Nonsense-mediated decay is not a major mechanism for regulating the uORF-containing mRNAs in Arabidopsis

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 1-10

Supplementary Tables 1-4

ABSTRACT

Background:

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.

Results:

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, Arabidopsis TuORFs are associated with 38%, 14%, and 43% reductions in translation efficiency, mRNA half-lives, and protein levels, respectively. Surprisingly, we found TuORF-containing mRNAs in plants have 51% higher transcript levels, which is opposite to the observations in animals. The higher transcript levels of TuORF-containing mRNAs are persistent in diverse tissues and developmental stages across different plants. We present multiple lines of evidence supporting the hypothesis that uORF translation generally does not trigger NMD in Arabidopsis. First, TuORF-containing mRNAs are not increased 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.

Conclusions:

In summary, our work reveals that Arabidopsis TuORFs reduce mRNA stability and translation through mechanisms different from NMD. Our results highlight a fundamental difference in gene regulation mediated by TuORFs in plants and animals.

56 **INTRODUCTION**

57 Upstream open reading frames (uORFs) are cis-acting translational repressors that
58 regulate many eukaryotic protein-coding transcripts. Mutations or modifications of uORFs are
59 associated with human diseases and plant traits [1–4]. Considering AUG as a translation start
60 codon, 30%–70% of genes contain at least one potential uORF in moss, Arabidopsis, fly,
61 mouse, and human [5–7]. Genes with predicted uORFs are enriched for signal transduction
62 proteins, transcription factors, and membrane proteins in animals [7] and for transcription factors
63 and protein kinases in plants [5], underlining the potential importance of uORFs in modulating
64 the expression of vital cellular regulators.

65 Many studies in eukaryotes have reported that uORF translation reduces main ORF
66 (mORF) translation [6, 8–11], and it is estimated to reduce protein production by 30%–80% [1,
67 8]. According to the scanning model of mRNA translation, the 43S pre-initiation complex, which
68 contains the small ribosomal subunit, several translation initiation factors, and initiator tRNA,
69 binds the 5' cap and scans the 5' untranslated region (UTR) to search for an optimal initiation
70 sequence [12]. When a uORF is translated, the ribosome may stall during elongation or at
71 termination, or it may dissociate from the mRNA after termination, thus reducing the translation
72 of the downstream mORF [13]. The mORFs downstream of uORFs may get translated through
73 leaky scanning or reinitiation [14].

74 uORF translation is also expected to trigger mRNA degradation, presumably through
75 nonsense-mediated decay (NMD). NMD is an evolutionarily conserved regulatory mechanism to
76 eliminate aberrant mRNAs, such as those that contain premature termination codons and
77 unusually long 3' UTRs [15, 16]. Two pathways have been established in mammals: the exon
78 junction complex (EJC)-dependent and long 3' UTR pathways. Normally, EJC and two NMD
79 factors (UP-FRAMESHIFT SUPPRESSOR 2 and 3; UPF2 and UPF3) deposited during splicing
80 of mRNAs will be evicted by the ribosome in the first round of translation. In the presence of a
81 premature termination codon, EJC-UPF2-UPF3 complexes downstream of the termination

82 codon remain attached to the mRNA. This allows UPF1, which is part of the surveillance
83 complex recruited at translation termination, to interact with UPF2 and UPF3 in proximity. The
84 interaction of UPF1, UPF2, and UPF3 activates the SUPPRESSOR WITH MORPHOLOGICAL
85 EFFECT ON GENITALIA1 (SMG1) kinase, which phosphorylates UPF1. This establishes the
86 first phase of NMD, substrate recognition, and activates downstream signaling and eventually
87 the second phase of NMD, substrate degradation, which involves SMG5, SMG6, and SMG7.
88 Phosphorylated UPF1 also inhibits translation initiation [17]. Alternatively, unusually long 3'
89 UTRs, which result in poor translation termination, are believed to promote NMD through
90 phosphorylated UPF1 independent of EJC [18].

91 The plant NMD pathway is best characterized in *Arabidopsis thaliana*. Arabidopsis
92 possesses homologous NMD factors involved in the first phase, including UPF1, UPF2, and
93 UPF3, and in the second phase, SMG7 [19]. However, Arabidopsis is missing the homologs of
94 animal SMG1, SMG5, and SMG6 [20]. The absence of key NMD components in plants
95 suggests significant variations in the NMD mechanisms between plants and animals. It remains
96 an open question how the differences in NMD components affect substrate recognition and
97 degradation in plants.

98 In Arabidopsis, putative NMD targets have been identified based on transcripts that
99 overly accumulate in NMD-defective mutants [21–24]. However, there is a concern that the
100 upregulation of some transcripts in NMD mutants might be caused by secondary effects of NMD
101 [25]. For example, NMD mutants affect numerous aspects of plant physiology, and some alleles
102 cause severe growth retardation and lethality due to constitutive immune responses [26–32].
103 Moreover, UPF1 has been reported to have multiple functions in plants, such as alternative
104 splicing regulation and translational control, in addition to NMD [25]. To identify *bona fide* NMD
105 targets, *upf1* and *smg7*, which are defective in the first and second phases of NMD,
106 respectively, were introduced into *pad4*, in which the immune response is inactivated [25, 28].
107 The 333 transcripts upregulated in both *upf1 pad4* and *smg7 pad4* were defined as high-

108 confidence NMD targets. It was shown these NMD targets are translationally regulated by UPF1
109 and that their turnover requires decapping and 5' to 3' decay [25].
110 As uORF translation could result in a transcript with an unusually long 3' UTR, uORF-
111 containing mRNAs are predicted to be targeted by NMD [16, 33]. Consistent with this
112 expectation, genome-wide studies in zebrafish, mouse, and human have found that transcripts
113 with translated uORFs have lower mRNA levels globally [6, 8]. Moreover, transcripts containing
114 translated uORFs are upregulated in knockdown of NMD factors in mouse embryonic stem cells
115 and human cell lines [34, 35], supporting the idea that these transcripts are NMD targets.
116 In plants, whether uORF translation triggers NMD is still under debate. Two genome-
117 wide studies suggest that uORF-containing mRNAs are targets of NMD, as transcripts
118 possessing predicted uORFs are enriched in the genes upregulated in Arabidopsis *upf1-1* and
119 *upf3-1* mutants [21, 23]. In contrast, three studies suggested that NMD only controls a small
120 fraction of uORF-containing mRNAs in plants, such as those containing uORFs with > 35
121 codons (~2% of uORF-containing genes), or those responsive to certain environmental stimuli,
122 such as polyamine or blue light [36–38]. Moreover, evidence from several gene-specific or
123 genome-wide studies suggests that uORFs do not trigger NMD. First, the mRNA levels of
124 several uORF-containing transcripts, including *NIP5;1* and *AtMHX*, are not increased in NMD
125 mutants [39, 40]. Second, a genome-wide study reported that genes with increasing numbers of
126 uORFs did not show lower transcript abundances in Arabidopsis and tomato [41]. Third, genes
127 with predicted uORFs are not over-represented in the consensus NMD target list identified from
128 three NMD mutants, including *upf1-5*, *upf3-1*, and *smg7-1* [27]. It was also reported that 47
129 mRNAs that contain Conserved Peptide uORFs (CPuORFs) behave differently from NMD
130 targets prior to their degradation mediated by XRN4 [42]. It is unknown whether some uORF-
131 containing mRNAs escape NMD or whether uORF-containing mRNAs are generally not targets
132 of NMD in plants. It has been concluded that the mRNA features studied so far are insufficient

133 to infer NMD targetability in plants [43]. Overall, it remains unclear how translated uORFs affect
134 mRNA stability and what the role of NMD in this process is in plants.

135 One limitation of the previous studies is that predicted uORFs, which include both
136 translated and untranslated uORFs, were commonly included in the analyses. Although many
137 uORFs have been predicted based on nucleotide sequences [7], determining which uORFs are
138 actively translated still requires experimental evidence. Recently, deep sequencing of ribosome
139 footprints – known as ribosome profiling or Ribo-seq – has been exploited to study genome-
140 wide mRNA translation [44, 45]. Several Ribo-seq studies in Arabidopsis have deployed
141 different strategies to identify translated uORFs (TuORFs) [9, 46–50]. Among the parameters
142 used, 3-nucleotide (3-nt) periodicity, which corresponds to translating ribosomes deciphering 3
143 nts per codon, is considered the most stringent criterion to identify high-confidence translation
144 events [44, 51–53]. Our previous work and a recent study using RiboTaper [51], which
145 computes the statistical significance of 3-nt periodicity, uncovered 187–1378 AUG-initiated
146 TuORFs in Arabidopsis [47, 50].

147 Here we further improved the Ribo-seq methodology and identified 2093 AUG-initiated
148 TuORFs in Arabidopsis seedlings. We systematically examined the assumed roles of TuORFs
149 in mORF translation and mRNA decay by integrating multiple genome-wide datasets. As
150 predicted, we found that TuORF mRNAs are associated with lower translation efficiency, shorter
151 mRNA half-lives, and lower protein abundance. Surprisingly, we found that, in contrast to the
152 animal paradigm, plant TuORF mRNAs have higher steady-state mRNA levels, and they are not
153 increased in the NMD mutants. Examinations of numerous parameters revealed that uORFs
154 and NMD regulate gene expression through separate mechanisms. Taken together, our results
155 reveal unexpected features and regulation of uORFs in Arabidopsis and suggest that plants and
156 animals have evolved different uORF-dependent mechanisms for gene regulation.
157
158

159 **RESULTS**

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

161 Previously we optimized Ribo-seq to reach high 3-nt periodicity, and we exploited this
162 unique feature of mRNA translation to identify translated ORFs in Arabidopsis [47]. Although the
163 data revealed many unannotated translation events, the number of TuORFs uncovered was
164 low, presumably due to insufficient Ribo-seq coverage within uORFs. We proposed that
165 increasing Ribo-seq coverage could improve the efficiency of translated ORF identification,
166 especially for relatively short ORFs, such as uORFs. Here we optimized the coverage of Ribo-
167 seq in 7-day-old Arabidopsis seedlings by minimizing the potential loss of ribosome footprints
168 during the purification process (see METHODS for details).

169 In addition to high correlations among biological replicates (**Figure S1A**), our new Ribo-
170 seq data displayed characteristics expected for high-quality data, including strong 3-nt
171 periodicity (92% in-frame reads), high enrichment for coding sequences (CDSs), and
172 characteristic ribosome footprint lengths (**Figure 1A-C**). The improved coverage is evident when
173 examining the number of unique ribosome footprints detected globally. At a sequencing depth of
174 20 million reads from randomly selected reads, the number of unique ribosome footprints in our
175 new data is 4.01-fold higher than that in our previous dataset (**Figure 1D**). The coverage
176 improvement is also evident in individual transcript profiles within both uORFs and mORFs. For
177 example, a TuORF in *REPRESSOR OF GA1* (*RGA1*) was identified in our current data but
178 missed in our previous study despite the mRNA levels being similar (**Figure S1B**).

179 The high-coverage data increased the number of TuORFs uncovered. In total, 2093
180 TuORFs were identified in 1768 genes based on significant 3-nt periodicity using RiboTaper and
181 Araport11 annotation (**Figure 1E-G** and **Table S1A**). This result is in strong contrast to the
182 relatively few TuORFs identified in Arabidopsis seedling shoots (93 TuORFs) and roots (138
183 TuORFs) from our previous data obtained using plants grown under similar conditions and using
184 the same statistical analysis and annotations (**Figure 1E** and **Table S1B-C**). The new data also

185 revealed translation from a higher number of Conserved Peptide uORFs (CPuORFs) and
186 annotated coding sequences (mORFs of protein-coding genes) (**Figure 1E** and **Table S1D-F**).
187 We manually plotted all TuORF genes to confirm strong 3-nt periodicity present in the expected
188 uORF regions using RiboPlotR [54] (examples shown in **Figure 1F-G, Figure S2**). Taken
189 together, these results show that our enhanced-coverage Ribo-seq data dramatically improved
190 the identification of TuORFs.

191 Previous analyses of predicted uORFs found that transcription factors and protein
192 kinases are overrepresented among uORF-containing genes [5, 55]. Consistent with these
193 analyses, functions related to transcription factors and protein kinases/phosphatases were
194 identified as being enriched in our Gene Ontology (GO) analysis of TuORF-containing genes
195 (**Table S2**). We found that many TuORF-containing genes encode extensively studied
196 regulators of plant growth and environmental responses. For example, this list included well-
197 known transcription factors, such as those involved in hormone signaling (RGA1 and AUXIN
198 RESPONSE FACTOR2/3/8), root development (SCARECROW), the circadian clock (TIMING
199 OF CAB EXPRESSION1), and light signaling (PHYTOCHROME INTERACTING FACTOR3/4/7)
200 [56–63]. The list also included well-characterized protein kinases, such as Phytochromes B and
201 E, ribosomal-protein S6 kinases (S6K1/2), and SNF1-related protein kinase 2 (SNRK2.1/2.5)
202 (**Figure S1B, S2A-N**) [64–69]. These results support the notion that TuORFs control important
203 cellular regulators in plants.
204

205 **TuORF-containing genes have higher mRNA abundance but are associated with lower**
206 **translation efficiency, shorter mRNA half-lives, and lower protein abundance**

207 To understand how uORF translation affects gene expression, we compared mRNA
208 abundance and translation efficiency of TuORF-containing genes (hereafter, TuORF genes),
209 genes containing sequence-predicted uORFs but undetected by RiboTaper (hereafter, UuORF
210 genes), and genes containing an annotated 5' UTR but no potential uORFs (hereafter, no-uORF

211 genes). Note that some UuORFs may be translated but missed in our analysis because their 3-
212 nt periodicity is insufficient to pass the statistical test: these include extremely short uORFs,
213 those translated at a low frequency, and those overlapping with other uORFs, thus disrupting
214 the 3-nt periodicity. To study the global effects of uORFs on mRNA stability and protein
215 abundance, we also compared genes with TuORFs, UuORFs, or no-uORFs, using available
216 mRNA half-life data determined by metabolic labeling in Arabidopsis seedlings and two
217 quantitative proteomics datasets from shoots and roots of Arabidopsis seedlings [70, 71]. The
218 growth conditions and source of the data are summarized in **Table S3**. We presented the data
219 in box plots to show the statistical results and in cumulative plots to show the trends of different
220 gene groups (**Figure 2A-D, Figure S3, S4**).

221 Studies in vertebrates have shown that uORF translation is associated with lower
222 steady-state mRNA abundance, consistent with the expectation that TuORF mRNAs are
223 regulated by NMD [6, 8]. Surprisingly, we found that TuORF genes in Arabidopsis have higher
224 steady-state mRNA levels than those of no-uORF genes (**Figure 2A, S3A**), despite having
225 shorter mRNA half-lives as expected (**Figure 2C, S3C**). The median mRNA level of TuORF
226 genes is 51% higher than that of no-uORF genes. The higher steady-state mRNA levels and
227 shorter mRNA half-lives together suggest that TuORF genes are highly transcribed. To examine
228 whether the high transcript levels of TuORF genes are conserved in other plant species, we
229 reanalyzed our previous tomato Ribo-seq data from seedling roots [72] and found the TuORF
230 genes in tomato also have higher steady-state mRNA levels relative to no-uORF genes (**Figure**
231 **2E**). Given that Arabidopsis and tomato diverged more than 100 million years ago [73], our
232 results suggest that TuORF-associated higher mRNA abundance could be prevalent in plants.
233 To investigate whether our observation is specific to seedlings or common to different
234 developmental stages/tissues/genetic backgrounds in Arabidopsis and tomato, we examined
235 the mRNA levels of the uORF-containing genes in 128 independent RNA-seq studies collected
236 by the EMBO-EBI expression atlas [74]. Consistent with our Arabidopsis seedlings and Tomato

237 seedling root data, we found the TuORF genes are expressed at higher levels in diverse
238 tissues, developmental stages, and ecotypes in Arabidopsis and tomato (**Figure 2G and S5A-**
239 **C**), supporting the hypothesis that TuORF genes are preferentially expressed at high levels in
240 these plants under a wide range of sampling conditions.

241 CPuORFs are uORFs that encode evolutionarily conserved peptides, and they were
242 identified independently from our approach by evolutionary conservation [75–79]. Many of the
243 CPuORFs were also identified as TuORFs in our study (**Figure 1E, 1G**). We found that the
244 median mRNA abundance of CPuORF genes is 2.2-fold higher than that of no-uORF genes
245 (**Figure S6**), consistent with our observations that genes with TuORFs are expressed at higher
246 levels. Our observation of the high mRNA abundance of plant TuORF genes reveals that plants
247 and animals implement TuORFs differently in gene regulation.

248 We cannot rule out the possibility that TuORFs are more likely to be identified in highly
249 expressed transcripts, as TuORF translation levels are positively correlated with transcript
250 abundance (r = 0.38, **Figure S7A**). However, high mRNA abundance alone is insufficient to
251 determine whether a potential uORF is translated, as many UuORF genes are highly expressed
252 as well and the distributions of TuORF and UuORF mRNA levels largely overlap (**Figure S7B**).
253 The fact that the TuORFs identified are functionally associated with lower translation efficiency,
254 shorter mRNA half-lives, and less protein accumulation than UuORF genes (**Figure 2B-D, F,**
255 **S4**) support the idea that TuORFs are selectively translated to repress the translation and
256 mRNA stability.

257 Interestingly, even though globally TuORF genes in Arabidopsis start with 51% more
258 mRNAs than no-uORF genes, they produce 41%–43% less protein, on average (**Figure 2A, 2D,**
259 **S3, S4**), likely due to the combined effects of a 38% reduction in translation efficiency and a
260 14% reduction in mRNA half-lives (**Figure 2B, 2C**). These observations suggest that the global
261 repression of TuORF gene protein synthesis is determined by multilayer post-transcriptional
262 regulation. Notably, for the four parameters examined here, UuORF genes consistently have

263 levels between those of TuORF genes and no-uORF genes: they show a 4% increase in mRNA
264 levels and 11%, 6%, and 21% reductions in translation efficiency, mRNA half-lives, and protein
265 abundance, respectively (**Figure 2A-D, S3A-D**). These results indicate that UuORFs have
266 smaller effects on gene expression than TuORFs and highlight the importance of distinguishing
267 TuORFs and UuORFs in genome-wide analysis.
268
269 **TuORF genes are larger and likely evolutionarily older than other genes**
270 We next investigated whether there are general structural differences among TuORF,
271 UuORF, and no-uORF mRNAs. Considering the directionality of ribosome scanning, we
272 hypothesized TuORFs might be associated with longer 5' UTRs. Indeed, we found that the
273 median 5' UTR length of TuORF mRNAs is 2.5-fold of that of no-uORF mRNAs (**Figure 3A**),
274 which is consistent with a previous report that uORFs in longer 5' UTRs are more likely to be
275 translated [9]. However, unexpectedly, we found that TuORF mRNAs also have longer CDSs
276 and 3'UTRs, as well as longer transcript lengths, than UuORF mRNAs (**Figure 3B-D**). No-uORF
277 mRNAs are the shortest in all categories examined (**Figure 3A-D**). These results support the
278 hypothesis that TuORFs are more likely to be found in longer 5' UTRs and reveal that TuORF
279 mRNAs tend to encode larger proteins (**Figure 3B**). The enrichment of TuORF genes for
280 transcription factors, protein kinases, and related functions in GO term analysis (**Table S2**) also
281 implies TuORF genes tend to encode proteins with conserved domains. It is known that
282 evolutionarily older genes are generally longer, expressed at higher levels, and have protein
283 sequences that are under stronger purifying selection [80, 81]. Our observations that TuORF
284 genes possess many features of evolutionarily older genes imply that older genes are more
285 likely to be regulated through mRNA translation and degradation mediated by TuORFs.
286
287 **TuORF mRNAs are not upregulated in the NMD mutants**

288 The established paradigm in vertebrates is that uORF translation triggers NMD and thus
289 results in low steady-state mRNA levels [6, 8, 34, 35]. Despite having higher steady-state
290 mRNA levels (**Figure 2A**), TuORF genes in Arabidopsis also have shorter mRNA half-lives
291 (**Figure 2C**). Therefore, we investigated whether TuORF-containing transcripts in Arabidopsis
292 are subject to NMD.

293 If TuORFs trigger NMD, one would expect TuORF mRNAs to be among the known NMD
294 targets and their mRNA levels to be increased in NMD-deficient mutants. We first examined
295 whether TuORF mRNAs were reported in the high-confidence NMD target list [25]. Among the
296 333 NMD targets that were upregulated in both *upf1 pad4* and *smg7 pad4* adult leaves [25], 248
297 genes with an annotated 5' UTR were expressed in our data (**Table S4A**). However, only 49 of
298 1748 TuORF genes were identified in the high-confidence NMD target list (**Table S4B**). The 49
299 genes represent 2.8% of TuORF-containing genes and 20% of the NMD targets. These results
300 suggest that TuORFs generally do not trigger NMD; rather, NMD targets are more likely to be
301 regulated by TuORFs than other genes (Chi-square test with Yates' correction: p-value =
302 7.555e-10). One caveat is that the poor overlap between the two gene lists could be due to
303 differences in growth conditions, developmental stages, and sample handling in different labs
304 (**Table S3**).

305 To further investigate if TuORF mRNAs are NMD targets, we examined whether they are
306 upregulated in NMD-deficient mutants by analyzing four published RNA-seq datasets for NMD
307 mutants [23, 25, 30, 82]. Although these datasets were collected from different NMD mutants
308 and at various developmental stages (**Table S3**), the high-confidence NMD targets are
309 consistently and significantly upregulated in all the NMD mutants examined compared with the
310 control plants (i.e., 3.05-, 1.54-, 4.98-, and 2.65-fold for NMD targets without TuORFs; 3.04-,
311 2.23-, 5.04-, and 2.66-fold for NMD targets with TuORFs)(**Figure 4A-D**). In contrast, TuORF
312 mRNA levels show no or little change in the NMD mutants (i.e., 1.01-, 0.98-, 1.10- and 1.10-
313 fold). Thus, only the NMD targets, but not TuORF mRNAs, are dramatically increased in the

314 NMD mutants, supporting the idea that uORF translation in Arabidopsis generally does not
315 trigger NMD.
316 Given different uORFs are translated at various levels (See **Figure 1F-G, S1B, S2A-N**
317 for examples), it is possible only the highly repressive TuORFs efficiently trigger NMD. To test
318 this hypothesis, we split the TuORF genes into two groups based on their uORF relative
319 translation efficiency (RTE) (i.e., TE of an uORF is normalized to TE of its mORF). However,
320 even the higher TuORF RTE group shows no or little effects on the mRNA accumulation in the
321 NMD mutants (i.e., 1.13-, 0.99-, 1.15- and 1.11-fold)(**Figure 4E-H**). We also tested whether
322 having more TuORFs in a transcript efficiently triggers NMD. We split the TuORF genes into two
323 groups based on the number of TuORFs detected. We found that transcripts with 2~5 TuORFs
324 do not lead to significant mRNA increase in the NMD mutants, either (i.e., 0.93-, 0.98-, 0.99-
325 and 1.09-fold)(**Figure 4I-L**). Together, these results further support the idea that uORF
326 translation in Arabidopsis generally does not induce NMD.
327
328 **Lengths of the TuORF and 3' UTR alone are insufficient to predict NMD targetability**
329 A previous study suggested that uORFs that encode a peptide longer than 35 amino
330 acids are likely to trigger NMD in Arabidopsis [36]. To test whether TuORF length is associated
331 with NMD targetability, we analyzed TuORF lengths in the NMD targets and non-NMD targets.
332 We found that the NMD target TuORF genes tend to have longer uORF peptides than non-NMD
333 target TuORF genes (median length 28 amino acids vs. 20.5 amino acids, $p = 9.874e-07$,
334 Wilcoxon signed-rank test). However, the distributions of the uORF peptide lengths in NMD
335 targets and non-NMD targets largely overlap (**Figure S8**), indicating that TuORF length alone is
336 insufficient to predict whether a TuORF triggers NMD of a transcript.
337 We also examined whether the high-confidence NMD targets have longer 3' UTRs, a
338 feature that potentially triggers NMD. Interestingly, the NMD targets do not have significantly
339 longer 3' UTRs, regardless of whether they contain TuORFs (**Figure S9**). Taken together, our

340 results are consistent with a recent suggestion that no definite RNA features studied to date are
341 sufficient to predict NMD targets in plants [43].
342
343 **TuORF mRNAs and NMD targets have distinct expression patterns**
344 To further characterize the roles of NMD and TuORFs in gene expression regulation, we
345 compared the expression patterns of the NMD targets and TuORF mRNAs (**Figure 5**).
346 Consistent with the expectation that NMD targets are being actively degraded, we found that
347 NMD targets have lower mRNA abundance than other genes; this is in strong contrast to
348 TuORF mRNAs, which have higher mRNA abundance than other genes (**Figure 5A, 5E**). While
349 both NMD targets and TuORF genes are associated with lower translation efficiency, shorter
350 mRNA half-lives, and lower protein abundance than other genes (**Figure 5B-D, 5F-H**), the NMD
351 targets are associated with even lower protein levels than TuORF genes (**Figure 5D, 5H**).
352 If TuORFs effectively trigger NMD, one would expect TuORF-containing genes to have
353 translation efficiency similar to TuORF-containing NMD targets. While both NMD targets and
354 TuORF mRNAs have lower translation efficiency (**Figure 5B, 5F**), TuORF-containing NMD
355 targets have significantly lower translation efficiency than either NMD targets without TuORFs or
356 TuORF-containing non-NMD target genes (**Figure 5B, 5F**). This additive repression of
357 translation efficiency in TuORF-containing NMD targets suggests that NMD and TuORFs
358 suppress translation through separate mechanisms and supports that TuORFs do not effectively
359 trigger NMD. Our observation is consistent with the expectation that TuORFs prevent the
360 scanning pre-initiation complex from reaching the mORF while NMD suppresses translation
361 through UPF1 and/or other translational repressors [17, 25, 83].
362 Together, the differences in the expression patterns between TuORF mRNAs and NMD
363 targets further support the idea that TuORFs are not an mRNA feature that efficiently triggers
364 NMD in Arabidopsis.
365

366 **TuORF mRNAs and NMD targets are degraded through different mRNA decay**
367 **mechanisms**

368 mRNA decay can occur in the 5' to 3' or 3' to 5' direction, which is carried out by different
369 machinery [84]. We next investigated the mechanisms underlying the decay of TuORF mRNAs
370 and NMD targets. A genome-wide survey of mRNA decay rates in wild type and mutants
371 defective in either 5' to 3' decay or 3' to 5' decay, or both, has been conducted in Arabidopsis
372 [85]. In that survey, the *vcs-7* mutant, a null mutant of *VARICOSE*, which encodes a scaffold
373 protein for the mRNA decapping complex, was used to examine the 5' to 3' mRNA decay; the
374 *sov* mutant, defective in *SUPPRESSOR OF VARICOSE*, which encodes a ribonuclease II, was
375 used to examine 3' to 5' decay; and the *vcs/sov* double mutant was used to examine decay
376 affected in both directions [86–89]. We exploited this published dataset [85] to study the genetic
377 requirements for the mRNA decay of TuORF mRNAs, NMD targets, TuORF-containing NMD
378 targets, and other mRNAs.

379 We first compared the decay rates of each gene group across the four genetic
380 backgrounds (**Figure 6**). Similar to mRNAs that are non-NMD targets and lack TuORFs (a.k.a.
381 other mRNAs, **Figure 6D, 6H**), TuORF mRNAs have dramatically reduced decay rates in the
382 *vcs* mutant compared to the wild type (**Figure 6A, 6E**), indicating that decapping and 5' to 3'
383 decay are critical for degrading these transcripts. In the *vcs/sov* double mutant, TuORF mRNAs
384 have slightly but significantly lower decay rates than in the *vcs* mutant (**Figure 6A and 6E**),
385 suggesting *SOV* also plays a minor role in degrading TuORF mRNAs.

386 In contrast, the mRNA decay rates of NMD targets without TuORFs are only significantly
387 lower in the *vcs/sov* double mutant, suggesting that NMD targets are efficiently degraded by
388 both the 5' to 3' and 3' to 5' pathways (**Figure 6B, 6F**). In other words, in the absence of either
389 *SOV* or *VCS*, the remaining pathway could completely take over the decay of NMD targets
390 (**Figure 6B, 6F**). Although there are fewer NMD targets containing TuORFs, which resulted in
391 small sample size and statistical insignificance (**Figure 6C, 6G**), the decay rates in this group of

392 genes in the four genetic backgrounds show trends similar to those of NMD targets without
393 TuORFs (**Figure 6B, 6F**).
394 Next, we compared the decay rates of the four gene groups within each genetic
395 background (**Figure S10)**. In all four genetic backgrounds, NMD targets have higher decay
396 rates than TuORF mRNAs and other genes (**Figure S10A-H**). This pattern is still persistent in
397 the *vcs/sov* double mutant **(Figure S10D, S10H)**, suggesting additional decay pathway(s) is
398 involved in degrading the NMD targets in the absence of both VCS and SOV.
399 Taken together, our results suggest that TuORF mRNAs and NMD targets are
400 preferentially degraded through different decay pathways. NMD targets consistently have higher
401 decay rates than TuORF mRNAs and appear to be regulated by additional pathway(s) other
402 than *VCS* and *SOV* (see the model in **Figure 6I**).
403
404 **DISCUSSION**
405 In this study, we identified Arabidopsis TuORFs using improved Ribo-seq data and
406 systematically examined their presumed roles in gene expression. While our results support the
407 expected roles of uORFs in repressing mORF translation and mRNA stability, our study reveals
408 distinct features and regulation of uORFs in plants compared with animals.
409 Our observations that plant TuORF genes have higher expression levels (**Figure 2**),
410 encode larger proteins (**Figure 3**), and are enriched for evolutionarily conserved domains
411 (**Table S2**) raised an interesting question as to whether older genes are more likely to be
412 controlled at the protein levels through mRNA degradation and translation via uORFs. Higher
413 mRNA abundance may allow TuORF mRNAs to be highly translated under certain conditions
414 that require a high level of specific proteins. For example, it has been suggested that
415 derepression of downstream translation is a general mechanism of uORF-mediated expression
416 under stress [90]. Thus, TuORFs could enable fast and dynamic regulation of mORF translation.
417 Future work monitoring the translation of TuORFs and their mORFs under different conditions

418 will help decipher the regulation and physiological roles of specific TuORFs.

419 If TuORFs trigger NMD in Arabidopsis, we would expect that TuORF mRNAs behave
420 similarly to NMD targets. In this report, we presented six observations suggesting that
421 Arabidopsis TuORF mRNAs and NMD targets are two distinct groups of transcripts: (1) NMD
422 targets, but not TuORF mRNAs, are upregulated in various NMD mutants (**Figure 4**). (2)
423 TuORF mRNAs have significantly higher transcript levels, while the NMD targets have lower
424 transcript levels than other genes (**Figure 5A**). (3) NMD and TuORFs additively repress
425 translation, suggesting they repress translation through separate mechanisms (**Figure 5B, 5F**).
426 (4) While NMD targets are efficiently degraded through both 5' to 3' and 3' to 5' decay, TuORF
427 mRNAs, like other genes, mainly rely on 5' to 3' decay (**Figure 6**). (5) The decay rates of NMD
428 targets are consistently higher than those of TuORF mRNAs (**Figure S10)**. (6) NMD targets also
429 appear to be degraded by additional pathway(s) in the absence of both *VCS* and *SOV* (**Figure**
430 **S10H**). These differences between TuORF mRNAs and NMD targets suggest that NMD is not a
431 major mechanism regulating TuORF mRNAs.

432 The fact that Arabidopsis is missing key NMD factors found in animals, including SMG1,
433 SMG5, and SMG6, suggests significant variations in the NMD mechanisms between plants and
434 animals. In particular, SMG1 is the kinase responsible for UPF1 phosphorylation required for
435 substrate recognition in animal NMD [15, 16]. The lack of SMG1 and the use of an alternative
436 kinase for Arabidopsis UPF1 phosphorylation [91] may lead to a difference in substrate
437 recognition: while TuORF mRNAs are recognized as NMD targets in animals, plant TuORF
438 mRNAs are not.

439 Despite the difference in NMD components in plants and animals, our findings that
440 Arabidopsis NMD targets are degraded through both 5' to 3' decay and 3' to 5' decay is
441 consistent with the two-directional decay of NMD targets in animals. Note that plant SOV is the
442 homolog of DIS3L2, the exonuclease responsible for 3' to 5' decay of NMD targets in animals
443 [92, 93]. Two-directional decay could be advantageous to quickly remove aberrant mRNAs from

444 the cytoplasm to prevent further translation and facilitate the recycling of NMD machinery.

445 Genes controlled by both NMD and TuORFs might encode critical regulators, and their
446 protein levels require precise regulation. These genes include key enzymes in gibberellin
447 hormone and polyamine biosynthesis and many transcription factors (**Figure 1F-G, and Table**
448 **S4B**). For example, one of the transcription factors, *SUPPRESSOR OF ACAULIS 51* (*SAC51*),
449 was discovered in a genetic screen for suppressors of the dwarf phenotype of the *acaulis5*
450 mutant [94]. *SAC51* contains a CPuORF that is controlled by the thermospermine [95]. When
451 the CPuORF was mutated, it increased the transcript level of *SAC51* [94]. *SAC51* encodes a
452 bHLH transcription factor that has been proposed to activate the genes required for stem
453 elongation.

454 It is well established that there is an important interplay between translation efficiency
455 and mRNA stability. Two mechanisms have been proposed to explain this phenomenon: the
456 "stalled ribosome-triggered decay model," which predicts that ribosome stalling or slowing
457 promotes mRNA degradation, and the "translation factor-protected model," which predicts that
458 active translation promotes mRNA stability because the translation initiation factors protect the
459 mRNA from the decapping complex [96–101]. Consistent with the first model, several CPuORF-
460 containing mRNAs in Arabidopsis were reported to have strong ribosome stalling patterns, and
461 the endonucleolytic cleavages upstream of the stalled ribosomes were observed [102, 103].
462 While most of the TuORFs we identified did not display clear stalling patterns (e.g., **Figure 1F-**
463 **G, S1B, S2A-N**), it remains possible that ribosomes have lower translation speeds, on average,
464 during uORF translation. Because ribosomes will quickly approach termination after initiation
465 due to the short length of uORFs, and because both initiation and termination are major time-
466 consuming steps in translation [104, 105], the average translation speed per codon for uORFs is
467 likely to be slower than that for longer ORFs. Alternatively, if cells mainly monitor the translation
468 efficiency of the mORF, then the poorer mORF translation caused by uORFs could explain the
469 faster degradation of TuORF mRNAs predicted by this model. In contrast, since ribosomes

470 actively engage with TuORFs, the “translation factor-protected” model does not fit TuORF
471 mRNAs, especially because their degradation relies on decapping and 5’ to 3’ decay in plants.
472 Further investigation into TuORF-mediated mRNA degradation will help elucidate the role of
473 uORFs in regulating mRNA stability.
474
475 **CONCLUSION**
476 Consistent with the observations in animals, we found that TuORFs in Arabidopsis reduce
477 mRNA stability and translation. However, unlike in animals, uORF translation generally does not
478 activate NMD in Arabidopsis. Our results reveal major differences in TuORF-mediated gene
479 expression between plants and animals.

480 **MATERIALS AND METHODS**

481 **Plant growth conditions and lysate preparation**

482 Arabidopsis Col-0 seeds were surface sterilized with 70% ethanol for 5 min, followed by 33%
483 bleach and 0.03% Tween 20 for 10 min, then rinsed with sterile water 5 times. The seeds were
484 imbibed at 4°C in the dark for 2 days, then grown hydroponically in sterile liquid media (2.15 g/L
485 Murashige and Skoog salt, 1% sucrose, 0.5 g/L MES, pH 5.7) while shaking at 85 rpm under a
486 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
487 7 days. At Zeitgeber time 4 (4 h after lights on), DMSO corresponding to 0.1% of the media
488 volume was added to the media (these were mock samples of our large-scale experiment). After
489 20 and 60 min, three biological replicates (~300 plants per sample) were harvested at each time
490 point and immediately flash frozen in liquid nitrogen.

491

492 Plant lysates were prepared as previously described [47]. Briefly, per 0.1 g of ground tissue
493 powder was resuspended in 400 μL of lysis buffer (100 mM Tris-HCl [pH 8], 40 mM KCl, 20 mM
494 $MgCl_2$, 2% [v/v] polyoxyethylene [10] tridecyl ether [Sigma, P2393], 1% [w/v] sodium
495 deoxycholate [Sigma, D6750], 1 mM dithiothreitol, 100 μg/mL cycloheximide [Sigma, C4859],
496 100 μg/mL chloramphenicol [Sigma R4408], and 10 units/mL DNase I [Epicenter, D9905K]).
497 The lysates were spun at 3,000 g for 3 min, and the supernatant was transferred to a new tube
498 and subsequently centrifuged at 20,000 g for 10 min. The supernatant was transferred to a new
499 tube and the RNA concentration was determined with 10x dilutions using the Qubit RNA HS
500 assay (Thermo Fisher Scientific; Q32852). Aliquots of 100 μL and 200 μL of the lysates were
501 made, and they were flash frozen in liquid nitrogen and stored at -80°C until further processing.

502

503 **Ribo-seq library construction**

504 Ribosome footprints were obtained using 200 μL of the lysates described above, and
505 sequencing libraries were constructed according to our previous method [47] with the following

506 modifications: briefly, after RNase I digestion (50 units nuclease per 40 µg of RNA; the nuclease
507 was included in TruSeq Mammalian Ribo Profile Kit, Illumina, RPHMR12126), adding 15 µl of
508 SUPERase-IN (Invitrogen, AM2696), and passing through a size exclusion column (Illustra
509 MicroSpin S-400 HR Columns; GE Healthcare; 27-5140-01), RNA > 17 nt was isolated with the
510 RNA Clean & Concentrator-25 kit (Zymo Research, R1017) and separated on 15% urea-TBE
511 gels (Invitrogen, EC68852BOX). Gel slices between 28 and 30 nt were isolated, and the RNAs
512 were purified as previously described. Next, rRNA depletion was performed using RiboZero
513 Plant Leaf kit (Illumina, MRZPL1224) in one quarter of the recommended reaction volume. The
514 major change we made here was to gel-purify the ribosome footprints prior to the rRNA
515 depletion. In our previous protocol, we purified the RNA twice (first >17 nt and then <200 nt) to
516 enrich ribosome footprints before the rRNA depletion and gel purification. We reasoned that
517 some footprints might be lost during the two RNA purification steps. In addition, the rRNA
518 depletion step requires a limited amount of input RNA according to the manufacturer's
519 recommendations. Therefore, we expected modifying the order of these steps could maximize
520 the input and minimize the footprint loss during the processes. Ribo-seq libraries were then
521 constructed using the TruSeq Mammalian Ribo Profile Kit (illumina, RPHMR12126) as
522 previously described with nine cycles of PCR amplification. Libraries with equal molarity were
523 pooled and sequenced on a Hi-Seq 4000 sequencer using single-end 50-bp sequencing.
524

525 **RNA-seq library construction**

526 Total RNA greater than 200 nt was purified from 100 µL of the lysates described above using
527 the RNA Clean & Concentrator-25 kit (Zymo Research, R1017) as previously described [47].
528 RNA integrity was evaluated using a Bioanalyzer (Agilent) RNA pico chip, and RNA integrity
529 numbers (RINs) ranging from 7.2 to 7.7 were observed among the samples. A total of 4 µg of
530 RNA per sample was subjected to rRNA depletion using the RiboZero Plant Leaf kit (Illumina,
531 MRZPL1224) following the manufacturer's recommendations. Then, 100 ng of rRNA-depleted

532 RNA was fragmented to around 200-nt long based on the RIN reported by the Bioanalyzer, and
533 strand-specific sequencing libraries were made using the NEBNext Ultra II Directional RNA
534 Library Prep Kit (New England Biolabs, E7760S) with eight cycles of amplification. Libraries of
535 equal molarity were pooled and sequenced on a Hi-Seq 4000 using paired-end 100-bp
536 sequencing.
537
538 **Data pre-processing and analysis**
539 Data pre-processing and analysis were performed similarly to that previously described [47],
540 except that the Araport11 annotation was used in this study. Briefly, for Ribo-seq libraries, the
541 adaptor (AGATCGGAAGAGCACACGTCT) was clipped with fastx_clipper (FASTX toolkit
542 v0.0.14) (http://hannonlab.cshl.edu/fastx_tool- kit/). For both RNA-seq and Ribo-seq, we used
543 Bowtie2 (v2.3.4.1) [106] to remove rRNA/tRNA/snRNA/snoRNA sequences. Both the RNA-seq
544 and Ribo-seq reads were mapped to the transcriptome with STAR aligner [107] (RNA-seq
545 parameters: --outFilterMismatchNmax 2 --outFilterMultimapNmax 20 --outFilterType BySJout --
546 alignSJoverhangMin 8 --alignSJDBoverhangMin 2; all parameters used for Ribo-seq were the
547 same except --outFilterMismatchNmax 1). We then used the bam files from both RNA-seq and
548 Ribo-seq for RiboTaper (v1.3.1a) [51] to identify translated ORFs. The distribution of Ribo-seq
549 reads in different genome features was calculated using Ribo-seQC. The Ribo-seq metaplot
550 was created using the create_metaplots.bash function in RiboTaper. The Ribo-seq read lengths
551 and offsets for RiboTaper were 24, 25, 26, 27, 28 and 8, 9, 10, 11, 12, respectively. Combining
552 both 20-min and 60-min samples, the total reads mapped to the genome for Ribo-seq and RNA-
553 seq were 298.01 million and 180.67 million, respectively. The discovered uORFs and mORFs
554 were extracted from the RiboTaper output ORF_max_filt file. To calculate translation efficiency,
555 we first used STAR to map the RNA-seq and Ribo-seq reads to the CDS of annotated coding
556 genes. The resulting bam files were used to quantify the transcripts per million (TPM) of each
557 gene with RSEM (v1.3.1) [108]. Then, translation efficiency was calculated by dividing the Ribo-

558 seq TPM by the RNA-seq TPM. The root and shoot data from Hsu et al., 2016 [47] were
559 reanalyzed using the Araport11 annotation. The P_sites_all files from RiboTaper were
560 processed using the following code: cut -f 1,3,6 P_sites_all | sort | uniq -c | sed -r 's/^(*[^]+)
561 +/\1\t/' > output.txt to combine the read counts at each P-site. Data visualization and statistical
562 analysis were performed in R (v4.0.3) [109]. In particular, the gene levels from the Ribo-seq and
563 RNA-seq reads were visualized with RiboPlotR [54].
564
565 The mRNA half-life and mRNA decay rate data were downloaded from Szabo et al. and
566 Sorenson et al., respectively [70, 85]. The quantitative proteomics data for Arabidopsis shoots
567 and roots were downloaded from Song et al. [71]. The expression levels (TPM) used for
568 investigating mRNA levels of uORF genes in different tissues, developmental stages, and
569 ecotypes in Arabidopsis and in tomato were directly downloaded from the EMBO-EBI
570 expression atlas (E-MTAB-7978: Arabidopsis tissue atlas, E-GEOD-53197: 17 Arabidopsis
571 thaliana accessions, E-MTAB-4812: tomato root, leaf, flower [two stages], and fruit [six stages],
572 and E-MTAB-4813: three longitudinal sections of six stages during tomato fruit development)
573 [74]. Our previous Ribo-seq and RNA-seq data for Arabidopsis roots and shoots were described
574 in [47]; our previous Ribo-seq and RNA-seq data for tomato roots were described in [72]. The
575 published RNA-seq data for the NMD mutants *upf1* (*lba1*) *upf3-1* and its wild-type control (GEO:
576 GSE41432), *smg7-1 pad4* and its *pad4* control (GEO: GSE55884), *upf1-3 pad4* and its *pad4*
577 control (GEO: GSE146340), and *upf2-10* and its wild-type control (SRA: SRP056795) were
578 downloaded, and the mRNA levels were quantified by Kalisto [110] and the TPM values were
579 further analyzed to compare the steady-state mRNA levels in different categories of genes.
580
581 **AVAILABILITY OF DATA AND MATERIALS**
582 All raw and processed sequencing data generated in this study have been submitted to the
583 NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession

number GSE183264. The genomic locations of TuORFs are available on TAIR genome browser (https://www.arabidopsis.org). Codes for this study are available at https://github.com/hsinyenwu/Wu-and-Hsu-2022-TuORF-nonNMD

COMPETING INTEREST STATEMENT

The authors declare that they have no competing interests.

FUNDING

This work was supported by a National Science Foundation grant (2051885) and a Michigan State University startup grant to PYH.

ACKNOWLEDGEMENTS

We thank Robert Last and Melissa Lehti-Shiu for critical reading and helpful comments on the manuscript. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by an NIH S10 OD018174 Instrumentation Grant.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:7507–12.
2. Xu G, Yuan M, Ai C, Liu L, Zhuang E, Karapetyan S, et al. uORF-mediated translation allows engineered plant disease resistance without fitness costs. Nature. 2017;545:491–4.

611 3. Zhang H, Si X, Ji X, Fan R, Liu J, Chen K, et al. Genome editing of upstream open reading
612 frames enables translational control in plants. Nature Biotechnology. 2018;36:894–8.
613 4. Lee DSM, Park J, Kromer A, Baras A, Rader DJ, Ritchie MD, et al. Disrupting upstream
614 translation in mRNAs is associated with human disease. Nature Communications 2021 12:1.
615 2021;12:1–14.
616 5. Kim BH, Cai X, Vaughn JN, Von Arnim AG. On the functions of the h subunit of eukaryotic
617 initiation factor 3 in late stages of translation initiation. Genome Biology. 2007;8:R60.
618 6. Johnstone TG, Bazzini AA, Giraldez AJ. Upstream ORFs are prevalent translational
619 repressors in vertebrates. The EMBO Journal. 2016;35:706–23.
620 7. Zhang H, Wang Y, Wu X, Tang X, Wu C, Lu J. Determinants of genome-wide distribution and
621 evolution of uORFs in eukaryotes. Nature Communications. 2021;12:1–17.
622 8. Chew GL, Pauli A, Schier AF. Conservation of uORF repressiveness and sequence features
623 in mouse, human and zebrafish. Nature Communications. 2016;7:11663.
624 9. Liu MJ, Wu SH, Wu JF, Lin WD, Wu YC, Tsai TY, et al. Translational landscape of
625 photomorphogenic Arabidopsis. Plant Cell. 2013;25:3699–710.
626 10. Wu HYL, Song G, Walley JW, Hsu PY. The tomato translational landscape revealed by
627 transcriptome assembly and ribosome profiling. Plant Physiology. 2019;181:367–80.
628 11. Lin Y, May GE, Kready H, Nazzaro L, Mao M, Spealman P, et al. Impacts of uORF codon
629 identity and position on translation regulation. Nucleic Acids Research. 2019;47:9358–67.
630 12. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5′-untranslated regions of
631 eukaryotic mRNAs. Science. 2016;352:1413–6.
632 13. Dever TE, Ivanov IP, Sachs MS. Conserved Upstream Open Reading Frame Nascent
633 Peptides That Control Translation. Annual Review of Genetics. 2020;54:237–64.
634 14. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999;234:187–208.
635 15. He F, Jacobson A. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts
636 Is only Part of the Story. Annual Review of Genetics. 2015;49:339–66.
637 16. Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by
638 nonsense-mediated mRNA decay. Nature Reviews Molecular Cell Biology 2019 20:7.
639 2019;20:406–20.
640 17. Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, Maquat LE. Upf1 Phosphorylation
641 Triggers Translational Repression during Nonsense-Mediated mRNA Decay. Cell.
642 2008;133:314–27.
643 18. Kurosaki T, Li W, Hoque M, Popp MWL, Ermolenko DN, Tian B, et al. A Post-Translational
644 regulatory switch on UPF1 controls targeted mRNA degradation. Genes and Development.

645 2014;28:1900–16.
646 19. Raxwal VK, Riha K. Nonsense mediated RNA decay and evolutionary capacitance.
647 Biochimica et Biophysica Acta. 2016;1859:1538–43.
648 20. Lloyd JPB, Davies B. SMG1 is an ancient nonsense-mediated mRNA decay effector. The
649 Plant journal : for cell and molecular biology. 2013;76:800–10.
650 21. Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, et al. Genome-wide
651 suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. Proceedings of
652 the National Academy of Sciences of the United States of America. 2009;106:2453–8.
653 22. Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, et al.
654 Alternative splicing and nonsense-mediated decay modulate expression of important regulatory
655 genes in Arabidopsis. Nucleic Acids Research. 2012;40:2454–69.
656 23. Drechsel G, Kahles A, Kesarwani AK, Stauffer E, Behr J, Drewe P, et al. Nonsense-
657 mediated decay of alternative precursor mRNA splicing variants is a major determinant of the
658 Arabidopsis steady state transcriptome. The Plant Cell. 2013;25:3726–42.
659 24. Degtiar E, Fridman A, Gottlieb D, Vexler K, Berezin I, Farhi R, et al. The feedback control of
660 UPF3 is crucial for RNA surveillance in plants. Nucleic Acids Research. 2015;43:4219–35.
661 25. Raxwal VK, Simpson CG, Gloggnitzer J, Entinze JC, Guo W, Zhang R, et al. Nonsense-
662 Mediated RNA Decay Factor UPF1 Is Critical for Posttranscriptional and Translational Gene
663 Regulation in Arabidopsis. The Plant Cell. 2020;32:2725–41.
664 26. Jeong HJ, Kim YJ, Kim SH, Kim YH, Lee IJ, Kim YK, et al. Nonsense-mediated mRNA
665 decay factors, UPF1 and UPF3, contribute to plant defense. Plant and Cell Physiology.
666 2011;52:2147–56.
667 27. Rayson S, Arciga-Reyes L, Wootton L, de Torres Zabala M, Truman W, Graham N, et al. A
668 role for nonsense-mediated mrna decay in plants: Pathogen responses are induced in
669 arabidopsis thaliana nmd mutants. PLoS ONE. 2012;7:e31917.
670 28. Riehs-Kearnan N, Gloggnitzer J, Dekrout B, Jonak C, Riha K. Aberrant growth and lethality
671 of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like
672 response. Nucleic Acids Research. 2012;40:5615–24.
673 29. Filichkin SA, Cumbie JS, Dharmawardhana P, Jaiswal P, Chang JH, Palusa SG, et al.
674 Environmental stresses modulate abundance and timing of alternatively spliced circadian
675 transcripts in Arabidopsis. Molecular Plant. 2015;8:207–27.
676 30. Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, et al. Gene-specific
677 translation regulation mediated by the hormone-signaling molecule EIN2. Cell. 2015;163:684–
678 97.

679 31. Chiam N-C, Fujimura T, Sano R, Akiyoshi N, Hiroyama R, Watanabe Y, et al. Nonsense-
680 Mediated mRNA Decay Deficiency Affects the Auxin Response and Shoot Regeneration in
681 Arabidopsis. Plant and Cell Physiology. 2019;60:2000–14.
682 32. Yoine M, Ohto MA, Onai K, Mita S, Nakamura K. The lba1 mutation of UPF1 RNA helicase
683 involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and
684 altered sugar signalling in Arabidopsis. The Plant Journal. 2006;47:49–62.
685 33. Kertész S, Kerényi Z, Mérai Z, Bartos I, Pálfy T, Barta E, et al. Both introns and long 3′-
686 UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. Nucleic
687 Acids Research. 2006;34:6147–57.
688 34. Hurt JA, Robertson AD, Burge CB. Global analyses of UPF1 binding and function reveal
689 expanded scope of nonsense-mediated mRNA decay. Genome Research. 2013;23:1636–50.
690 35. Baird TD, Cheng KCC, Chen YC, Buehler E, Martin SE, Inglese J, et al. ICE1 promotes the
691 link between splicing and nonsense-mediated mRNA decay. eLife. 2018;7.
692 36. Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D. Plant upstream ORFs can trigger
693 nonsense-mediated mRNA decay in a size-dependent manner. Plant Molecular Biology.
694 2009;71:367–78.
695 37. Uchiyama-Kadokura N, Murakami K, Takemoto M, Koyanagi N, Murota K, Naito S, et al.
696 Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of
697 the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in Arabidopsis thaliana. Plant
698 and Cell Physiology. 2014;55:1556–67.
699 38. Kurihara Y, Makita Y, Kawashima M, Fujita T, Iwasaki S, Matsui M. Transcripts from
700 downstream alternative transcription start sites evade uORF-mediated inhibition of gene
701 expression in Arabidopsis. Proceedings of the National Academy of Sciences of the United
702 States of America. 2018;115:7831–6.
703 39. Saul H, Elharrar E, Gaash R, Eliaz D, Valenci M, Akua T, et al. The upstream open reading
704 frame of the Arabidopsis AtMHX gene has a strong impact on transcript accumulation through
705 the nonsense-mediated mRNA decay pathway. The Plant Journal. 2009;60:1031–42.
706 40. Tanaka M, Sotta N, Yamazumi Y, Yamashita Y, Miwa K, Murota K, et al. The minimum open
707 reading frame, AUG-Stop, induces boron-dependent ribosome stalling and mRNA degradation.
708 The Plant Cell. 2016;28:2830–49.
709 41. Li YR, Liu MJ. Prevalence of alternative AUG and non-AUG translation initiators and their
710 regulatory effects across plants. Genome Research. 2020;30:1418–33.
711 42. Nagarajan VK, Kukulich PM, Von Hagel B, Green PJ. RNA degradomes reveal substrates
712 and importance for dark and nitrogen stress responses of Arabidopsis XRN4. Nucleic Acids

713 Research. 2019;47:9216–30.
714 43. Ohtani M, Wachter A. NMD-Based Gene Regulation - A Strategy for Fitness Enhancement
715 in Plants? Plant and Cell Physiology. 2019;60:1953–60.
716 44. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo
717 of translation with nucleotide resolution using ribosome profiling. Science. 2009;324:218–23.
718 45. Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of
719 protein synthesis. Nature Reviews Molecular Cell Biology. 2015;16:651–64.
720 46. Juntawong P, Girke T, Bazin J, Bailey-Serres J. Translational dynamics revealed by
721 genome-wide profiling of ribosome footprints in Arabidopsis. Proceedings of the National
722 Academy of Sciences of the United States of America. 2014;111:E203–12.
723 47. Hsu PY, Calviello L, Wu H-YL, Li F-W, Rothfels CJ, Ohler U, et al. Super-resolution
724 ribosome profiling reveals unannotated translation events in Arabidopsis. Proceedings of the
725 National Academy of Sciences of the United States of America. 2016;113:E7126–35.
726 48. Hu Q, Merchante C, Stepanova AN, Alonso JM, Heber S. Genome-Wide Search for
727 Translated Upstream Open Reading Frames in Arabidopsis Thaliana. IEEE Transactions on
728 Nanobioscience. 2016;15:150–9.
729 49. Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J. Global analysis
730 of ribosome-associated noncoding RNAs unveils new modes of translational regulation.
731 Proceedings of the National Academy of Sciences of the United States of America.
732 2017;114:E10018–27.
733 50. Kurihara Y, Makita Y, Shimohira H, Fujita T, Iwasaki S, Matsui M. Translational landscape of
734 protein-coding and non-protein-coding RNAs upon light exposure in Arabidopsis. Plant and Cell
735 Physiology. 2020;61:536–45.
736 51. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, et al. Detecting
737 actively translated open reading frames in ribosome profiling data. Nature Methods.
738 2016;13:165–70.
739 52. Xu Z, Hu L, Shi B, Geng S, Xu L, Wang D, et al. Ribosome elongating footprints denoised
740 by wavelet transform comprehensively characterize dynamic cellular translation events. Nucleic
741 Acids Research. 2018;46:e109.
742 53. Harnett D, Meerdink E, Calviello L, Sydow D, Ohler U. Genome-Wide Analysis of Actively
743 Translated Open Reading Frames Using RiboTaper/ORFquant. Methods in Molecular Biology.
744 2021;2252:331–46.
745 54. Wu H-YL, Hsu PY. RiboPlotR: a visualization tool for periodic Ribo-seq reads. Plant
746 Methods. 2021;17:1–7.

747 55. Von Arnim AG, Jia Q, Vaughn JN. Regulation of plant translation by upstream open reading
748 frames. Plant Science. 2014;214:1–12.
749 56. Silverstone AL, Ciampaglio CN, Sun TP. The Arabidopsis RGA gene encodes a
750 transcriptional regulator repressing the gibberellin signal transduction pathway. The Plant Cell.
751 1998;10:155–69.
752 57. Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, et al. Cloning of the
753 Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. Science.
754 2000;289:768–71.
755 58. Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, et al. The
756 SCARECROW gene regulates an asymmetric cell division that is essential for generating the
757 radial organization of the Arabidopsis root. Cell. 1996;86:423–33.
758 59. Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW. AUXIN RESPONSE
759 FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission
760 in Arabidopsis thaliana. Development. 2005;132:4563–74.
761 60. Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, Guilfoyle TJ, et al. Auxin response
762 factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. Development.
763 2005;132:4107–18.
764 61. Ni M, Tepperman JM, Quail PH. PIF3, a phytochrome-interacting factor necessary for
765 normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. Cell.
766 1998;95:657–67.
767 62. Huq E, Quail PH. PIF4, a phytochrome-interacting bHLH factor, functions as a negative
768 regulator of phytochrome B signaling in Arabidopsis. EMBO Journal. 2002;21:2441–50.
769 63. Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, et al. The Arabidopsis
770 phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to
771 prolonged red light by modulating phyB levels. The Plant Cell. 2008;20:337–52.
772 64. Sharrock RA, Quail PH. Novel phytochrome sequences in Arabidopsis thaliana: structure,
773 evolution, and differential expression of a plant regulatory photoreceptor family. Genes &
774 Development. 1989;3:1745–57.
775 65. Devlin PF, Patel SR, Whitelam GC. Phytochrome E Influences Internode Elongation and
776 Flowering Time in Arabidopsis. The Plant Cell. 1998;10:1479–87.
777 66. Zhang SH, Lawton MA, Hunter T, Lamb CJ. atpk1, a novel ribosomal protein kinase gene
778 from Arabidopsis. I. Isolation, characterization, and expression. Journal of Biological Chemistry.
779 1994;269:17586–92.
780 67. Mizoguchi T, Hayashida N, Yamaguchi-Shinozaki K, Kamada H, Shinozaki K. Two genes

781 that encode ribosomal-protein S6 kinase homologs are induced by cold or salinity stress in
782 Arabidopsis thaliana. FEBS Letters. 1995;358:199–204.
783 68. Park YS, Hong SW, Oh SA, Kwak JM, Lee HH, Nam HG. Two putative protein kinases from
784 Arabidopsis thaliana contain highly acidic domains. Plant Molecular Biology. 1993;22:615–24.
785 69. Boudsocq M, Barbier-Brygoo H, Laurière C. Identification of nine sucrose nonfermenting 1-
786 related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana.
787 Journal of Biological Chemistry. 2004;279:41758–66.
788 70. Szabo EX, Reichert P, Lehniger MK, Ohmer M, de Francisco Amorim M, Gowik U, et al.
789 Metabolic labeling of RNAs uncovers hidden features and dynamics of the Arabidopsis
790 transcriptome. The Plant Cell. 2020;32:871–87.
791 71. Song G, Hsu PY, Walley JW. Assessment and Refinement of Sample Preparation Methods
792 for Deep and Quantitative Plant Proteome Profiling. Proteomics. 2018;18:e1800220.
793 72. Wu H-YL, Song G, Walley JW, Hsu PY. Translational landscape in tomato revealed by
794 transcriptome assembly and ribosome profiling. bioRxiv. 2019;181:367–80.
795 73. Clarke JT, Warnock RCM, Donoghue PCJ. Establishing a time-scale for plant evolution. The
796 New Phytologist. 2011;192:266–301.
797 74. Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S, et al.
798 Expression Atlas update: From tissues to single cells. Nucleic Acids Research. 2020;48:D77–
799 83.
800 75. Hayden CA, Jorgensen RA. Identification of novel conserved peptide uORF homology
801 groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and
802 preferential association with transcription factor-encoding genes. BMC Biology. 2007;5:32.
803 76. Jorgensen RA, Dorantes-Acosta AE. Conserved peptide upstream open reading frames are
804 associated with regulatory genes in angiosperms. Frontiers in Plant Science. 2012;3 AUG:191.
805 77. Takahashi H, Takahashi A, Naito S, Onouchi H. BAIUCAS: A novel BLAST-based algorithm
806 for the identification of upstream open reading frames with conserved amino acid sequences
807 and its application to the Arabidopsis thaliana genome. Bioinformatics. 2012;28:2231–41.
808 78. Vaughn JN, Ellingson SR, Mignone F, Arnim A von. Known and novel post-transcriptional
809 regulatory sequences are conserved across plant families. RNA. 2012;18:368–84.
810 79. Van Der Horst S, Snel B, Hanson J, Smeekens S. Novel pipeline identifies new upstream
811 ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5′ leader
812 of mRNAs in Arabidopsis thaliana. RNA. 2019;25:292–304.
813 80. Wolfa YI, Novichkovb PS, Kareva GP, Koonina E V., Lipmana DJ. The universal distribution
814 of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different

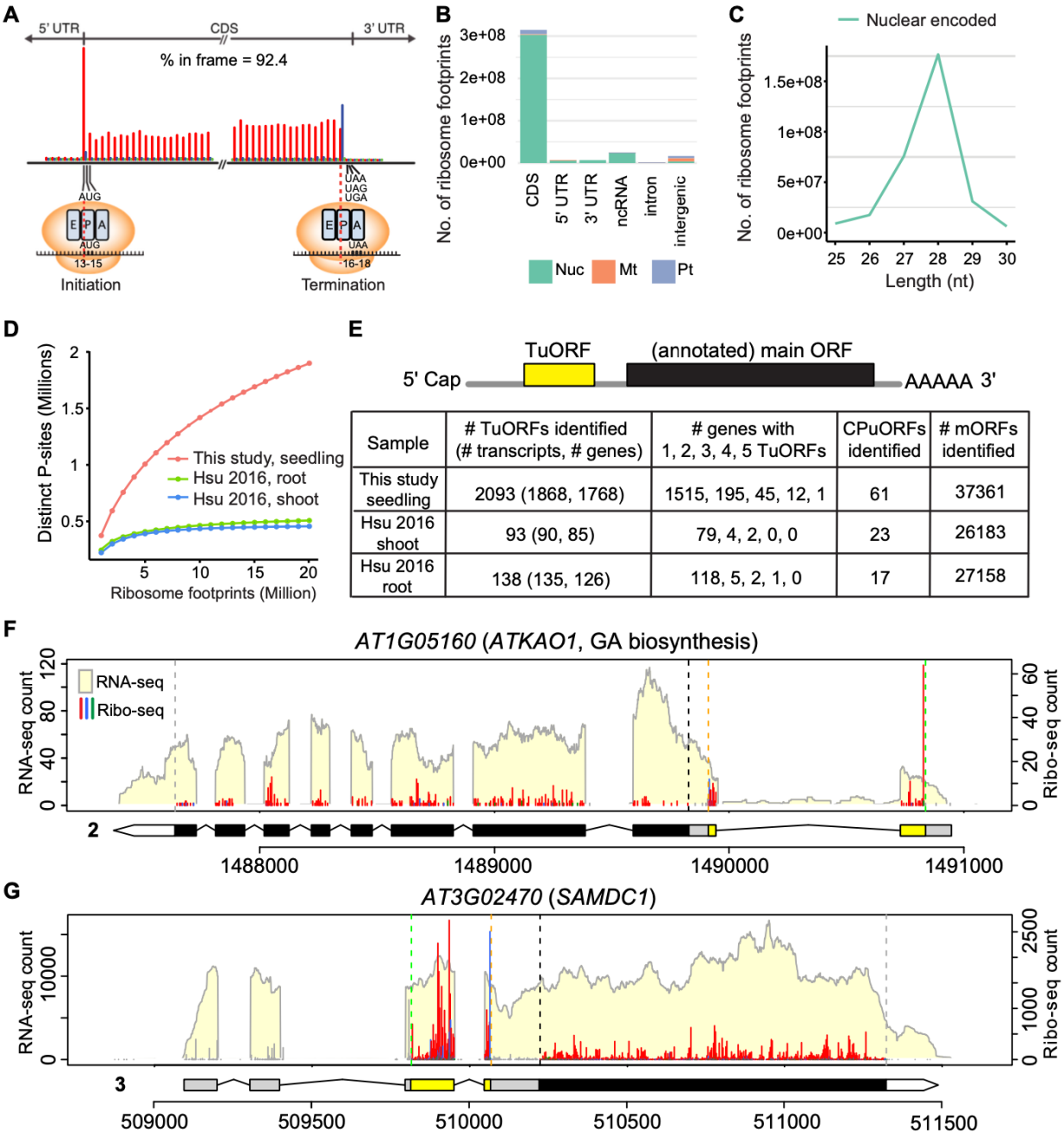
815 apparent ages. Proceedings of the National Academy of Sciences of the United States of
816 America. 2009;106:7273–80.
817 81. Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, Zhang D. Young genes out of the male: An
818 insight from evolutionary age analysis of the pollen transcriptome. Molecular Plant. 2015;8:935–
819 45.
820 82. Gloggnitzer J, Akimcheva S, Srinivasan A, Kusenda B, Riehs N, Stampfl H, et al. Nonsense-
821 mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial
822 defense. Cell Host and Microbe. 2014;16:376–90.
823 83. Zinshteyn B, Sinha NK, Enam SU, Koleske B, Green R. Translational repression of NMD
824 targets by GIGYF2 and EIF4E2. PLoS Genetics. 2021;17 10 October:e1009813.
825 84. Sieburth LE, Vincent JN. Beyond transcription factors: Roles of mrna decay in regulating
826 gene expression in plants [version 1; referees: 3 approved]. F1000Research. 2018;7:1940.
827 85. Sorenson RS, Deshotel MJ, Johnson K, Adler FR, Sieburth LE. Arabidopsis mRNA decay
828 landscape arises from specialized RNA decay substrates, decapping-mediated feedback, and
829 redundancy. Proceedings of the National Academy of Sciences of the United States of America.
830 2018;115:E1485–94.
831 86. Deyholos MK, Cavaness GF, Hall B, King E, Punwani J, Van Norman J, et al. Varicose, a
832 WD-domain protein, is required for leaf blade. Development. 2003;130:6577–88.
833 87. Goeres DC, Van Norman JM, Zhang W, Fauver NA, Spencer M Lou, Sieburth LE.
834 Components of the Arabidopsis mRNA decapping complex are required for early seedling
835 development. The Plant Cell. 2007;19:1549–64.
836 88. Xu J, Yang JY, Niu QW, Chua NH. Arabidopsis DCP2, DCP1, and VARICOSE form a
837 decapping complex required for postembryonic development. The Plant Cell. 2006;18:3386–98.
838 89. Zhang W, Murphy C, Sieburth LE. Conserved RNaseII domain protein functions in
839 cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes.
840 Proceedings of the National Academy of Sciences of the United States of America.
841 2010;107:15981–5.
842 90. Andreev DE, Arnold M, Kiniry SJ, Loughran G, Michel AM, Rachinskii D, et al. TASEP
843 modelling provides a parsimonious explanation for the ability of a single uORF to derepress
844 translation during the integrated stress response. eLife. 2018;7:e32563.
845 91. Kerényi F, Wawer I, Sikorski PJ, Kufel J, Silhavy D. Phosphorylation of the N- and C-
846 terminal UPF1 domains plays a critical role in plant nonsense-mediated mRNA decay. Plant
847 Journal. 2013;76:836–48.
848 92. Chang HM, Triboulet R, Thornton JE, Gregory RI. A role for the Perlman syndrome

849 exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature. 2013;497:244–8.
850 93. Lubas M, Damgaard CK, Tomecki R, Cysewski D, Jensen TH, Dziembowski A.
851 Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human
852 cytoplasmic mRNA. The EMBO journal. 2013;32:1855–68.
853 94. Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T. The dwarf
854 phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an upstream ORF of a
855 bHLH gene. Development. 2006;133:3575–85.
856 95. Cai Q, Fukushima H, Yamamoto M, Ishii N, Sakamoto T, Kurata T, et al. The SAC51 family
857 plays a central role in thermospermine responses in arabidopsis. Plant and Cell Physiology.
858 2016;57:1583–92.
859 96. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a
860 major determinant of mRNA stability. Cell. 2015;160:1111–24.
861 97. Radhakrishnan A, Chen YH, Martin S, Alhusaini N, Green R, Coller J. The DEAD-Box
862 Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell.
863 2016;167:122-132.e9.
864 98. Beelman CA, Parker R. Differential effects of translational inhibition in cis and in trans on the
865 decay of the unstable yeast MFA2 mRNA. Journal of Biological Chemistry. 1994;269:9687–92.
866 99. LaGrandeur T, Parker R. The cis acting sequences responsible for the differential decay of
867 the unstable MFA2 and stable PGK1 transcipts in yeast include the context of the translational
868 start codon. RNA. 1999;5:420–33.
869 100. Schwartz DC, Parker R. mRNA Decapping in Yeast Requires Dissociation of the Cap
870 Binding Protein, Eukaryotic Translation Initiation Factor 4E. Molecular and Cellular Biology.
871 2000;20:7933–42.
872 101. Chan LY, Mugler CF, Heinrich S, Vallotton P, Weis K. Non-invasive measurement of
873 mRNA decay reveals translation initiation as the major determinant of mRNA stability. eLife.
874 2018;7:e32536.
875 102. Hou C-Y, Lee W-C, Chou H-C, Chen A-P, Chou S-J, Chen H-M. Global analysis of
876 truncated RNA ends reveals new insights into ribosome stalling in plants. The Plant Cell.
877 2016;28:2398–416.
878 103. Yu X, Willmann MR, Anderson SJ, Gregory BD. Genome-wide mapping of uncapped and
879 cleaved transcripts reveals a role for the nuclear mrna cap-binding complex in cotranslational
880 rna decay in Arabidopsis. The Plant Cell. 2016;28:2385–97.
881 104. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-Limiting Steps in Yeast Protein
882 Translation. Cell. 2013;153:1589–601.

883 105. Lawson MR, Lessen LN, Wang J, Prabhakar A, Corsepius NC, Green R, et al.
884 Mechanisms that ensure speed and fidelity in eukaryotic translation termination. Science.
885 2021;373:876–82.
886 106. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods.
887 2012;9:357–9.
888 107. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
889 universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
890 108. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or
891 without a reference genome. BMC Bioinformatics. 2011;12:323.
892 109. R Core Team (2013). R: A language and environment for statistical computing. R
893 Foundation for Statistical Computing, Vienna, Austria. 2017.
894 110. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
895 quantification. Nature biotechnology. 2016;34:525–7.
896

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

Figure 1. Enhanced Ribo-seq coverage improves uORF identification

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

903 indicate they are in the first (expected), second, and third reading frames, respectively.
904 The majority of footprints were mapped to the CDS in the expected reading frame
905 (92.4% in frame).
906 (B) Genomic features mapped by ribosome footprints. Reads that mapped to nuclear (Nuc)-,
907 mitochondrial (Mt)-, and plastid (Pt)-encoded genes are shown.
908 (C) Length distribution of ribosome footprints. Reads that mapped to nuclear-encoded genes
909 are presented.
910 (D) Distinct P-sites detected in 1 to 20 million randomly selected ribosome footprints from
911 our current and previous datasets.
912 (E) Numbers of TuORFs, CPuORFs, and mORFs identified in our current and previous
913 datasets.
914 (F-G) RNA-seq and Ribo-seq profiles of *ATKAO1 and SAMDC1* from our current data. RNA-
915 seq coverage is shown with a light-yellow background. Ribo-seq reads are presented
916 with their first nt at the P-site, and they are colored in red, blue, and green to indicate
917 they are in the first (expected), second, and third reading frames, respectively. Within the
918 gene models, yellow boxes represent the TuORFs, black boxes represent the annotated
919 mORFs, and gray and white regions indicate 5' UTRs and 3' UTRs, respectively. The
920 specific isoform being considered is indicated to the left of the gene model. Within the
921 profiles, green and orange vertical dashed lines represent translation start and stop,
922 respectively, for the TuORF. Black and gray vertical dashed lines represent translation
923 start and stop, respectively, for the annotated mORF.
924
925
926
927
928
929

Figure 2

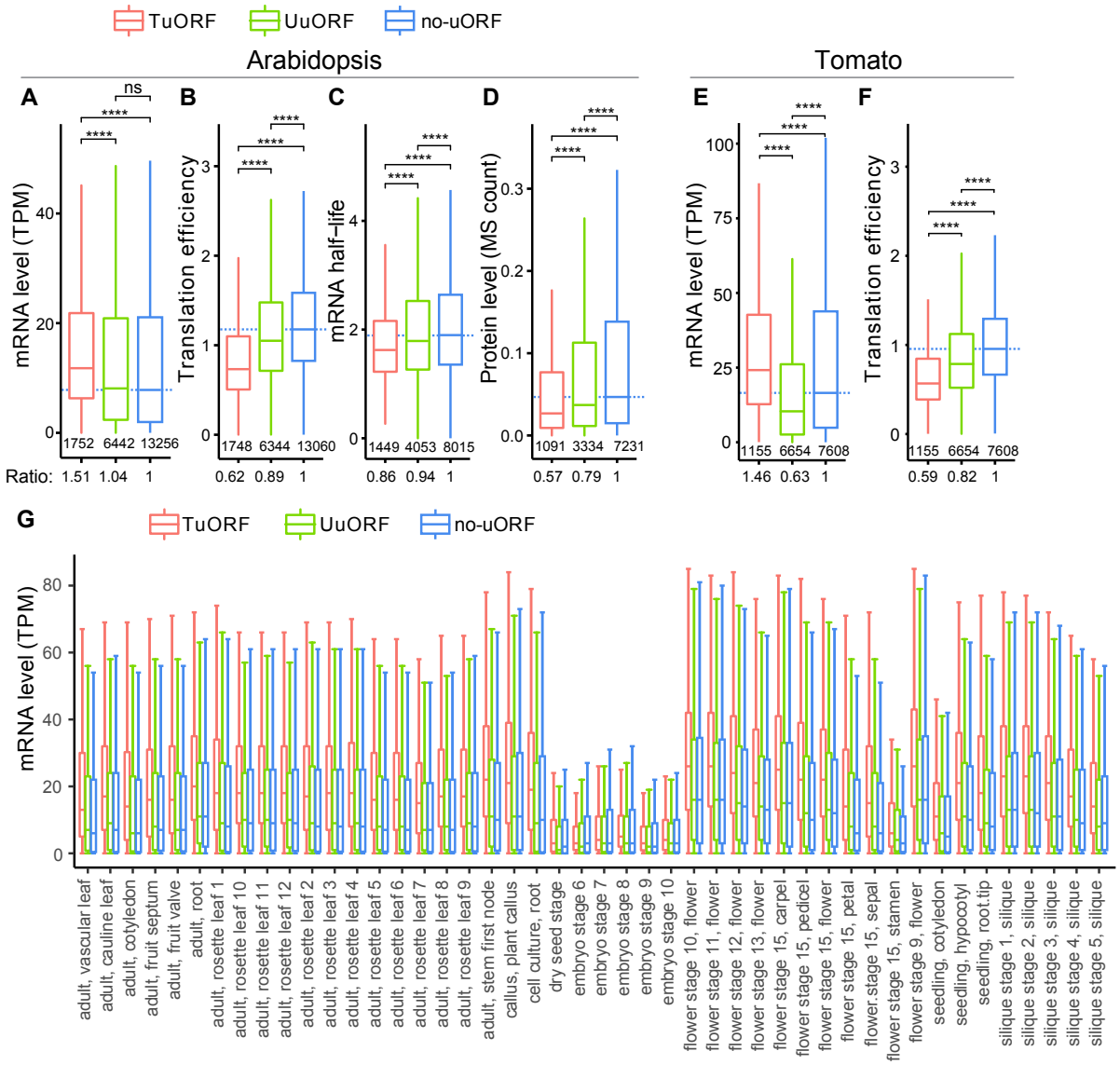


Figure 2. TuORF-containing genes have higher mRNA abundance but are associated with lower translation efficiency, shorter mRNA half-lives, and lower protein abundance

(A-D) The mRNA levels, translation efficiency, mRNA half-lives, and protein levels of TuORF, UuORF, and no-uORF genes in Arabidopsis seedlings. The mRNA half-life data, which were measured with metabolic labeling in Arabidopsis seedlings, were extracted from [70], and the protein abundance as determined by quantitative proteomics in Arabidopsis seedling shoots were extracted from [71]. Within the boxplots, the dashed lines mark the

938 median level of the no-uORF genes; the number of genes in each category is listed below
939 the lower whisker. The ratios underneath the plots indicate the median of each group
940 normalized to that of the no-uORF genes. The statistical significance was determined by
941 Wilcoxon signed-rank test. The adjusted p-value was determined using the Benjamini and
942 Yekutieli procedure to control for the false discovery rate in multiple testing (* 0.05 > p >
943 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-4, **** 1e-4 > p > 0). These statistical
944 analyses and graphic layouts are the same for all boxplots in the study.
945 (E-F) The mRNA levels and translation efficiency of TuORF, UuORF, and no-uORF genes from
946 a reanalysis of our previous tomato root data [72]. The dashed lines mark the median level
947 of the no-uORF genes.
948 (G) The mRNA levels of TuORF, UuORF, and no-uORF genes in 46 Arabidopsis tissues and
949 developmental stages from independent studies. The TPM values from RNA-seq
950 quantification were downloaded from the EMBO-EBI expression atlas [74]. Additional
951 Arabidopsis and tomato RNA-seq datasets were presented in Figure S5.
952
953
954

Figure 3

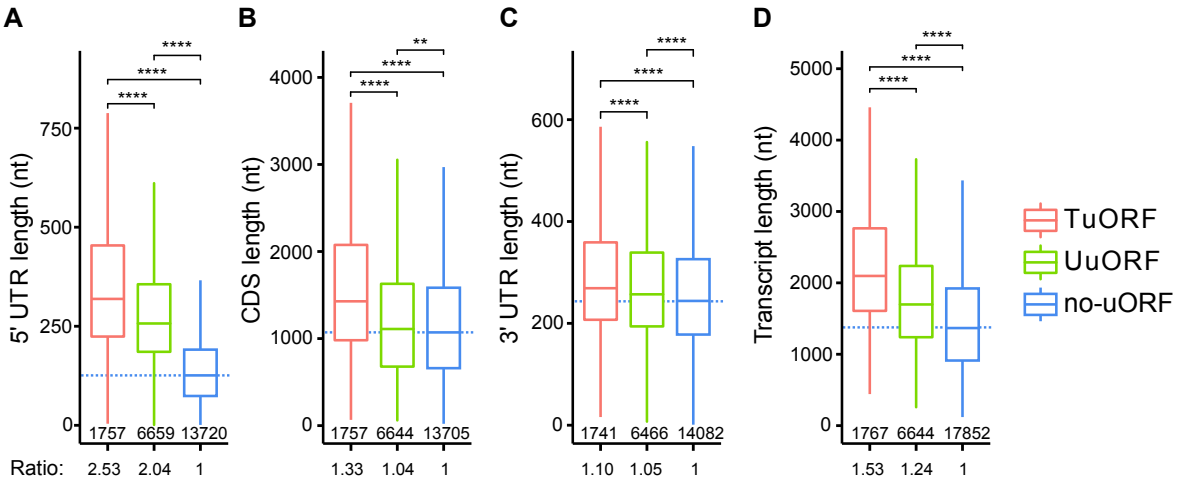


Figure 3. TuORF genes are larger than other genes

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

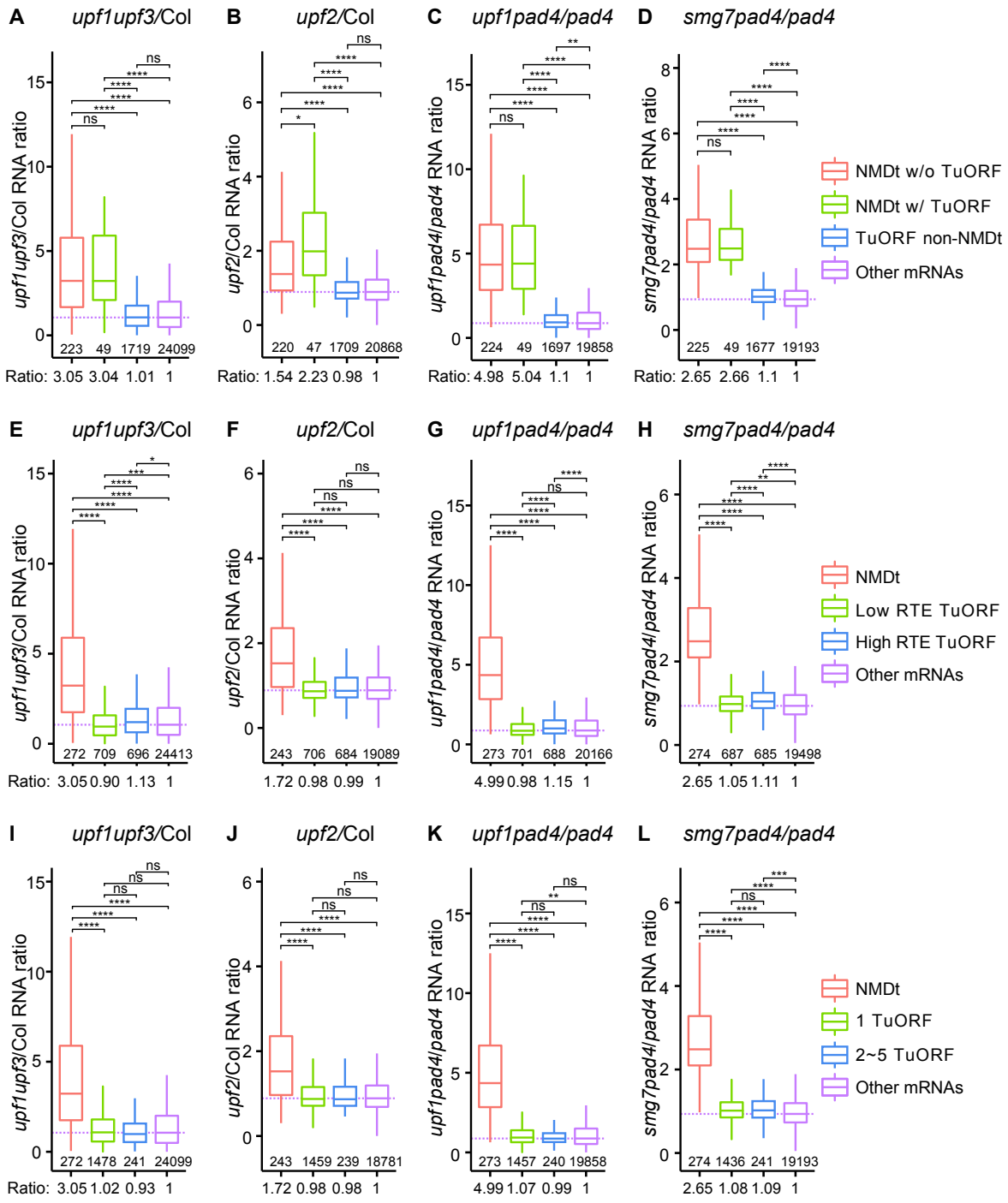


Figure 4: TuORF mRNAs are not upregulated in the NMD mutants

The ratios of mRNA levels between NMD mutants and their control plants from four RNA-seq datasets [23, 25, 30, 82] were used to examine the effects of the NMD pathway on NMD targets

967 (NMDt) and TuORF genes. (A) *upf1* (*lba1* allele) *upf3-1* double mutant compared with wild type
968 (Col-0). (B) *upf2-10* compared with wild type (Col-0). (C) *upf1-3 pad4* compared with *pad4*. (D)
969 *smg7-1 pad4* compared to *pad4*. See Table S3 for specific growth conditions for each dataset.
970 (E-H) Same datasets as 4A-D, but the TuORF genes were grouped based on low or high
971 relative TuORF translation efficiency (RTE) (i.e., TE_{TuORF} / TE_{mORF}). (I-L) Same datasets as 4A-
972 D, but the TuORF genes were grouped based on the TuORF number(s). Within the boxplots,
973 the dashed lines mark the median level of other mRNAs; the number of genes in each category
974 is listed below the lower whisker. The ratios underneath the plots indicate the median of each
975 group normalized to that of other mRNAs. The statistical significance was determined as
976 described in the Figure 2 legend.

Figure 5

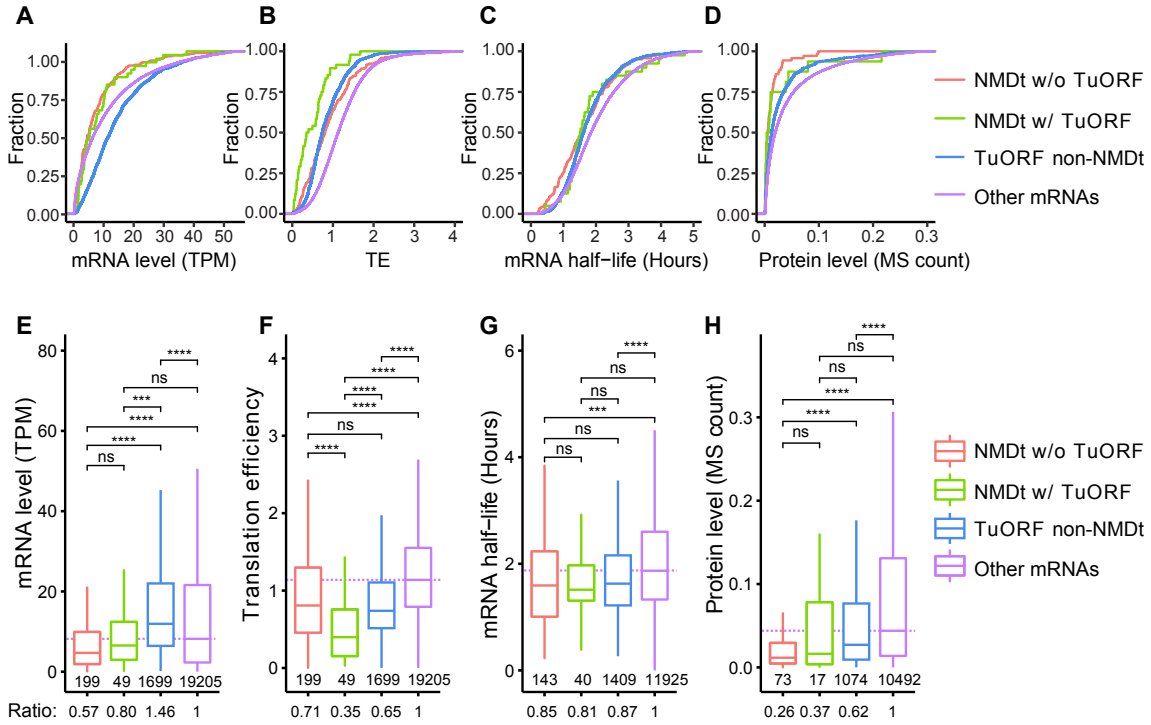


Figure 5. Expression patterns of NMD targets and TuORF mRNAs.

Cumulative plots (A-D) and boxplots (E-H) for the mRNA levels (A, E), translation efficiency (B, F), mRNA half-lives (C, G), and protein levels (D, H) 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

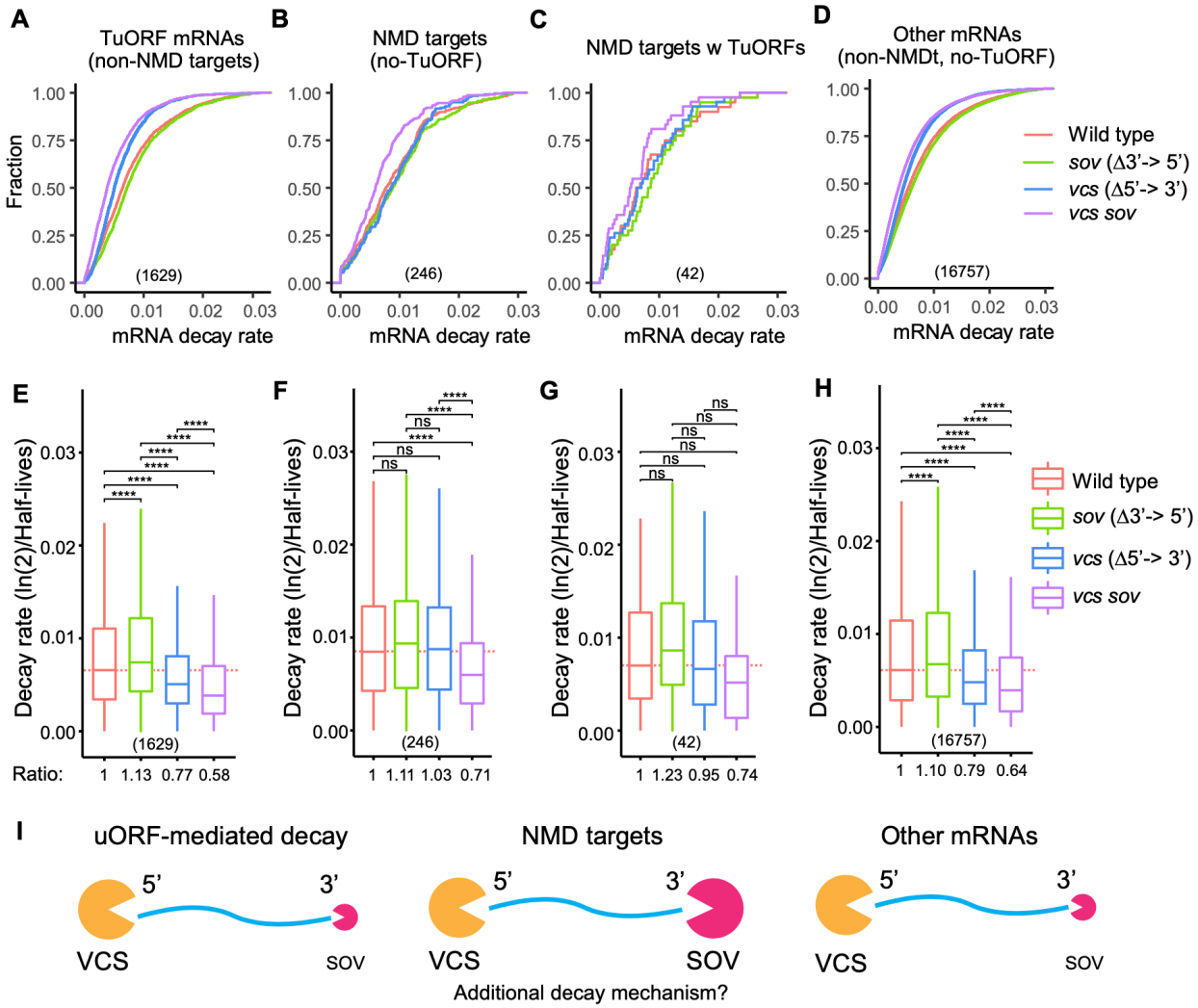


Figure 6. TuORF mRNAs and NMD targets are degraded through different mRNA decay mechanisms

Cumulative plots (A-D) and boxplots (E-H) showing the mRNA decay rates for NMD targets and TuORF mRNAs 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 [85]. 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 [87, 89]. 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 [85]. Within the boxplots, the dashed lines mark the median level of

998 wild type; the number of genes in each category is listed below the lower whisker. The ratios
999 underneath the plots indicate the median of each group normalized to that of wild type. The
1000 statistical significance was determined as described in the Figure 2 legend. (I) Model of mRNA
1001 decay for NMD targets, TuORF mRNAs, and other mRNAs. The orange packman indicates 5'-
1002 to-3' decay machinery, and the hot pink packman indicates 3'-to-5' decay machinery.
1003

1004 **SUPPLEMENTARY DATA**

1005 **Figure S1.** Correlations among samples and comparisons between our previous and current
1006 datasets

1007 **Figure S2.** Examples of TuORFs in important regulatory genes.

1008 **Figure S3.** Cumulative plots for the mRNA levels, translation efficiency, mRNA half-lives and
1009 protein abundance of uORF-containing genes

1010 **Figure S4.** Protein abundance of the uORF-containing genes in Arabidopsis root

1011 **Figure S5.** mRNA levels of uORF-containing genes in different ecotypes, growth stages, and
1012 tissues in Arabidopsis and tomato

1013 **Figure S6.** mRNA levels of CPuORF-containing genes compared to those of no-uORF genes

1014 **Figure S7.** (A) TuORF translation levels are positively correlated with the mRNA levels. (B) The
1015 distributions of the mRNA levels of TuORF and UuORF genes in Arabidopsis seedlings. Vertical
1016 dashed lines indicate the median value of each group.

1017 **Figure S8.** Lengths of the uORF peptides for NMD target and non-NMD target TuORFs

1018 **Figure S9.** Lengths of UTRs and CDS for NMD targets and TuORF genes

1019 **Figure S10.** The mRNA decay rates of NMD targets and TuORF-containing mRNAs in different
1020 genetic backgrounds

1021 **Table S1**. Translated ORFs identified by RiboTaper

1022 - **Table S1A-C.** Translated uORFs identified in the current dataset and our previous
1023 Arabidopsis shoot and root datasets
1024 - **Table S1D-F.** Translated mORFs identified in the current dataset and our previous
1025 Arabidopsis shoot and root datasets

1026 **Table S2.** GO-term analysis of TuORF genes

1027 **Table S3.** Summary of the large-scale datasets used in this study

1028 **Table S4.** NMD target gene list

1029 - **Table S4A.** High-confidence NMD targets expressed in our data

1030 • **Table S4B.** 49 genes identified as high-confidence NMD target TuORF genes

SUPPLEMENTARY FIGURES

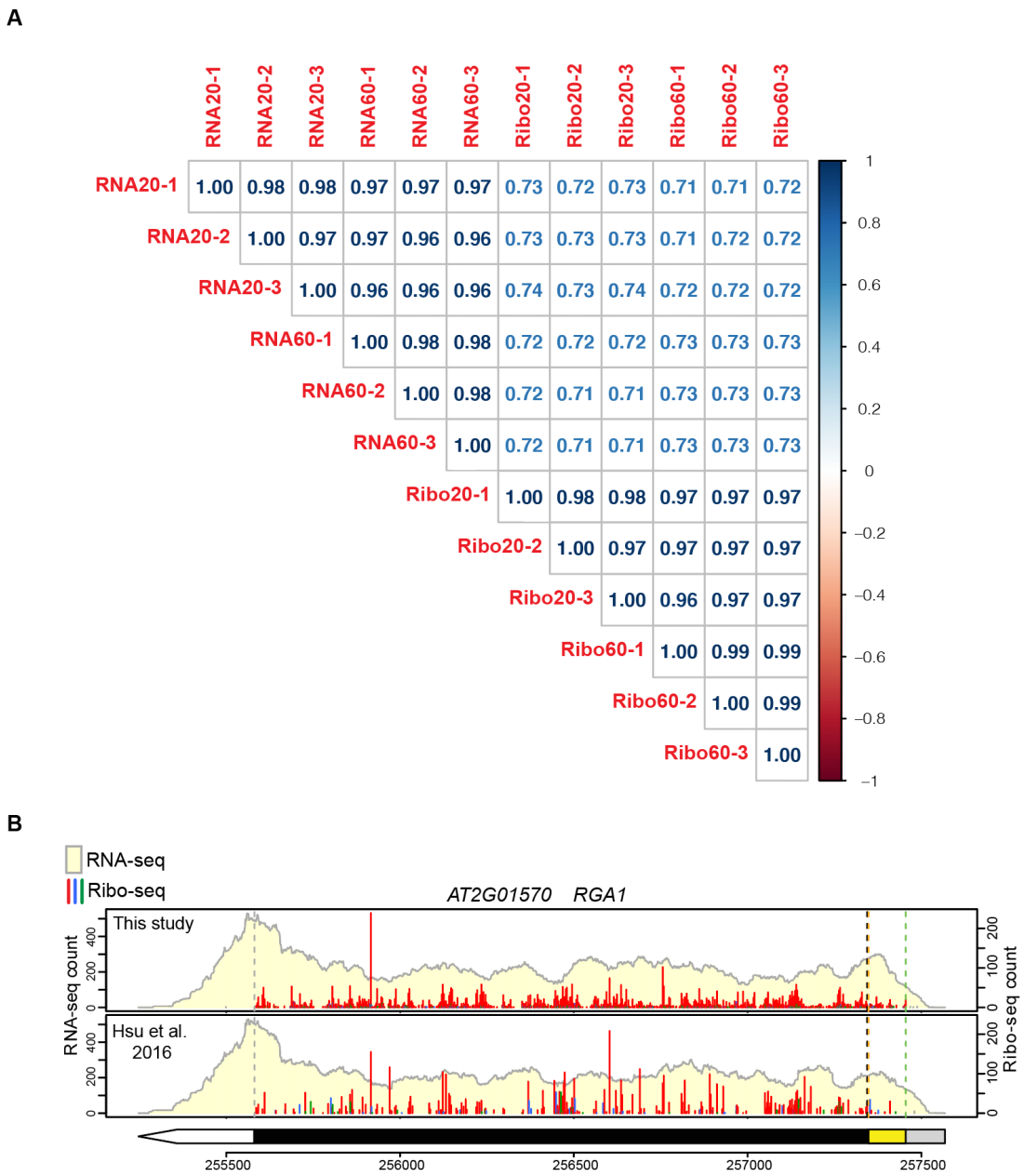
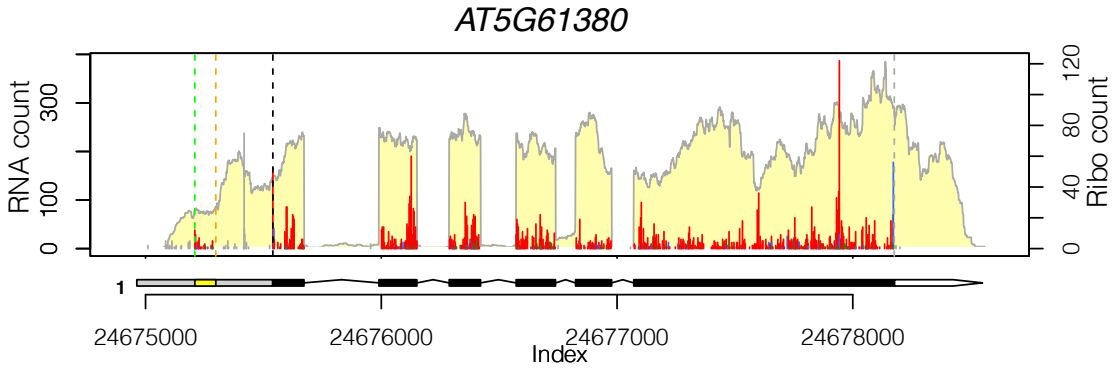


Figure S1. Correlations among samples and comparisons between previous and current RNA-seq and Ribo-seq datasets. (A) Pearson correlations among three biological replicates of our current RNA-seq and Ribo-seq analyses of samples harvested after 20 or 60 min of DMSO treatment (these are control samples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

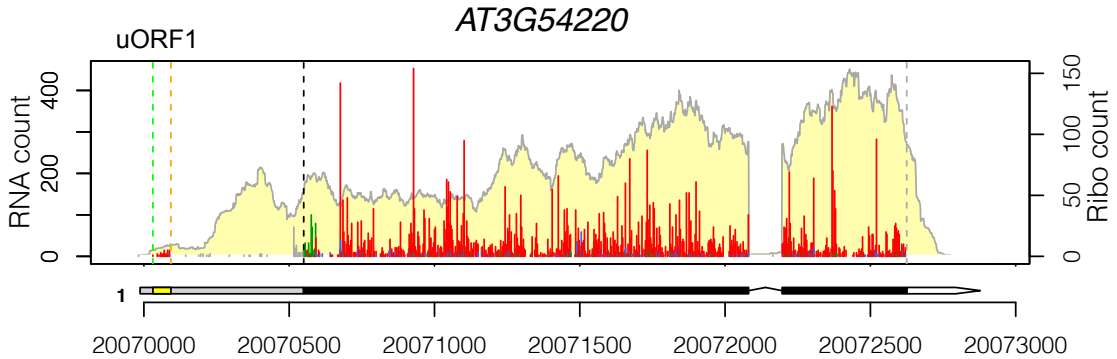
both the uORF and mORF regions in our current dataset; *RGA1* is shown as an example. Note that the mRNA levels are similar between the previous data (Hsu et al. 2016 [1], shoot) and the current data. Data representations are the same as those described in the Figure 1F legend.

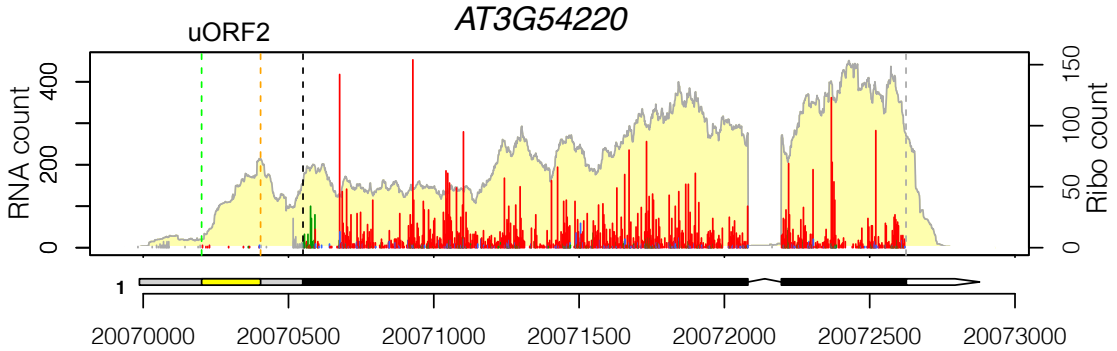
Figure S2. TuORFs in known important regulatory genes

A. *TOC1*

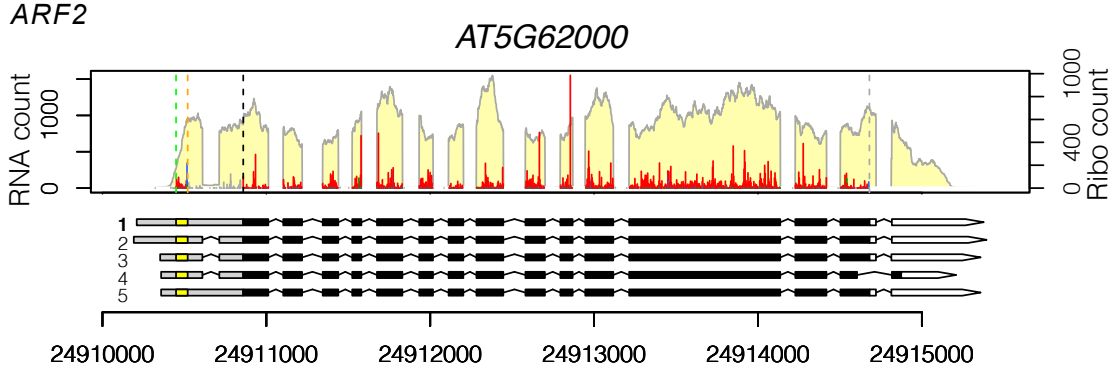


B. *SCARECROW*

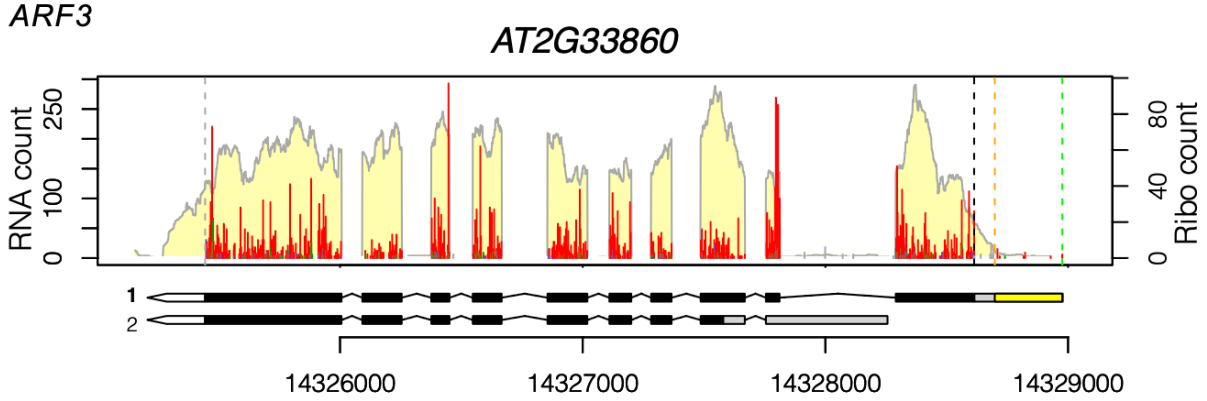




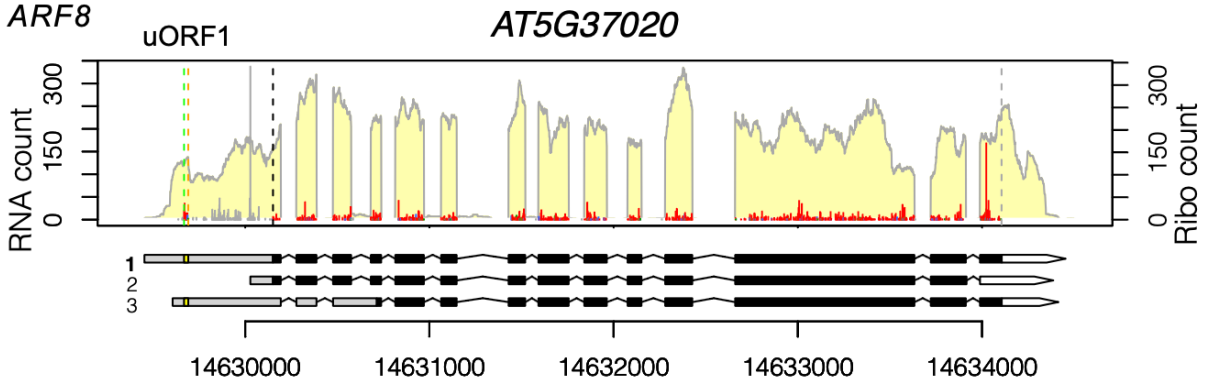
C. *ARF2*

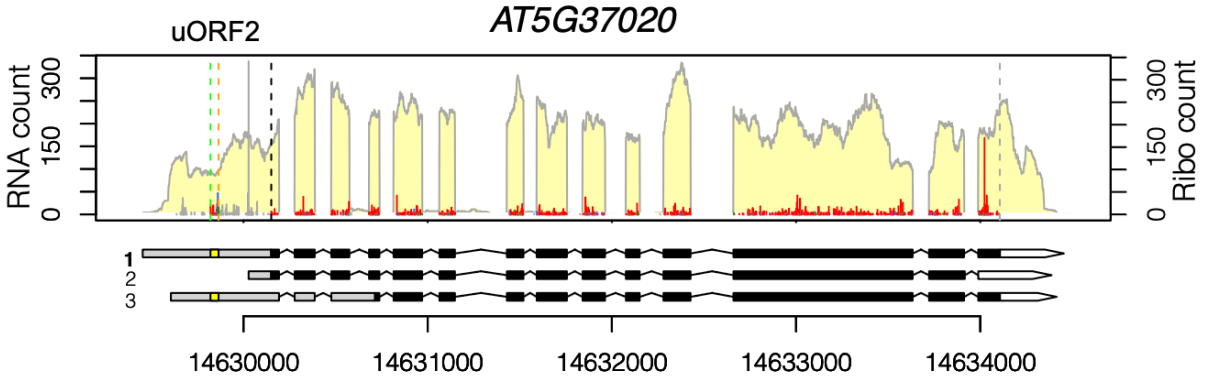


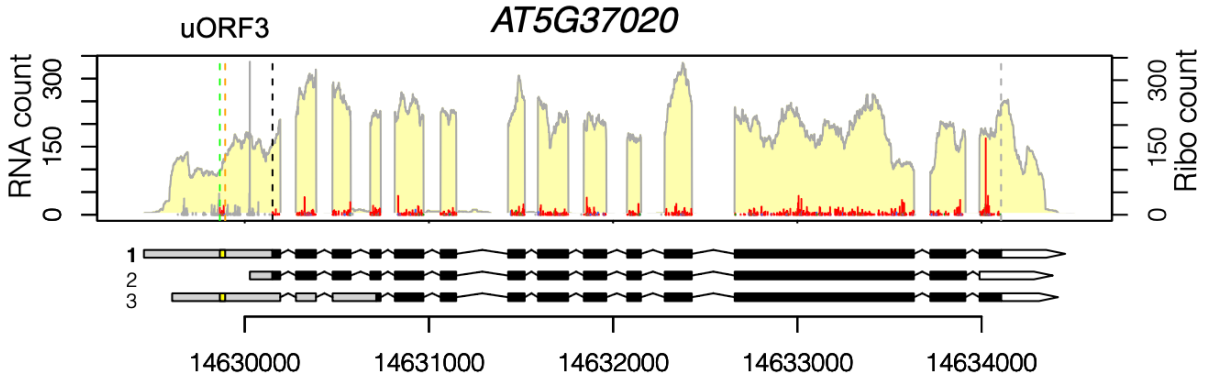
D. *ARF3*


AT2G33860
RNA count
0
100
250
Ribo count
0
40
80
1
2
14326000
14327000
14328000
14329000

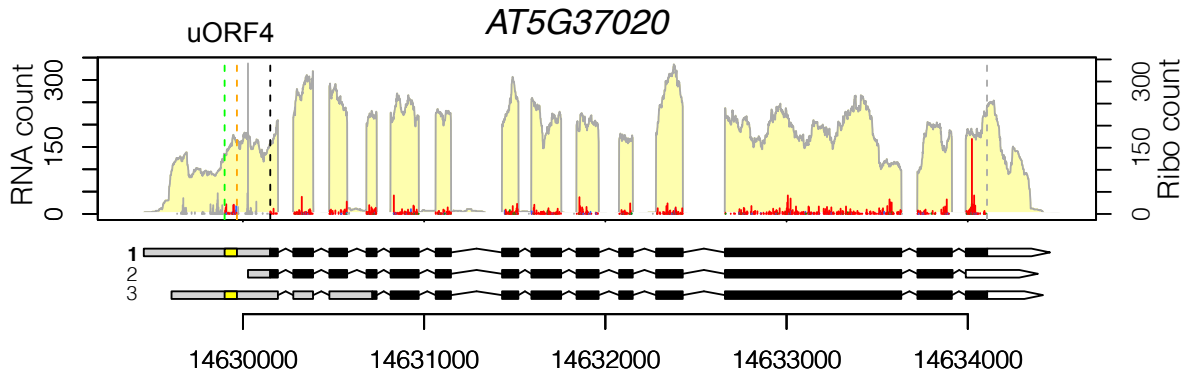
E. *ARF8*


uORF1
AT5G37020
RNA count
0
150
300
Ribo count
0
150
300
1
2
3
14630000
14631000
14632000
14633000
14634000

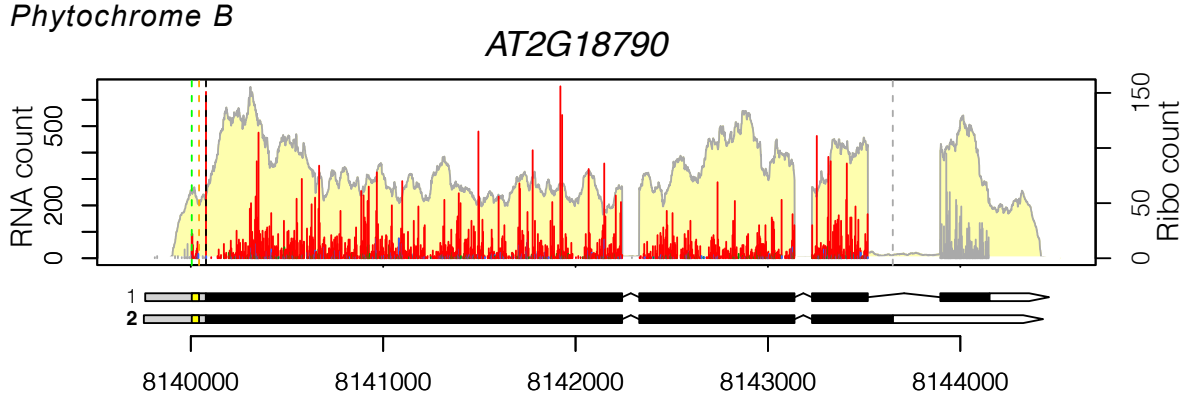

uORF2
AT5G37020
RNA count
0
150
300
Ribo count
0
150
300
1
2
3
14630000
14631000
14632000
14633000
14634000


uORF3
AT5G37020
RNA count
0
150
300
Ribo count
0
150
300
1
2
3
14630000
14631000
14632000
14633000
14634000

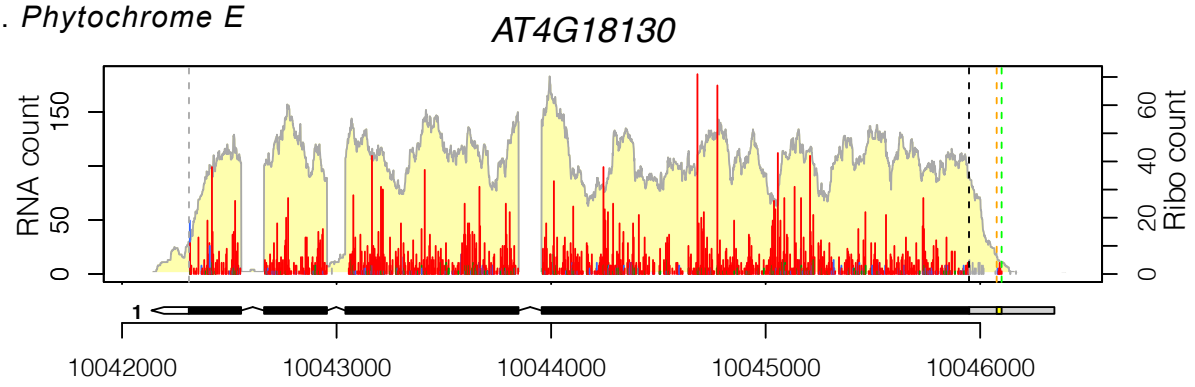
E. *ARF8* (continued)



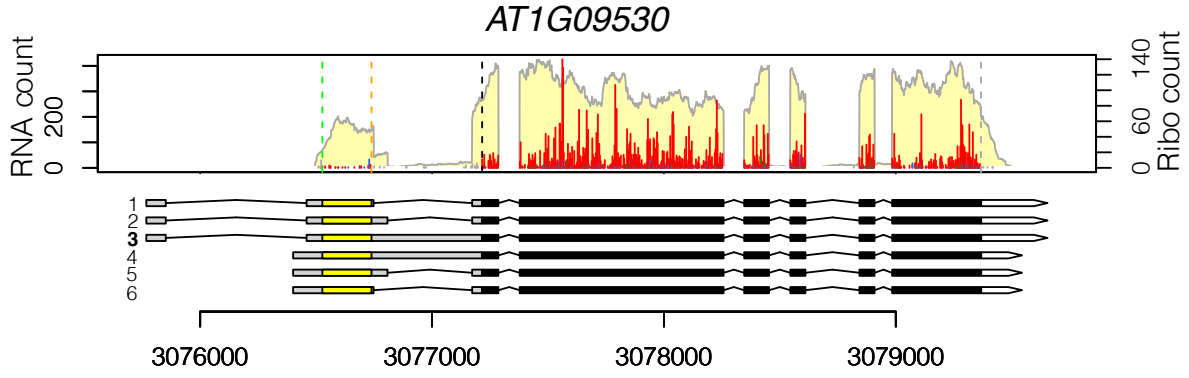
F. *Phytochrome B*



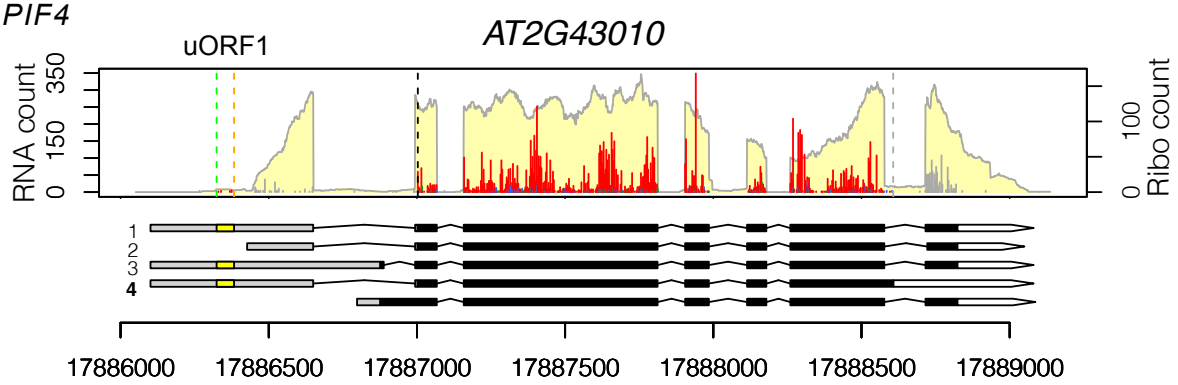
G. *Phytochrome E*

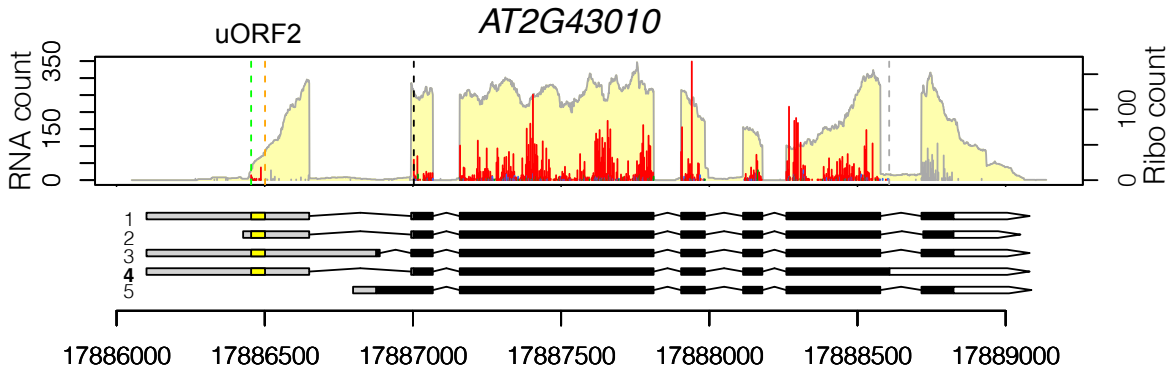


H. *PIF3*

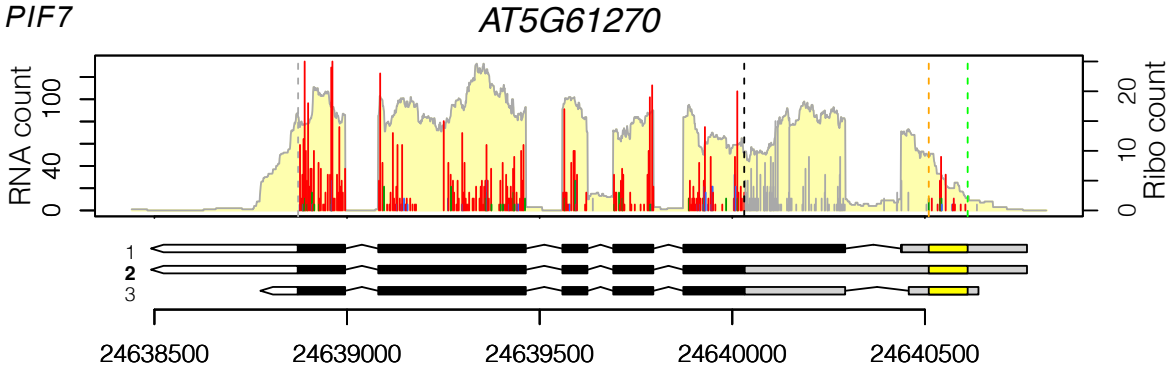


I. *PIF4*

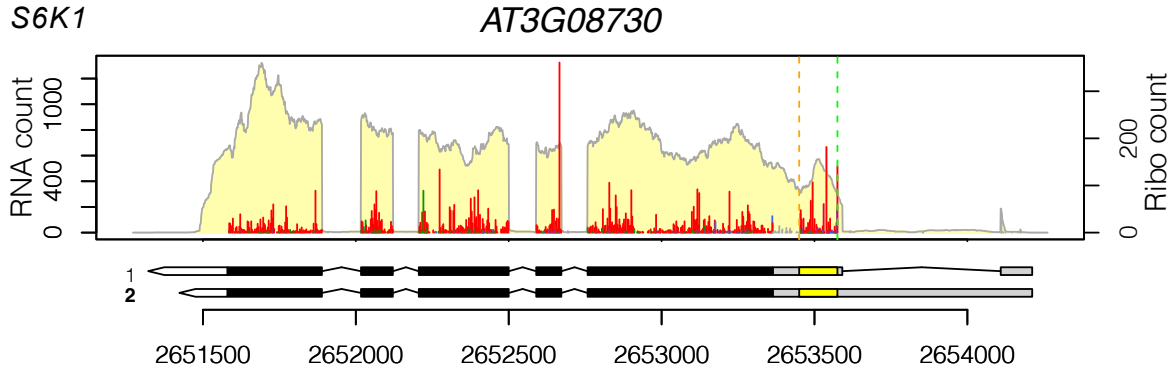




J. *PIF7*



K. *S6K1*



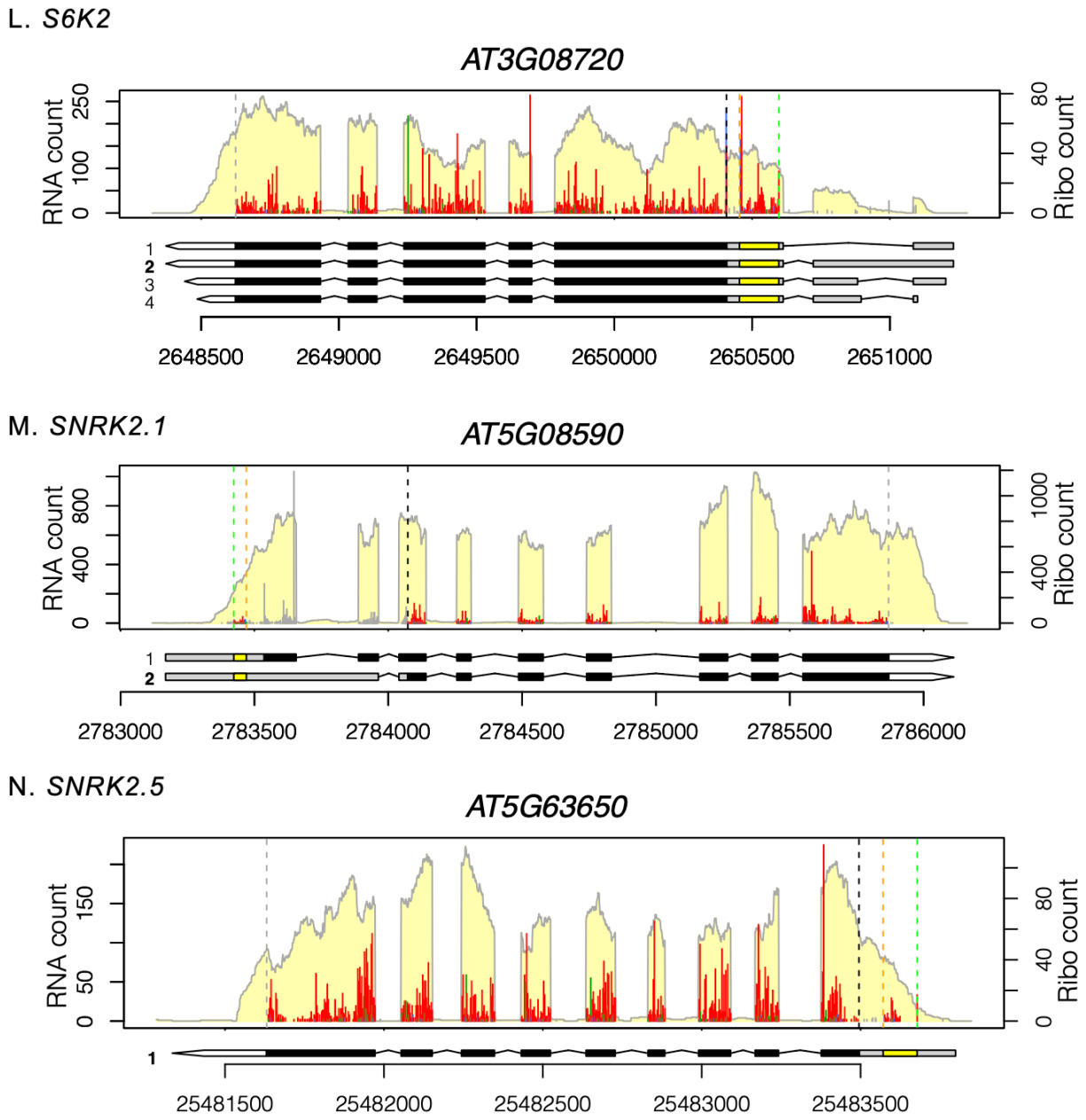


Figure S2. Examples of TuORFs in important regulatory genes. Data are presented as described in the Figure 1F legend. Isoform numbers are shown to the left of the gene models, and the isoform being considered is bolded. For genes that have multiple TuORFs identified (B, E, and I), the 3-nt periodicity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

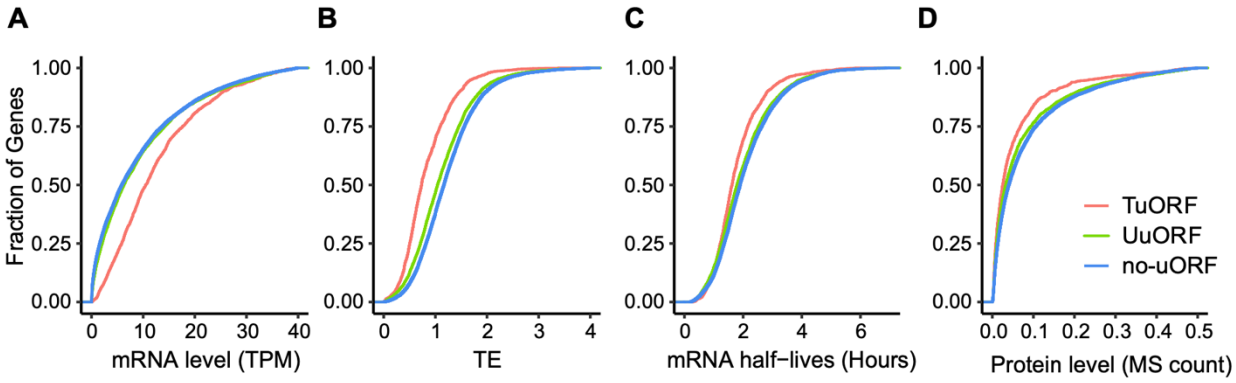


Figure S3. Cumulative plots for the mRNA levels (A), translation efficiency (B), mRNA half-lives (C) and protein abundance (D) of TuORF, UuORF and no-uORF genes.

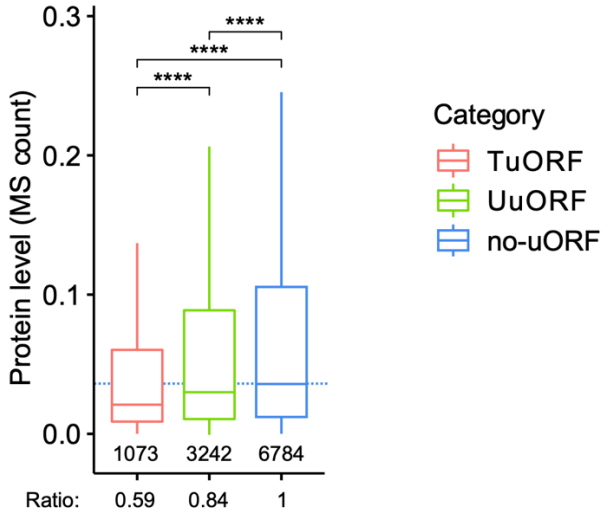


Figure S4. Protein abundance of the uORF-containing genes in Arabidopsis roots. The protein quantification data from Arabidopsis seedling roots were extracted from [2]. Within the boxplot, the dashed line marks the median level of no-uORF genes; the number of genes in each category is listed below the 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 as described in the Figure 2 legend.

Figure S5

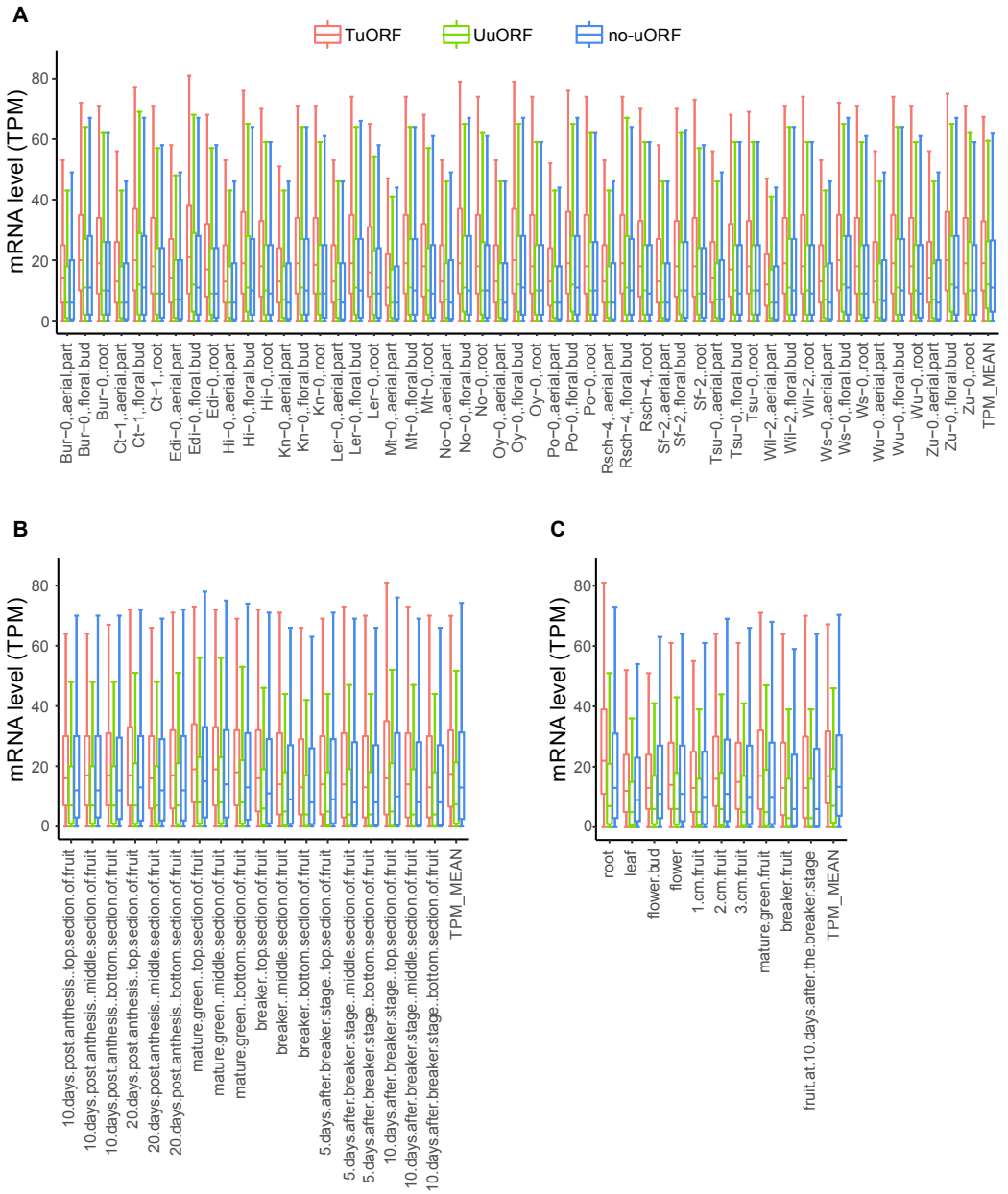


Figure S5. mRNA levels of uORF-containing genes in different tissues, developmental stages and ecotypes in Arabidopsis and tomato.

TuORF, UuORF, and no-uORF genes were compared.

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

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

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

The TPM values from RNA-seq quantification were extracted from the EMBO-EBI expression atlas [3] (see METHODS).

Figure S6

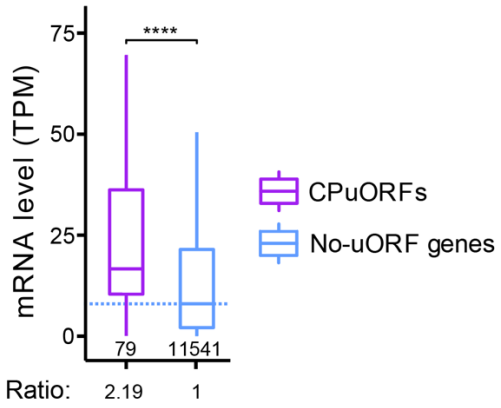


Figure S6. mRNA levels of CPuORF-containing genes compared with no-uORF genes. Within the boxplot, the dashed line marks the median level of no-uORF genes; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of no-uORF genes. The statistical significance was determined as described in the Figure 2 legend.

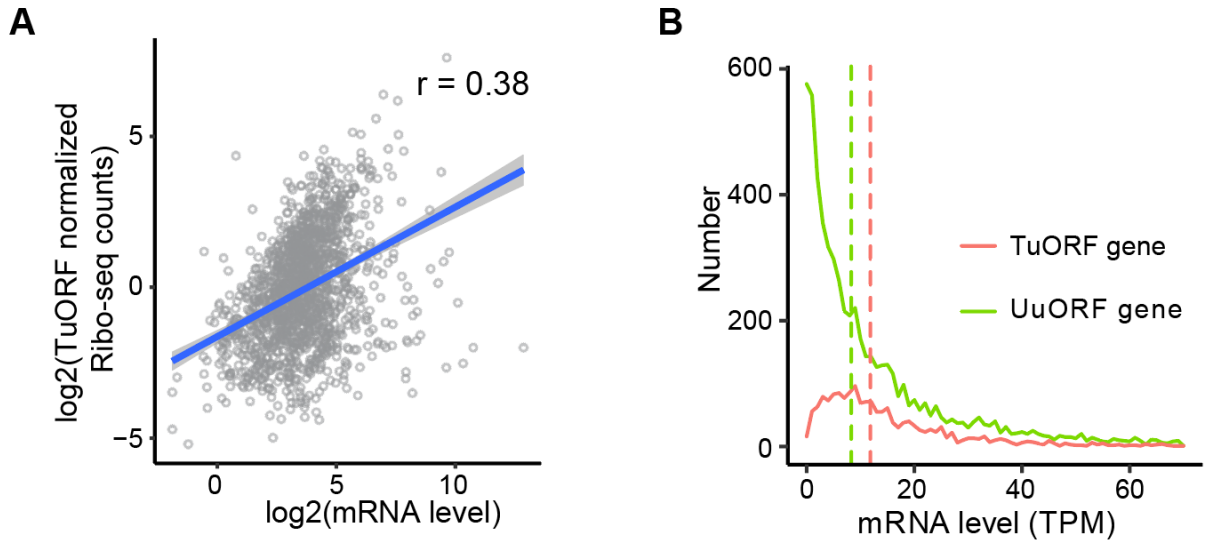


Figure S7. (A) TuORF translation levels are positively correlated with the mRNA levels. TuORF normalized read count is TuORF Ribo-seq count divided by TuORF length. (B) The distributions of the mRNA levels of TuORF and UuORF genes. Vertical dashed lines indicate the median value of each group.

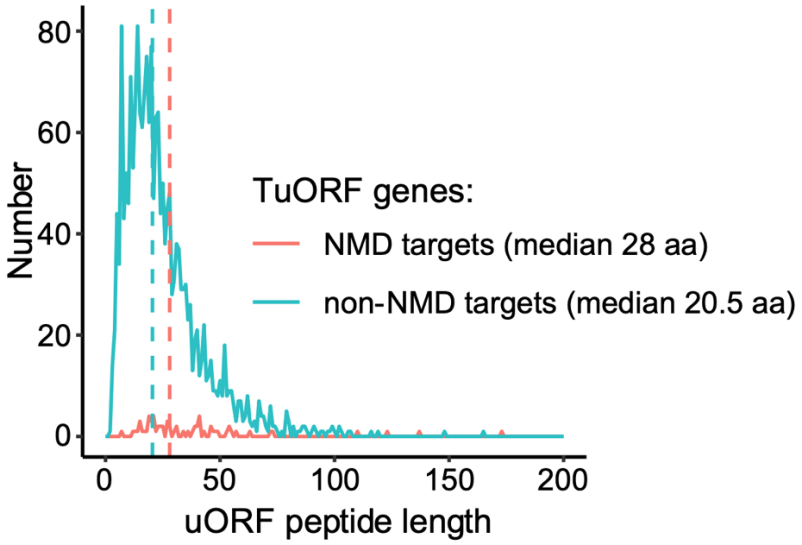


Figure S8. Distributions of uORF peptide lengths in NMD target and non-NMD target TuORF genes. Vertical dashed lines indicate the median value of each group.

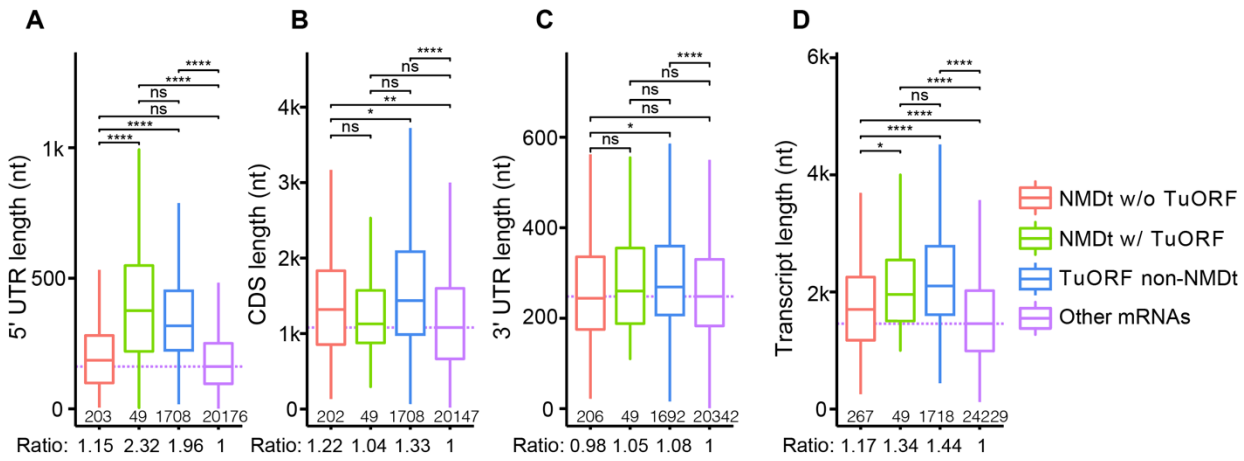


Figure S9. Lengths of UTRs and CDSs for NMD targets and TuORF mRNAs.

Lengths of the 5' UTRs (A), CDSs (B), 3' UTRs (C), and entire transcripts (D) of NMD targets (NMDt) without a TuORF, NMD targets with a TuORF, non-NMD 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.

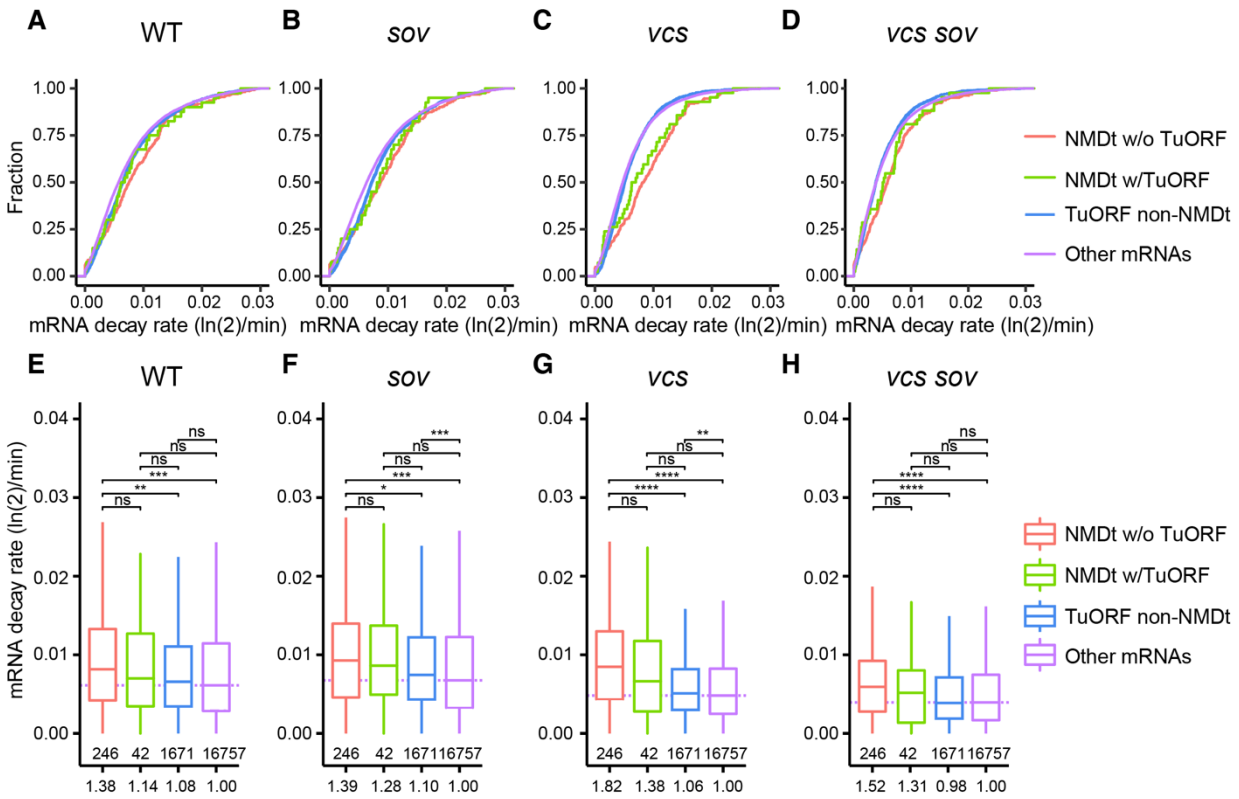


Figure S10. The mRNA decay rates of NMD targets and TuORF-containing mRNAs in different genetic backgrounds. Cumulative plots (A-D) and boxplots (E-H) showing the mRNA decay rates for NMD targets and TuORF mRNAs in wild type and mutants defective in either 5' to 3' decay (*vcs-7*), 3' to 5' decay (*sov*), or both (*vcs sov).* The mRNA decay rates were extracted from [4]. 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.

Reference:

1. Hsu PY, Calviello L, Wu H-YL, Li F-W, Rothfels CJ, Ohler U, et al. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America. 2016;113:E7126–35.

2. Song G, Hsu PY, Walley JW. Assessment and Refinement of Sample Preparation Methods for Deep and Quantitative Plant Proteome Profiling. Proteomics. 2018;18:e1800220.

3. Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S, et al. Expression Atlas update: From tissues to single cells. Nucleic Acids Research. 2020;48:D77–83.

4. Sorenson RS, Deshotel MJ, Johnson K, Adler FR, Sieburth LE. Arabidopsis mRNA decay landscape arises from specialized RNA decay substrates, decapping-mediated feedback, and redundancy. Proceedings of the National Academy of Sciences of the United States of America. 2018;115:E1485–94.