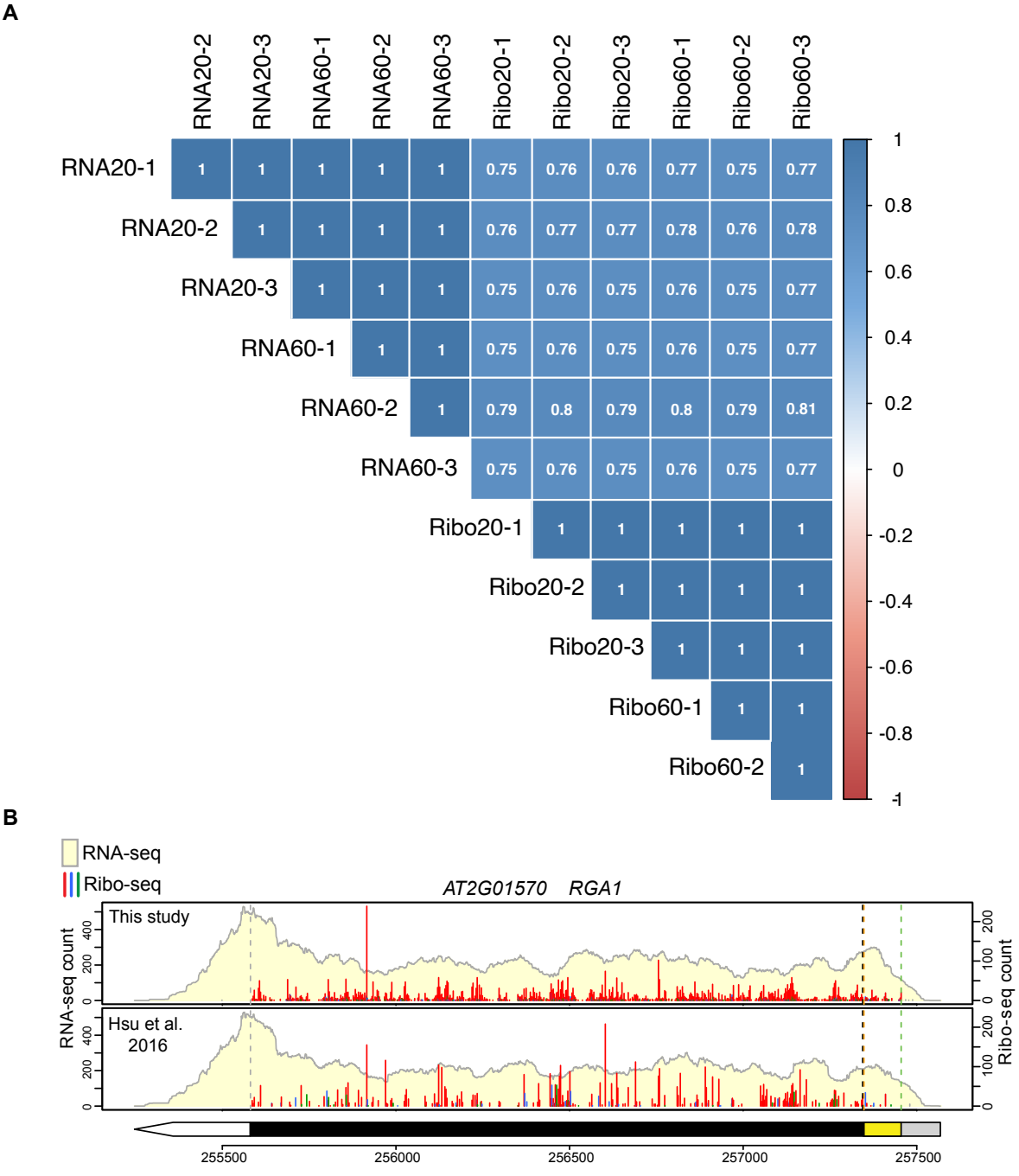
1013 **Table S4.** NMD target gene list (Excel)

1014 Table S4A. Overlap between Raxwal et al. high-confidence NMD targets and our TuORF genes

1015 Table S4B. 49 genes identified as high-confidence NMD target and TuORF genes

1016

1017 **SUPPLEMENTARY FIGURES**



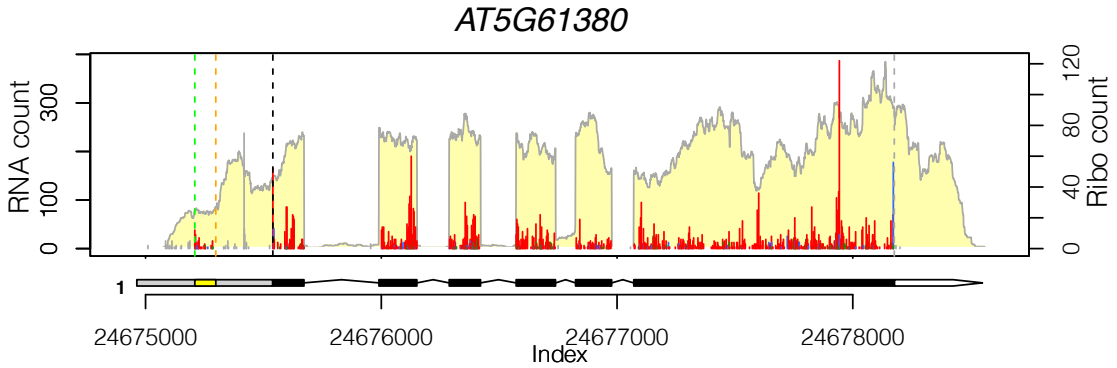
1018

1019 **Figure S1. Correlations among samples and comparisons between our previous and**
1020 **current RNA-seq and Ribo-seq datasets.** (A) Pearson correlation among three biological
1021 replicates of our current RNA-seq and Ribo-seq analyses of samples harvested after 20 or 60
1022 min of DMSO treatment (these are control samples of our large-scale experiment). TPM values
1023 in the CDSs of annotated protein-coding genes were used. (B) Higher Ribo-seq read coverage

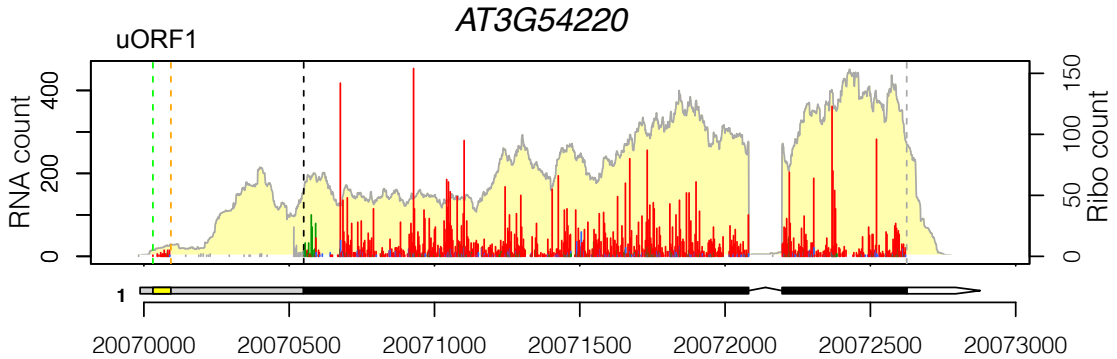
1024 was obtained in both the uORF and mORF regions in our current dataset; *RGA1* is shown as an
1025 example. Note that the mRNA levels are similar between our previous data (Hsu et al. 2016,
1026 shoot) and current data. Data representations are the same as those described in the Figure 1F
1027 legend.

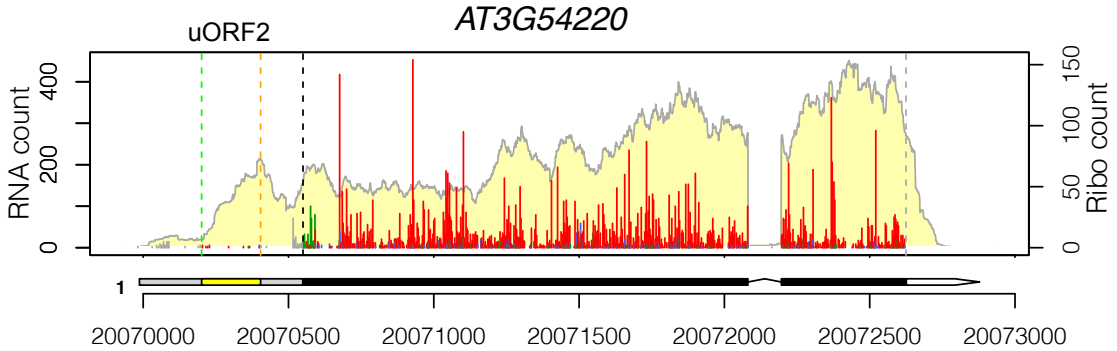
Figure S2. TuORFs in known important regulatory genes

A. *TOC1*

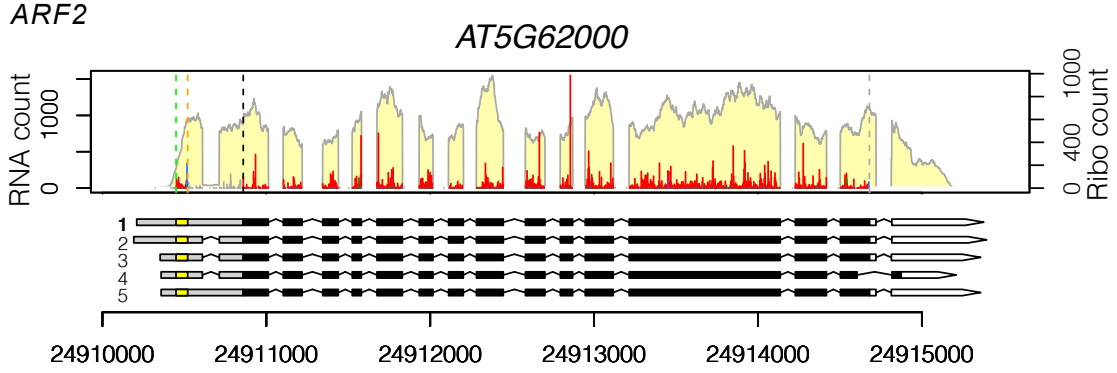


B. *SCARECROW*



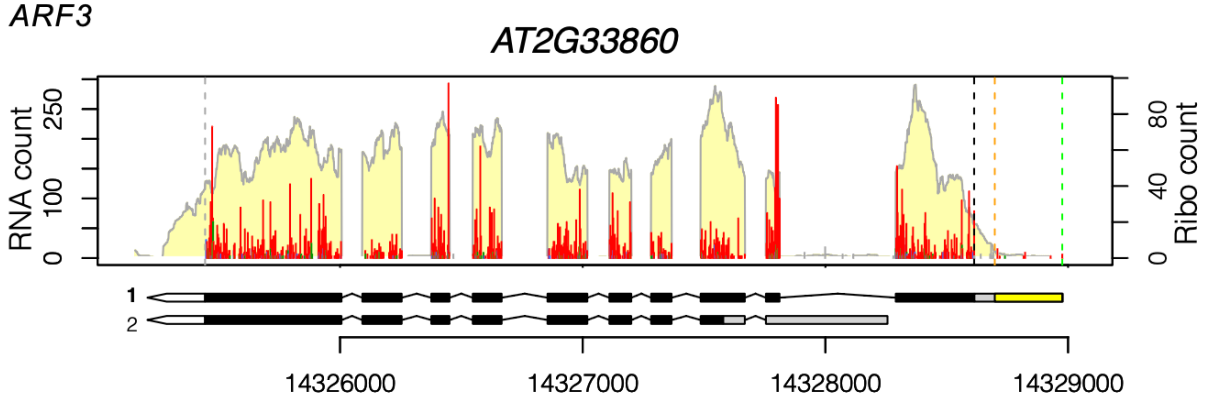


C. *ARF2*

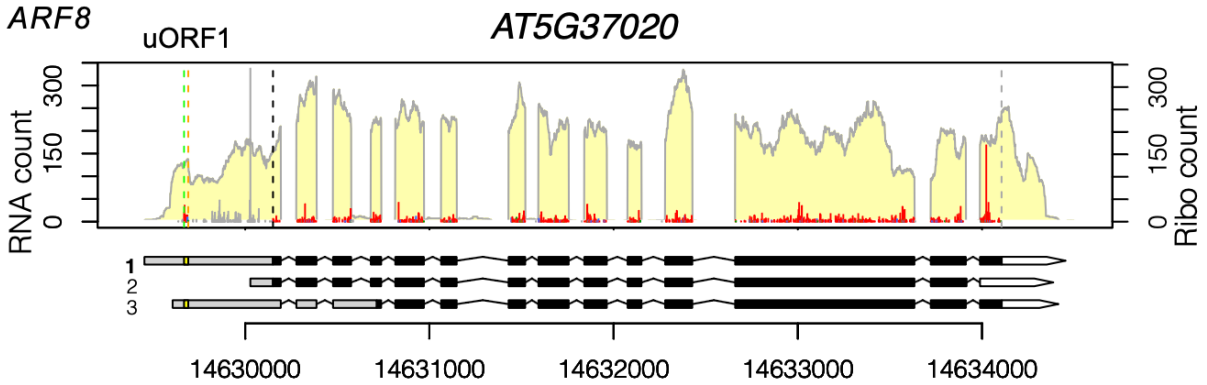


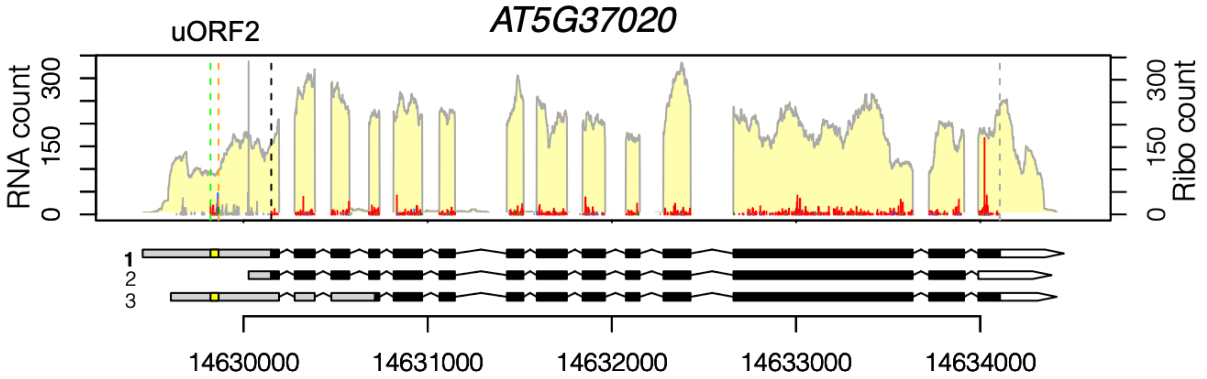
1028

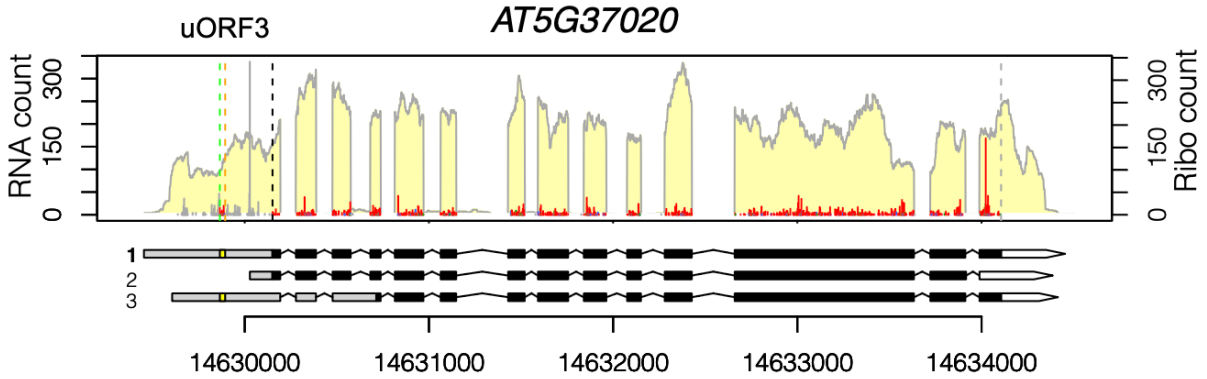
D. *ARF3*



E. *ARF8*



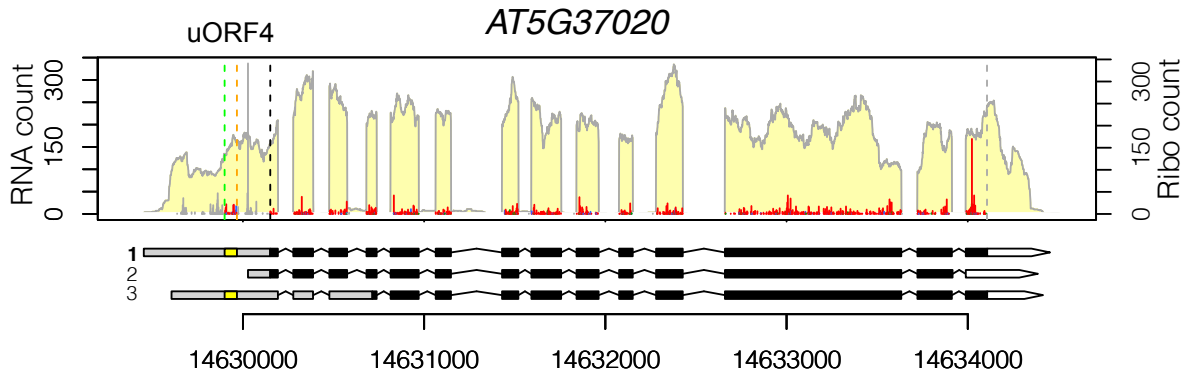




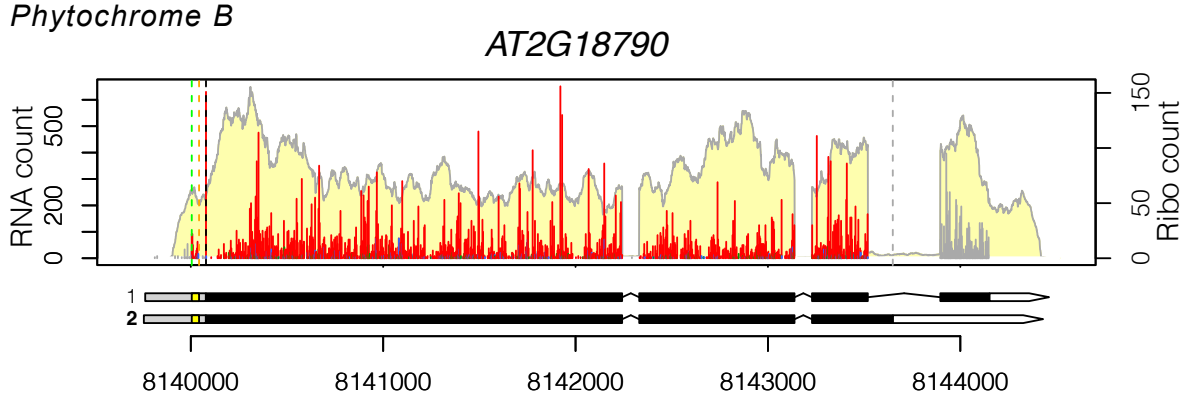
1029

1030

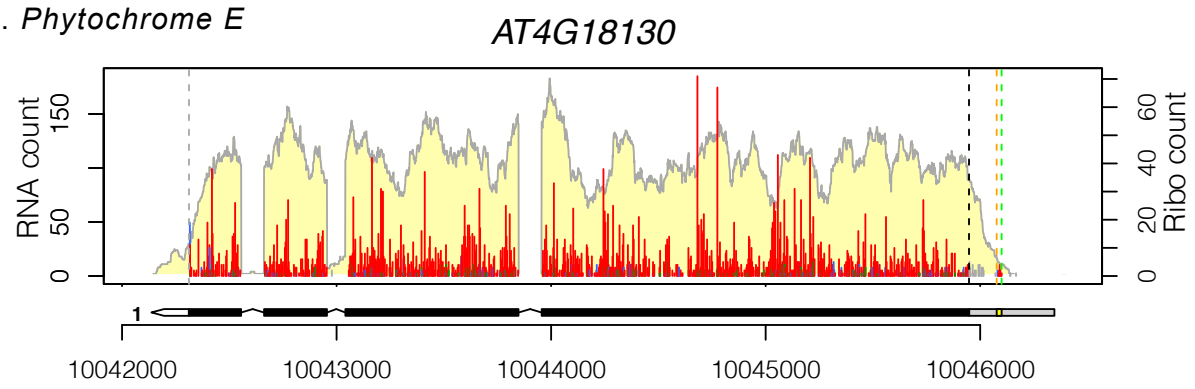
E. *ARF8* (continued)



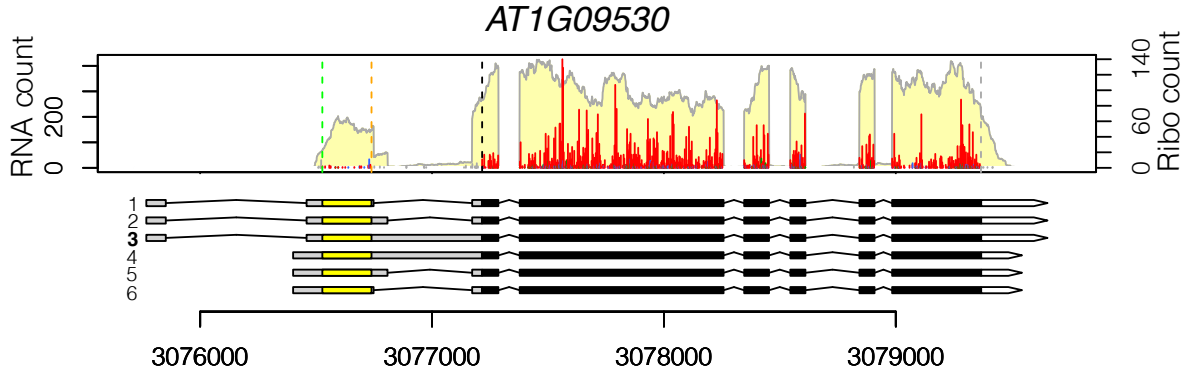
F. *Phytochrome B*



G. *Phytochrome E*

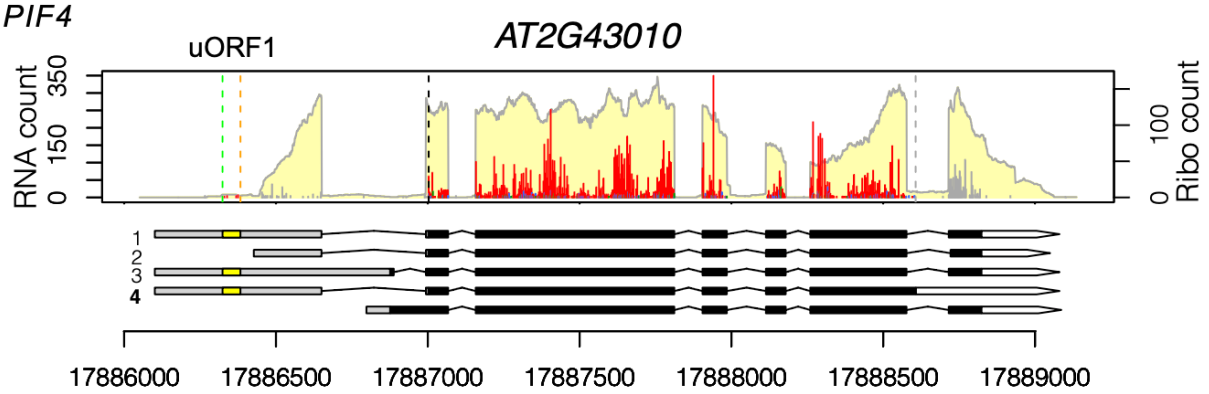


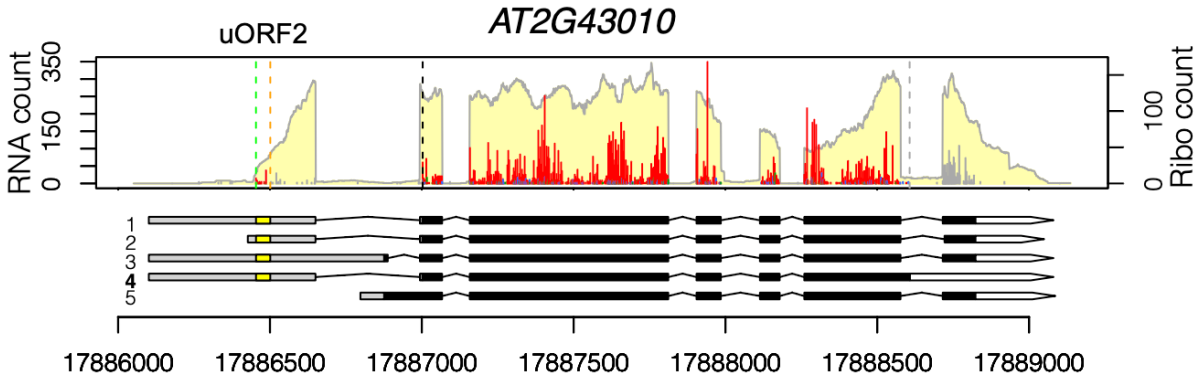
H. *PIF3*



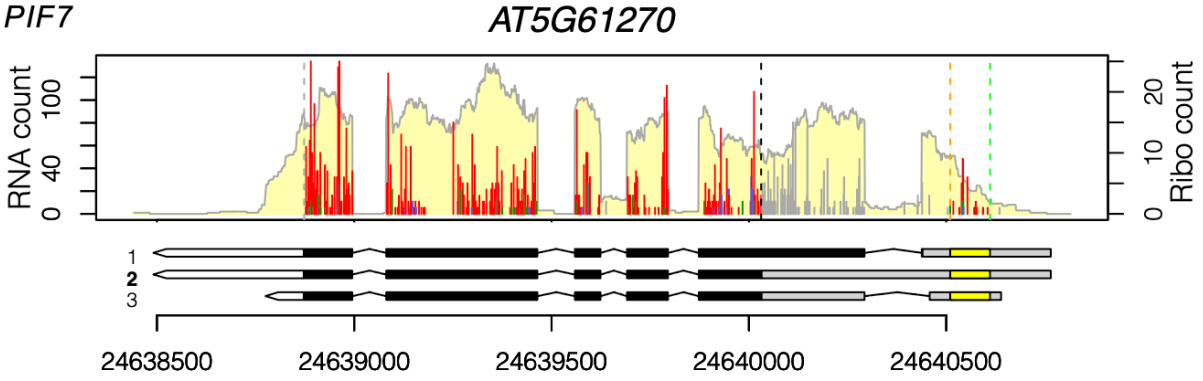
1031

I. *PIF4*

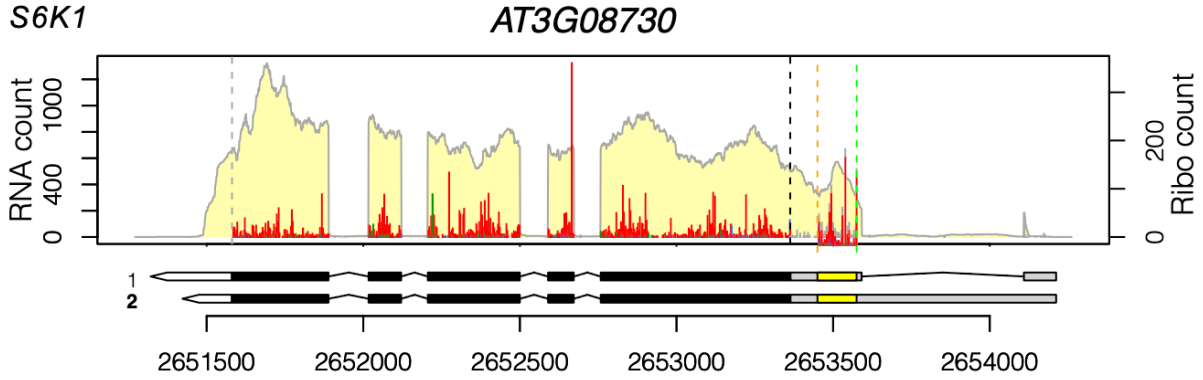




J. *PIF7*



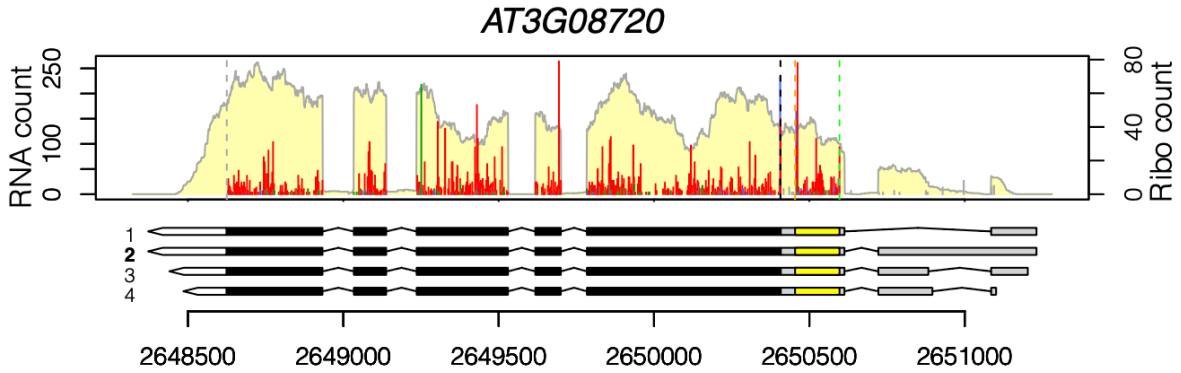
K. *S6K1*



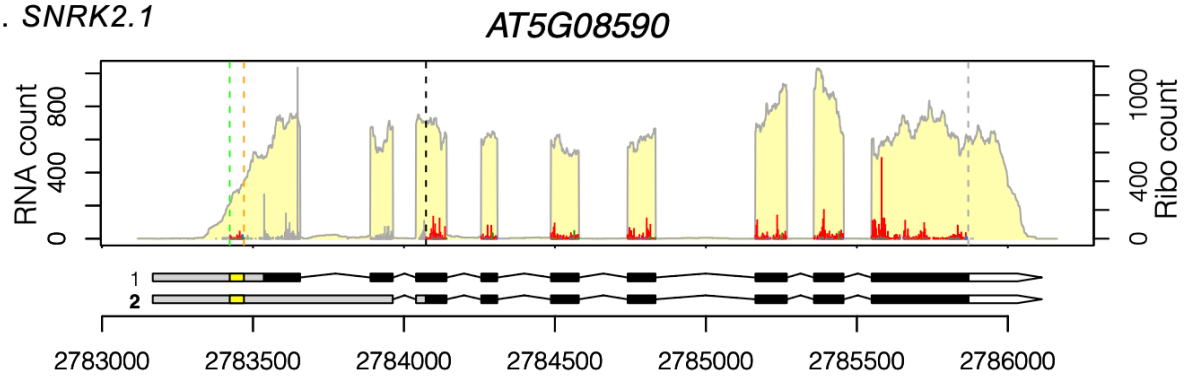
1032

1033

L. *S6K2*



M. *SNRK2.1*



N. *SNRK2.5*

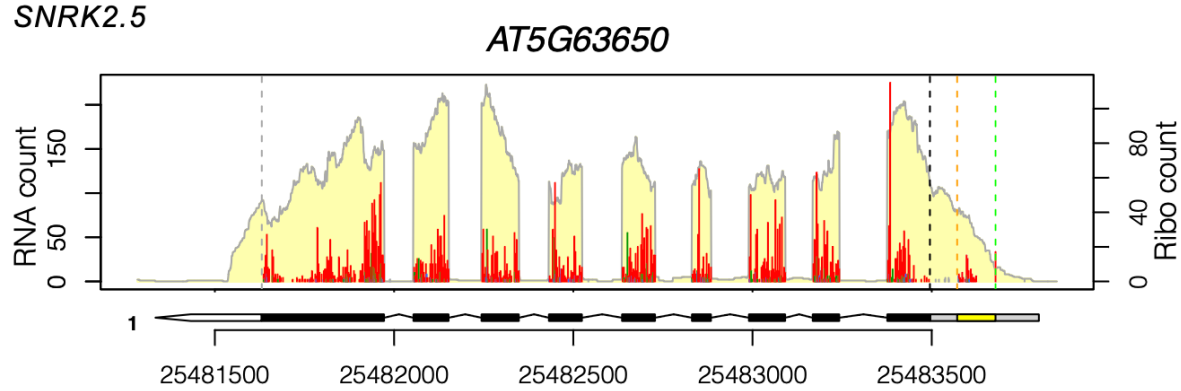
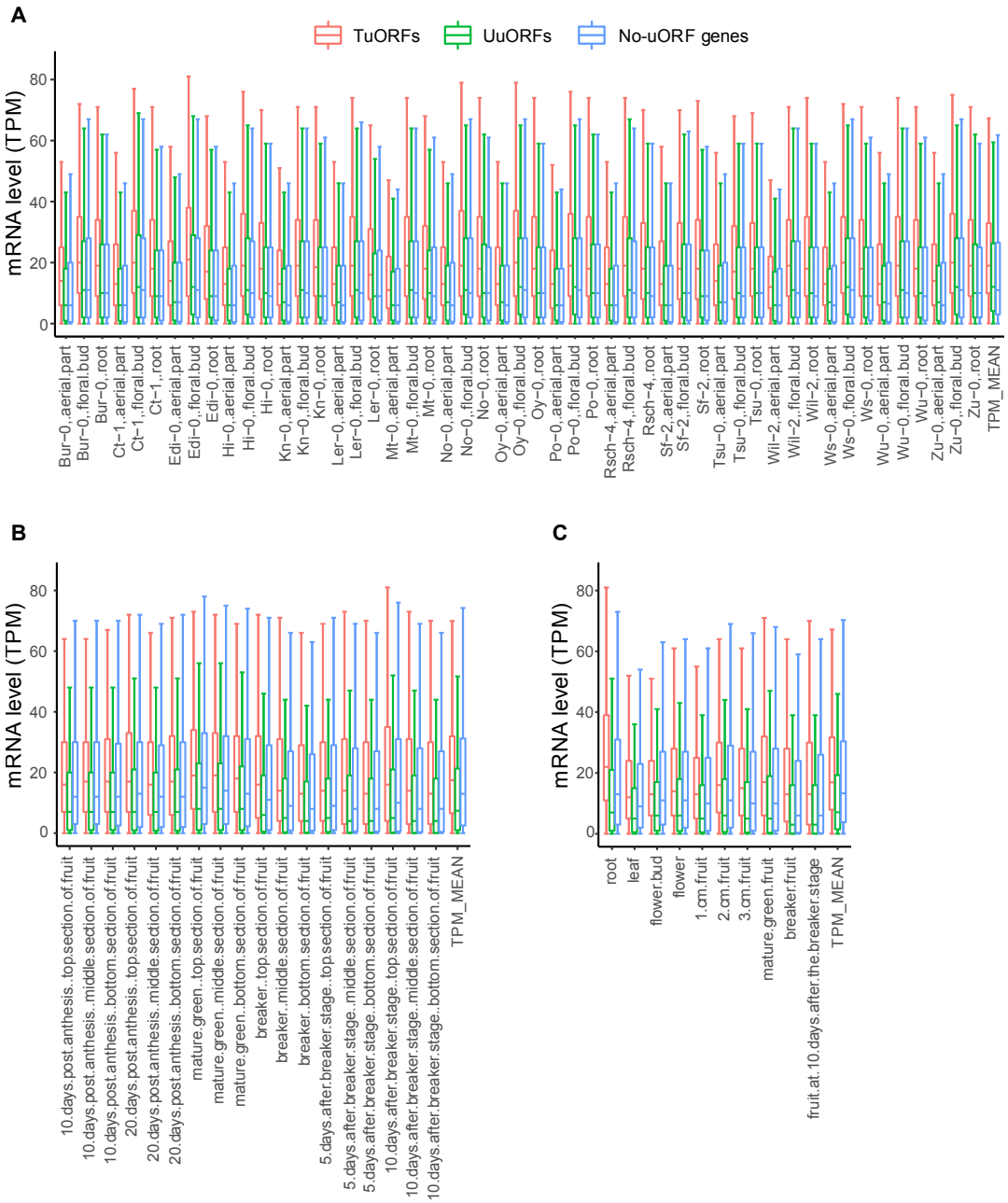


Figure S2. Examples of TuORFs in important regulatory genes. Data are presented as described in the Figure 1F legend. Isoform numbers are shown to the left of the gene models, and the isoform being considered is bolded. For genes that have multiple TuORFs identified (B, E, and I), the 3-nt periodicity of each TuORF is presented in a separate panel.



1041

1042 **Figure S3. mRNA levels of uORF-containing genes in different growth stages and tissues**

1043 **of various Arabidopsis ecotypes and tomato.**

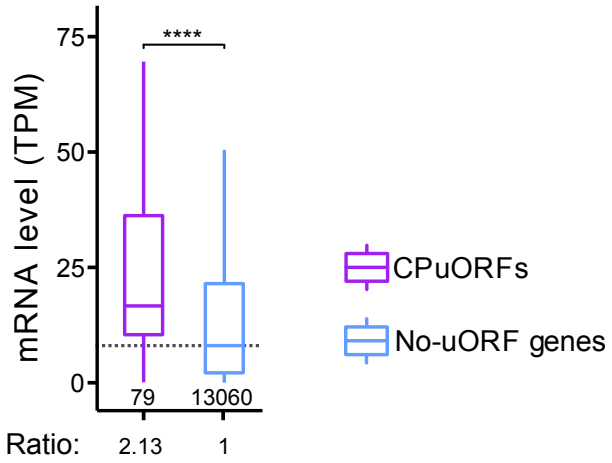
1044 (A) Arabidopsis roots and aerial and floral parts from different ecotypes.

1045 (B) Various developmental stages of tomato fruit (Heinz 1706 cultivar).

1046 (C) Roots, shoots, flowers, and fruits of tomato (Heinz 1706 cultivar).

1047 The TPM values from RNA-seq quantification were extracted from the EMBO-EBI expression

1048 atlas (Papatheodorou et al. 2020) (see MATERIALS AND METHODS).



1049

1050 **Figure S4. mRNA levels of CPuORF-containing genes compared with no-uORF genes.**

1051 Within the boxplot, the dashed line marks the median level of no-uORF genes; the number of

1052 genes in each category is listed below the lower whisker. The ratios underneath the plots

1053 indicate the median of each group normalized to that of no-uORF genes. The statistical

1054 significance was determined by Wilcoxon signed-rank test. The adjusted p-value was

1055 determined using the Benjamini and Yekutieli procedure to control for the false discovery rate in

1056 multiple testing (* 0.05 > p > 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-4, **** 1e-4 > p > 0).

1057

1058

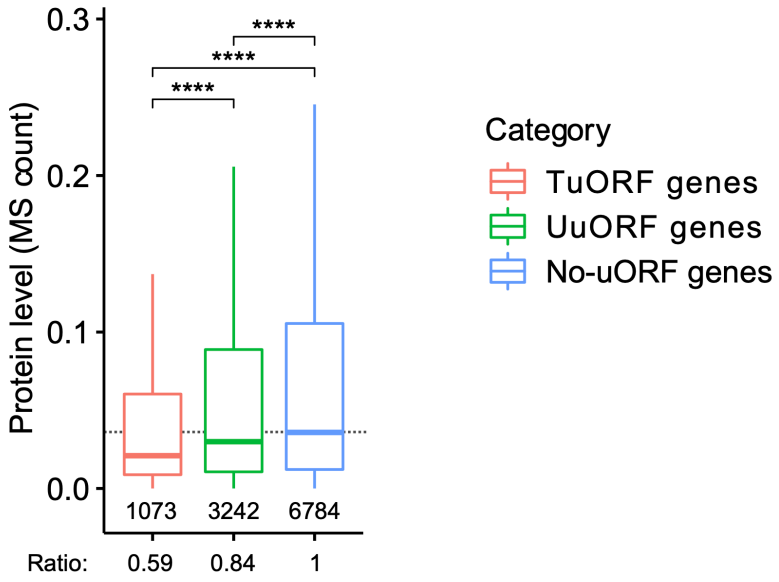
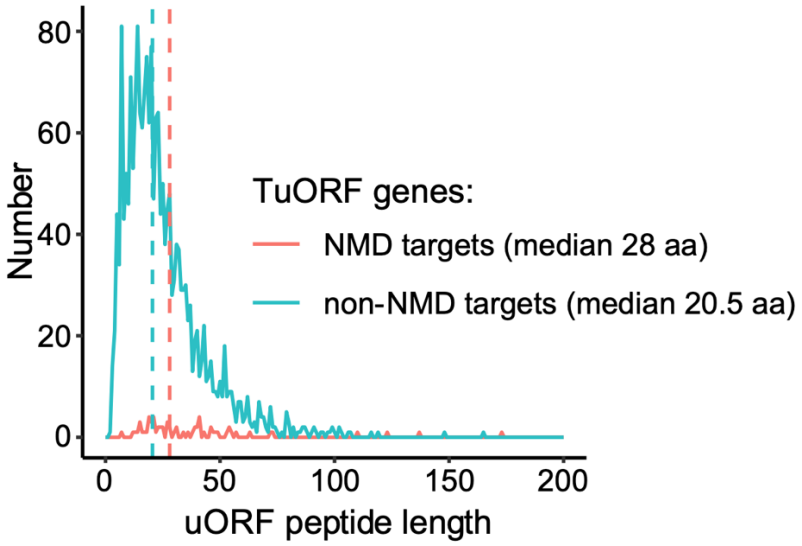


Figure S5. Protein abundance of the uORF-containing genes in Arabidopsis roots. The protein quantification data from Arabidopsis seedling roots were extracted from Song et al. (2018). Within the boxplot, the dashed line marks the median level of no-uORF genes; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of no-uORF genes. The statistical significance was determined by Wilcoxon signed-rank test. The adjusted p-value was determined using the Benjamini and Yekutieli procedure to control for the false discovery rate in multiple testing (* 0.05 > p > 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-4, **** 1e-4 > p > 0).



1068

1069 **Figure S6.** Distributions of uORF peptide lengths in NMD target and non-NMD target TuORF

1070 genes. Vertical dashed lines indicate the median value of each group.

1071

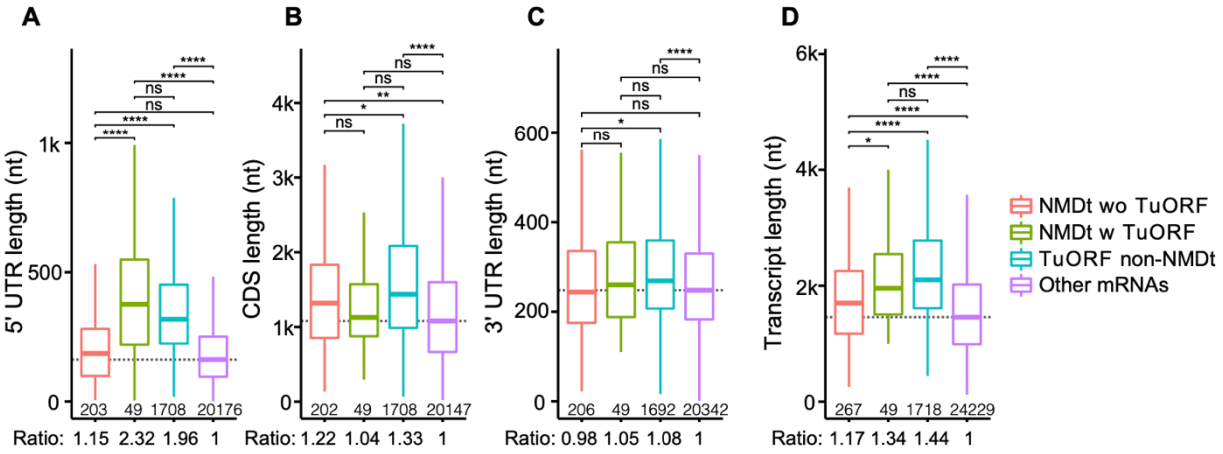


Figure S7. Lengths of UTRs and CDSs for NMD targets and TuORF genes.

Lengths of the 5' UTRs (A), CDSs (B), 3' UTRs (C), and entire transcripts (D) of NMD targets without a TuORF, NMD targets with a TuORF, non-NMD TuORF genes, and other genes. Within the boxplots, the dashed lines mark the median level of other mRNAs; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of other mRNAs. The statistical significance was determined by Wilcoxon signed-rank test. The adjusted p-value was determined using the Benjamini and Yekutieli procedure to control for the false discovery rate in multiple testing (* 0.05 > p > 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-4, **** 1e-4 > p > 0).

1083 **Table S3. Summary of datasets used in this study**
1084

Reference	Dataset	Sample	Genotype	Growth conditions
Current study	Current study RNA-seq Ribo-seq	7-day-old Arabidopsis seedlings	Col	16h-light/8h-dark, grown in 0.5x MS liquid media, collected between 4–5 h after light on
(Hsu et al., 2016)	Our previous data RNA-seq Ribo-seq	4-day-old Arabidopsis seedling shoots or roots	Col	16h-light/8h-dark, grown in 0.5x MS liquid media, collected at 3 h after light on
(Song et al., 2018)	Our previous quantitative proteomics data	4-day-old Arabidopsis seedling shoots or roots	Col	16h-light/8h-dark, grown in 0.5x MS liquid media, collected at 3 h after light on
(Szabo et al., 2020)	mRNA half-life (metabolic labeling)	5-day-old Arabidopsis seedlings	Col	Constant light, grown in 0.5x MS liquid media
(Wu et al., 2019)	Our previous tomato data RNA-seq Ribo-seq	4-day-old tomato seedling roots	Heinz 1706	16h-light/8h-dark, grown on 1x MS solid media, collected at 3 h after light on
(Drechsel et al., 2013)	NMD mutant RNA-seq 1	11-day-old Arabidopsis seedlings	Col, *upf1 (lba1)* *upf3-1*	Constant light, grown on 0.5x MS solid media
(Merchante et al., 2015)	NMD mutant RNA-seq 2	3-day-old Arabidopsis etiolated seedlings	Col, *upf2-10*	Darkness, grown on 1x MS solid media
(Raxwal et al., 2020)	NMD mutant RNA-seq 3	Arabidopsis 8-week-old leaves for *upf1 pad4*, 3–4-week-old leaves for *pad4*	*upf1-3 pad4*, *pad4*	16h-light/8h-dark, soil
(Gloggnitzer et al., 2014)	NMD mutant RNA-seq 4	Arabidopsis likely 4–5-week-old leaves	*smg7-1 pad4*, *pad4*	16h-light/8h-dark, soil
(Sorenson et al., 2018)	mRNA decay rate	5-day-old Arabidopsis seedlings	Wild type, *vcs*, *sov*, *vcs sov*	Constant light, grown on 0.5x MS solid media, treated with liquid solution

1085
1086