Title: Translation of upstream open reading frames generally does not trigger nonsense-mediated decay in Arabidopsis

Short title: uORF translation does not trigger NMD

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

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) is: Polly Hsu (pollyhsu@msu.edu).

Figures 1-5

Supplementary Figures 1-10

Supplementary Tables 1-4

ABSTRACT

Upstream ORFs (uORFs) are common regulatory elements in the 5' untranslated regions of eukaryotic mRNAs. 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. Here, we identified actively translated uORFs (TuORFs) in Arabidopsis through ribosome profiling and systematically examined their roles in gene expression. Like animal systems, Arabidopsis TuORFs are associated with 38%, 14%, and 43% reductions in translation efficiency, mRNA half-lives, and protein levels, respectively. However, we found TuORF-containing mRNAs have 51% higher transcript levels, and this phenomenon is persistent in diverse tissues and developmental stages across 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 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. Our results suggest that Arabidopsis TuORFs reduce mRNA stability and translation through mechanisms different from NMD and highlight a fundamental difference in gene regulation mediated by TuORFs in plants and animals.

INTRODUCTION

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

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

uORF translation is also expected to trigger mRNA degradation, presumably through nonsense-mediated decay (NMD). NMD is an evolutionarily conserved regulatory mechanism to eliminate aberrant mRNAs, such as those that contain premature termination codons or unusually long 3’ UTRs (He and Jacobson, 2015; Kurosaki et al., 2019). Two pathways have been established in mammals: the exon junction complex (EJC)-dependent and long 3’ UTR pathways. Normally, EJC and two NMD factors (UP-FRAMESHIFT SUPPRESSOR 2 and 3; UPF2 and UPF3) deposited during splicing of mRNAs will be evicted by the ribosome in the first round of translation. In the presence of a premature termination codon, EJC-UPF2-UPF3 complexes downstream of the termination codon remain attached to the mRNA. This allows UPF1, which is part of the surveillance complex recruited at translation termination, to interact with UPF2 and UPF3 in proximity. The interaction of UPF1, UPF2, and UPF3 activates the SUPPRESSOR WITH MORPHOLOGICAL EFFECT ON GENITALIA1 (SMG1) kinase, which

82 phosphorylates UPF1. This establishes the first phase of NMD, substrate recognition, and
83 activates downstream signaling and eventually the second phase of NMD, substrate
84 degradation, which involves SMG5, SMG6, and SMG7. Phosphorylated UPF1 also inhibits
85 translation initiation (Isken et al., 2008). Alternatively, unusually long 3' UTRs, which result in
86 poor translation termination, are believed to promote NMD through phosphorylated UPF1
87 independent of EJC (Kurosaki et al., 2014).
88 The plant NMD pathway is best characterized in *Arabidopsis thaliana*. Arabidopsis
89 possesses homologous NMD factors involved in the first phase, including UPF1, UPF2, and
90 UPF3, and in the second phase, SMG7 (Raxwal and Riha, 2016). However, Arabidopsis is
91 missing the homologs of animal SMG1, SMG5, and SMG6 (Lloyd and Davies, 2013). The
92 absence of key NMD components in plants suggests significant variations in the NMD
93 mechanisms between plants and animals. It remains an open question how the differences in
94 NMD components affect substrate recognition and degradation in plants.
95 In Arabidopsis, putative NMD targets have been identified based on transcripts that
96 overly accumulate in NMD-defective mutants (Kurihara et al., 2009; Kalyna et al., 2012;
97 Drechsel et al., 2013; Degtiar et al., 2015). However, there is a concern that the upregulation of
98 some transcripts in NMD mutants might be caused by secondary effects of NMD (Raxwal et al.,
99 2020). For example, NMD mutants affect numerous aspects of plant physiology, and some
100 alleles cause severe growth retardation and lethality due to constitutive immune responses
101 (Jeong et al., 2011; Rayson et al., 2012; Riehs-Kearnan et al., 2012; Filichkin et al., 2015;
102 Merchante et al., 2015; Chiam et al., 2019; Yoine et al., 2006). Moreover, UPF1 has been
103 reported to have multiple functions in plants, such as alternative splicing regulation and
104 translational control, in addition to NMD (Raxwal et al., 2020). 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 (Riehs-
107 Kearnan et al., 2012; Raxwal et al., 2020). The 333 transcripts upregulated in both *upf1 pad4*
108 and *smg7 pad4* were defined as high-confidence NMD targets. It was shown these NMD targets
109 are translationally regulated by UPF1 and that their turnover requires decapping and 5' to 3'
110 decay (Raxwal et al., 2020).
111 As uORF translation could mimic a transcript with an unusually long 3' UTR, uORF-
112 containing mRNAs are predicted to be targeted by NMD (Kertész et al., 2006; Kurosaki et al.,
113 2019). Consistent with this expectation, genome-wide studies in zebrafish, mouse, and human
114 have found that transcripts with translated uORFs have lower mRNA levels globally (Chew et
115 al., 2016; Johnstone et al., 2016). Moreover, transcripts containing translated uORFs are

116 upregulated in knockdown of NMD factors in mouse embryonic stem cells and human cell lines
117 (Hurt et al., 2013; Baird et al., 2018), supporting the idea that these transcripts are NMD targets.
118 In plants, whether uORF translation triggers NMD is still under debate. Two genome-
119 wide studies suggest that uORF-containing mRNAs are targets of NMD, as transcripts
120 possessing predicted uORFs are enriched in the genes upregulated in Arabidopsis *upf1-1* and
121 *upf3-1* mutants (Kurihara et al., 2009; Drechsel et al., 2013). In contrast, three studies
122 suggested that NMD only controls a small fraction of uORF-containing mRNAs in plants, such
123 as those containing uORFs with > 35 codons (~2% of uORF-containing genes), or those
124 responsive to certain environmental stimuli, such as polyamine or blue light (Nyikó et al., 2009;
125 Uchiyama-Kadokura et al., 2014; Kurihara et al., 2018). Moreover, evidence from several gene-
126 specific or genome-wide studies suggests that uORFs do not trigger NMD. First, the mRNA
127 levels of several uORF-containing transcripts, including *NIP5;1* and *AtMHX*, are not increased in
128 NMD mutants (Saul et al., 2009; Tanaka et al., 2016). Second, a genome-wide study reported
129 that genes with increasing numbers of uORFs did not show lower transcript abundances in
130 Arabidopsis and tomato (Li and Liu, 2020). Third, genes with predicted uORFs are not over-
131 represented in the consensus NMD target list identified from three NMD mutants, including
132 *upf1-5*, *upf3-1*, and *smg7-1* (Rayson et al., 2012). It was also reported that 47 mRNAs that
133 contain Conserved Peptide uORFs (CPuORFs) behave differently from NMD targets prior to
134 their degradation mediated by XRN4 (Nagarajan et al., 2019). It is unknown whether some
135 uORF-containing mRNAs escape NMD or whether uORF-containing mRNAs are generally not
136 targets of NMD in plants. It has been concluded that the mRNA features studied so far are
137 insufficient to infer NMD targetability in plants (Ohtani and Wachter, 2019). Overall, it remains
138 unclear how translated uORFs affect mRNA stability and what the role of NMD in this process is
139 in plants.
140 One limitation of the previous studies is that predicted uORFs, which include both
141 translated and untranslated uORFs, were commonly included in the analyses. Although many
142 uORFs have been predicted based on nucleotide sequences (Zhang et al., 2021), determining
143 which uORFs are actively translated still requires experimental evidence. Recently, deep
144 sequencing of ribosome footprints – known as ribosome profiling or Ribo-seq – has been
145 exploited to study genome-wide mRNA translation (Ingolia et al., 2009; Brar and Weissman,
146 2015). Several Ribo-seq studies in Arabidopsis have deployed different strategies to identify
147 translated uORFs (TuORFs) (Liu et al., 2013; Juntawong et al., 2014; Hsu et al., 2016; Hu et al.,
148 2016; Bazin et al., 2017; Kurihara et al., 2020). Among the parameters used, 3-nucleotide (3-nt)
149 periodicity, which corresponds to translating ribosomes deciphering 3 nts per codon, is

150 considered the most stringent criterion to identify high-confidence translation events (Ingolia et
151 al., 2009; Calviello et al., 2016; Xu et al., 2018; Harnett et al., 2021). Our previous work and a
152 recent study using RiboTaper (Calviello et al., 2016), which computes the statistical significance
153 of 3-nt periodicity, uncovered 187–1378 AUG-initiated TuORFs in Arabidopsis (Hsu et al., 2016;
154 Kurihara et al., 2020).

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

166 **RESULTS**

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

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

176 In addition to high correlations among biological replicates (**Figure S1A**), our new Ribo-
177 seq data displayed characteristics expected for high-quality data, including strong 3-nt
178 periodicity (92% in-frame reads), high enrichment for coding sequences (CDSs), and
179 characteristic ribosome footprint lengths (**Figure 1A-C**). The improved coverage is evident when
180 examining the number of unique ribosome footprints detected globally. At a sequencing depth of
181 20 million reads from randomly selected reads, the number of unique ribosome footprints in our
182 new data is 4.01-fold higher than that in our previous dataset (**Figure 1D**). The coverage
183 improvement is also evident in individual transcript profiles within both uORF and mORF
184 regions. For example, a TuORF in *REPRESSOR OF GA1* (*RGA1*) was identified in our current
185 data but missed in our previous study despite the mRNA levels being similar (**Figure S1B**).

186 The high-coverage data increased the number of TuORFs uncovered. In total, 2093
187 TuORFs were identified in 1768 genes based on significant 3-nt periodicity using RiboTaper
188 (with the default cutoff of RiboTaper, Multitaper F-test, $p < 0.05$) and Araport11 annotation
189 (**Figure 1E** and **Table S1A**). This result is in strong contrast to the relatively few TuORFs
190 identified in Arabidopsis seedling shoots (93 TuORFs) and roots (138 TuORFs) from our
191 previous data obtained using plants grown under similar conditions and using the same
192 statistical analysis and annotations (**Figure 1E** and **Table S1B-C**). The new data also revealed
193 translation from a higher number of Conserved Peptide uORFs (CPuORFs) and annotated
194 coding sequences (mORFs of protein-coding genes) (**Figure 1E** and **Table S1D-F**). We
195 manually plotted all TuORF genes to confirm strong 3-nt periodicity present in the expected
196 uORF regions using RiboPlotR (Wu and Hsu, 2021) (examples shown in **Figure 1F-G, Figure**
197 **S2**). Taken together, these results show that our enhanced-coverage Ribo-seq data
198 dramatically improved the identification of TuORFs.

199 Previous analyses of predicted uORFs found that transcription factors and protein
200 kinases are overrepresented among uORF-containing genes (Von Arnim et al., 2014; Kim et al.,
201 2007). Consistent with these analyses, functions related to transcription factors and protein
202 kinases/phosphatases were identified as being enriched in our Gene Ontology (GO) analysis of
203 TuORF-containing genes (**Table S2**). We found that many TuORF-containing genes encode
204 extensively studied regulators of plant growth and environmental responses. For example, this
205 list included well-known transcription factors, such as those involved in hormone signaling
206 (RGA1 and AUXIN RESPONSE FACTOR2/3/8), root development (SCARECROW), the
207 circadian clock (TIMING OF CAB EXPRESSION1), and light signaling (PHYTOCHROME
208 INTERACTING FACTOR3/4/7) (Silverstone et al., 1998; Strayer et al., 2000; Di Laurenzio et al.,
209 1996; Ellis et al., 2005; Nagpal et al., 2005; Ni et al., 1998; Huq and Quail, 2002; Leivar et al.,
210 2008). The list also included well-characterized protein kinases, such as Phytochromes B and
211 E, ribosomal-protein S6 kinases (S6K1/2), and SNF1-related protein kinase 2 (SNRK2.1/2.5)
212 (**Figure S1B, S2A-N**) (Sharrock and Quail, 1989; Devlin et al., 1998; Zhang et al., 1994;
213 Mizoguchi et al., 1995; Park et al., 1993; Boudsocq et al., 2004). These results support the
214 notion that TuORFs control important cellular regulators in plants.
215
216 **TuORF-containing genes have higher mRNA abundance but are associated with lower**
217 **translation efficiency, shorter mRNA half-lives, and lower protein abundance**
218 To understand how uORF translation affects gene expression, we compared mRNA
219 abundance and translation efficiency of TuORF-containing genes (hereafter, TuORF genes),
220 genes containing sequence-predicted uORFs but undetected by RiboTaper (hereafter, UuORF
221 genes), and genes containing an annotated 5' UTR but no potential uORFs (hereafter, no-uORF
222 genes). Note that some UuORFs may be translated but missed in our analysis because their 3-
223 nt periodicity is insufficient to pass the statistical test: these include extremely short uORFs,
224 those translated at a low frequency, and those overlapping with mORF or other uORFs and thus
225 disrupting the 3-nt periodicity. To study the global effects of uORFs on mRNA stability and
226 protein abundance, we also compared genes with TuORFs, UuORFs, or no-uORFs, using
227 available mRNA half-life data determined by metabolic labeling in Arabidopsis seedlings and
228 two quantitative proteomics datasets from shoots and roots of Arabidopsis seedlings (Szabo et
229 al., 2020; Song et al., 2018). The growth conditions and source of the data are summarized in
230 **Table S3**. Below we presented the data in cumulative plots to show the trends of different gene
231 groups and in box plots to show the statistical results (**Figure 2A-H and S3**).

232 Studies in vertebrates have shown that uORF translation is associated with lower
233 steady-state mRNA abundance, consistent with the expectation that TuORF mRNAs are
234 regulated by NMD (Chew et al., 2016; Johnstone et al., 2016). Surprisingly, we found that
235 TuORF genes in Arabidopsis are associated with higher steady-state mRNA levels than those
236 of UuORF and no-uORF genes (**Figure 2A, 2E**), despite having shorter mRNA half-lives as
237 expected (**Figure 2C, 2G**). The median mRNA level of TuORF genes is 51% higher than that of
238 no-uORF genes (**Figure 2E**). The higher steady-state mRNA levels and shorter mRNA half-lives
239 together suggest that TuORF genes are highly transcribed. To examine whether the high
240 transcript levels of TuORF genes are conserved in other plant species, we reanalyzed our
241 previous tomato Ribo-seq data from seedling roots (Wu et al., 2019) and found the TuORF
242 genes in tomato also have higher steady-state mRNA levels relative to no-uORF genes (**Figure**
243 **S4A**). Given that Arabidopsis and tomato diverged more than 100 million years ago (Clarke et
244 al., 2011), our results suggest that TuORF-associated higher mRNA abundance could be
245 prevalent in plants. To investigate whether our observation is specific to seedlings or common to
246 different developmental stages/tissues/genetic backgrounds in Arabidopsis and tomato, we
247 examined the mRNA levels of the uORF-containing genes in 128 independent RNA-seq studies
248 collected by the EMBO-EBI expression atlas (Papatheodorou et al., 2020). Consistent with our
249 Arabidopsis seedlings and tomato seedling root data, we found the TuORF genes are
250 associated with higher mRNA levels in diverse tissues, developmental stages, and ecotypes in
251 Arabidopsis and tomato (**Figure 2I and S5A-C**), supporting the hypothesis that TuORF genes
252 are preferentially expressed at high levels in these plants under a wide range of sampling
253 conditions.
254 CPuORFs are uORFs that encode evolutionarily conserved peptides, and they were
255 identified independently from our approach by evolutionary conservation (Hayden and
256 Jorgensen, 2007; Jorgensen and Dorantes-Acosta, 2012; Takahashi et al., 2012; Vaughn et al.,
257 2012; Van Der Horst et al., 2019). Many CPuORFs were also identified as TuORFs in our study
258 (**Figure 1E, 1G**). We found that the median mRNA abundance of CPuORF genes is 2.2-fold of
259 that of no-uORF genes (**Figure S6**), consistent with our observations that genes with TuORFs
260 are expressed at higher levels.
261 Next, we examined the translation efficiency, mRNA half-lives, and protein levels of
262 TuORF, UuORF, and no-uORF genes. Consistent with the animal studies, we found
263 Arabidopsis TuORF genes are associated with lower translation efficiency, shorter mRNA half-
264 lives, and less protein accumulation than UuORF and no-uORF genes (**Figure 2B-D, 2F-H, S3,**
265 **S4B**). Interestingly, even though globally TuORF genes start with 51% more mRNAs than no-

266 uORF genes (**Figure 2E**), they produce 41%–43% less protein (**Figure 2H, S3**), on average,
267 likely due to the combined effects of a 38% reduction in translation efficiency (**Figure 2F**) and a
268 14% reduction in mRNA half-lives (**Figure 2G**). These observations suggest that the global
269 repression of TuORF gene protein synthesis is determined by multilayer post-transcriptional
270 regulation. Notably, for the four parameters examined here, UuORF genes consistently have
271 levels between those of TuORF genes and no-uORF genes: they show a 4% increase in mRNA
272 levels and 11%, 6%, and 21% reductions in translation efficiency, mRNA half-lives, and protein
273 abundance, respectively (**Figure 2A-H**). These results indicate that UuORFs have smaller
274 effects on gene expression than TuORFs and highlight the importance of distinguishing
275 TuORFs and UuORFs in genome-wide analysis.
276
277 **TuORF genes are larger and likely evolutionarily older than other genes**
278 We next investigated whether there are general structural differences among TuORF,
279 UuORF, and no-uORF mRNAs. Considering the directionality of ribosome scanning, we
280 hypothesized TuORFs might be associated with longer 5' UTRs. Indeed, we found that the
281 median 5' UTR length of TuORF mRNAs is 2.5-fold of that of no-uORF mRNAs (**Figure S7A**),
282 which is consistent with a previous report that uORFs in longer 5' UTRs are more likely to be
283 translated (Liu et al., 2013). However, unexpectedly, we found that TuORF mRNAs also have
284 longer CDSs and 3'UTRs, as well as longer transcript lengths, than UuORF mRNAs (**Figure**
285 **S7B-D**). No-uORF mRNAs are the shortest in all categories examined (**Figure S7A-D**). These
286 results support the hypothesis that TuORFs are more likely to be found in longer 5' UTRs and
287 reveal that TuORF mRNAs tend to encode larger proteins (**Figure S7B**). The enrichment of
288 TuORF genes for transcription factors, protein kinases, and related functions in GO term
289 analysis (**Table S2**) also implies TuORF genes tend to encode proteins with conserved
290 domains. It is known that evolutionarily older genes are generally longer, expressed at higher
291 levels, and have protein sequences that are under stronger purifying selection (Wolfa et al.,
292 2009; Cui et al., 2015). Our observations that TuORF genes possess many features of
293 evolutionarily older genes imply that older genes are more likely to be regulated through mRNA
294 translation and degradation mediated by TuORFs.
295
296 **TuORF mRNAs are not upregulated in the NMD mutants**
297 The established paradigm in vertebrates is that uORF translation triggers NMD and thus
298 results in low steady-state mRNA levels (Chew et al., 2016; Johnstone et al., 2016; Hurt et al.,
299 2013; Baird et al., 2018). Despite having higher steady-state mRNA levels (**Figure 2A and 2E**),

300 TuORF genes in Arabidopsis also have shorter mRNA half-lives (**Figure 2C and 2G**).
301 Therefore, we investigated whether TuORF-containing transcripts in Arabidopsis are still subject
302 to NMD.
303 If TuORFs trigger NMD, one would expect TuORF mRNAs to be among the known NMD
304 targets and their mRNA levels to be increased in NMD-deficient mutants. We first examined
305 whether TuORF mRNAs were reported in the high-confidence NMD target list (Raxwal et al.,
306 2020). Among the 333 NMD targets that were upregulated in both *upf1 pad4* and *smg7 pad4*
307 adult leaves (Raxwal et al., 2020), 248 genes with an annotated 5' UTR were expressed in our
308 data (**Table S4A**). However, only 49 of 1748 TuORF genes were identified in the high-
309 confidence NMD target list (**Table S4B**). The 49 genes represent 2.8% of TuORF-containing
310 genes and 20% of the NMD targets. These results suggest that TuORFs generally do not trigger
311 NMD; rather, NMD targets are more likely to be regulated by TuORFs than other genes (Chi-
312 square test with Yates' correction: p-value = 7.555e-10). One caveat is that the poor overlap
313 between the two gene lists could be due to differences in growth conditions, developmental
314 stages, and sample handling in different labs (**Table S3**).
315 To further investigate if TuORF mRNAs are NMD targets, we examined whether they are
316 upregulated in NMD-deficient mutants by analyzing four published RNA-seq datasets for NMD
317 mutants (Drechsel et al., 2013; Merchante et al., 2015; Raxwal et al., 2020; Gloggnitzer et al.,
318 2014). Although these datasets were collected from different NMD mutants and at various
319 developmental stages (**Table S3**), the high-confidence NMD targets are consistently and
320 significantly upregulated in all the NMD mutants examined compared with the control plants
321 (i.e., 3.05-, 1.54-, 4.98-, and 2.65-fold for NMD targets without TuORFs; 3.04-, 2.23-, 5.04-, and
322 2.66-fold for NMD targets with TuORFs) (**Figure 3A-D**). In contrast, TuORF mRNA levels show
323 no or little change in the NMD mutants (i.e., 1.01-, 0.98-, 1.10- and 1.10-fold). Thus, only the
324 NMD targets, but not TuORF mRNAs, are dramatically increased in the NMD mutants,
325 supporting the idea that uORF translation in Arabidopsis generally does not trigger NMD.
326 Given different uORFs are translated at various levels (See **Figure 1F-G, S1B, S2A-N**
327 for examples), it is possible only the highly repressive TuORFs efficiently trigger NMD. To test
328 this hypothesis, we split the TuORF genes into two groups based on their uORF relative
329 translation efficiency (RTE) (i.e., TE of an uORF is normalized to TE of its mORF). However,
330 even the higher TuORF RTE group shows no or little effects on the mRNA accumulation in the
331 NMD mutants (i.e., 1.13-, 0.99-, 1.15- and 1.11-fold) (**Figure 3E-H**). We also tested whether
332 having more TuORFs in a transcript efficiently triggers NMD by splitting the TuORF genes into
333 two groups based on the number of TuORFs detected. We found that even transcripts with 2~5

334 TuORFs do not lead to significant mRNA increase in the NMD mutants, either (i.e., 0.93-, 0.98-,
335 0.99- and 1.09-fold)(**Figure 3I-L**). Together, these results support the idea that uORF translation
336 in Arabidopsis generally does not induce NMD.
337
338 **Lengths of the TuORF and 3' UTR alone are insufficient to predict NMD targetability**
339 A previous study suggested that uORFs that encode a peptide longer than 35 amino
340 acids are likely to trigger NMD in Arabidopsis (Nyikó et al., 2009). To test whether TuORF
341 length is associated with NMD targetability, we analyzed TuORF lengths in the NMD targets and
342 non-NMD targets. We found that the NMD target TuORF genes tend to have longer uORF
343 peptides than non-NMD target TuORF genes (median length 28 amino acids vs. 20.5 amino
344 acids, p = 9.874e-07, Wilcoxon signed-rank test). However, the distributions of the uORF
345 peptide lengths in NMD targets and non-NMD targets largely overlap (**Figure S8**), indicating
346 that TuORF length alone is insufficient to predict whether a TuORF triggers NMD of a transcript.
347 We also examined whether the high-confidence NMD targets have longer 3' UTRs, a
348 feature that potentially triggers NMD. We found that the NMD targets do not have significantly
349 longer 3' UTRs, regardless of whether they contain TuORFs (**Figure S9**). Taken together, our
350 results are consistent with the notion that no definite RNA features studied to date are sufficient
351 to predict NMD targets in plants (Ohtani and Wachter, 2019).
352
353 **TuORF mRNAs and NMD targets have distinct expression patterns**
354 To further characterize the roles of NMD and TuORFs in gene expression regulation, we
355 compared the expression patterns of the NMD targets and TuORF mRNAs (**Figure 4**).
356 Consistent with the expectation that NMD targets are being actively degraded, we found that
357 NMD targets have lower mRNA abundance than other mRNAs; this is in strong contrast to
358 TuORF mRNAs, which have higher mRNA abundance than other mRNAs (**Figure 4A, 4E**).
359 While both NMD targets and TuORF mRNAs are associated with lower translation efficiency,
360 shorter mRNA half-lives, and lower protein abundance than other mRNAs (**Figure 4B-D, 4F-H**),
361 the NMD targets are associated with even lower protein levels than TuORF mRNAs (**Figure 4D,**
362 **4H**).
363 If TuORFs effectively trigger NMD, one would expect TuORF mRNAs to have translation
364 efficiency similar to that of NMD targets with TuORFs, which would inherently be repressed by
365 both TuORFs and NMD. However, we found that NMD targets with TuORFs have significantly
366 lower translation efficiency than either NMD targets without TuORFs or TuORF mRNAs not
367 identified as NMD targets (**Figure 4B, 4F**). This additive repression of translation efficiency in

368 NMD targets with TuORFs suggests that NMD and TuORFs suppress translation through
369 separate mechanisms. This result also supports that TuORFs do not effectively trigger NMD.
370 Our observation is consistent with the expectation that TuORFs prevent the scanning pre-
371 initiation complex from reaching the mORF, while NMD suppresses translation through UPF1
372 and/or other translational repressors (Isken et al., 2008; Raxwal et al., 2020; Zinshteyn et al.,
373 2021).
374 Together, the differences in the expression patterns between TuORF mRNAs and NMD
375 targets further support the idea that TuORFs are not an mRNA feature that efficiently triggers
376 NMD in Arabidopsis.
377
378 **TuORF mRNAs and NMD targets are degraded through different mRNA decay**
379 **mechanisms**
380 We next investigated the mechanisms underlying the decay of TuORF mRNAs and NMD
381 targets. mRNA decay can occur in the 5' to 3' or 3' to 5' direction, which is carried out by
382 different machinery (Sieburth and Vincent, 2018). A genome-wide survey of mRNA decay rates
383 in wild type and mutants defective in either 5' to 3' decay or 3' to 5' decay, or both, has been
384 conducted in Arabidopsis (Sorenson et al., 2018). In that survey, the *vcs-7* mutant, a null mutant
385 of *VARICOSE*, which encodes a scaffold protein for the mRNA decapping complex, was used to
386 examine the 5' to 3' mRNA decay; the *sov* mutant, defective in *SUPPRESSOR OF VARICOSE*,
387 which encodes a ribonuclease II, was used to examine 3' to 5' decay; and the *vcs/sov* double
388 mutant was used to examine decay affected in both directions (Deyholos et al., 2003; Goeres et
389 al., 2007; Xu et al., 2006; Zhang et al., 2010). We exploited this published dataset (Sorenson et
390 al., 2018) to study the genetic requirements for the mRNA decay of TuORF mRNAs, NMD
391 targets, TuORF-containing NMD targets, and other mRNAs.
392 We first compared the decay rates of each gene group across the four genetic
393 backgrounds (**Figure 5**). Similar to mRNAs that are non-NMD targets and lack TuORFs (i.e.,
394 other mRNAs, **Figure 5D, 5H**), TuORF mRNAs have dramatically reduced decay rates in the
395 *vcs* mutant compared to the wild type (**Figure 5A, 5E**), indicating that decapping and 5' to 3'
396 decay are critical for degrading these transcripts. In the *vcs/sov* double mutant, TuORF mRNAs
397 have slightly but significantly lower decay rates than in the *vcs* mutant (**Figure 5A and 5E**),
398 suggesting *SOV* also plays a minor role in degrading TuORF mRNAs.
399 In contrast, the mRNA decay rates of NMD targets without TuORFs are only significantly
400 lower in the *vcs/sov* double mutant (**Figure 5B, 5F**), suggesting that NMD targets can be
401 efficiently degraded by either the 5' to 3' or 3' to 5' pathways. In other words, in the absence of

402 either *SOV* or *VCS*, the remaining pathway could completely take over the decay of NMD
403 targets (**Figure 5B, 5F**). Although there are fewer NMD targets containing TuORFs, which
404 resulted in small sample size and statistical insignificance (**Figure 5C, 5G**), the decay rates in
405 this group of genes in the four genetic backgrounds show trends similar to other NMD targets
406 without TuORFs (**Figure 5B, 5F**).
407 Next, we compared the decay rates of the four gene groups within each genetic
408 background (**Figure S10)**. In all four genetic backgrounds, NMD targets with or without TuORFs
409 consistently have higher decay rates than TuORF mRNAs and other mRNAs (**Figure S10A-H**).
410 This pattern is still persistent in the *vcs/sov* double mutant **(Figure S10D, S10H)**, suggesting
411 additional decay pathway(s) is involved in degrading the NMD targets in the absence of both
412 VCS and SOV.
413 Taken together, our results suggest that TuORF mRNAs and NMD targets are
414 preferentially degraded through different decay pathways. Furthermore, NMD targets
415 consistently have higher decay rates than TuORF mRNAs and appear to be regulated by
416 additional pathway(s) other than VCS and SOV (see the model in **Figure 5I**).
417
418 **DISCUSSION**
419 In this study, we identified Arabidopsis TuORFs using improved Ribo-seq data and
420 systematically examined their roles in gene expression. While our results support the expected
421 roles of uORFs in repressing mORF translation and mRNA stability, our study reveals distinct
422 features and regulation of uORFs in plants compared with animals.
423 Our observations that plant TuORF genes are associated with higher expression levels
424 (**Figure 2**), encode larger proteins (**Figure S7B**), and are enriched for evolutionarily conserved
425 domains (**Table S2**) raised an interesting question as to whether older genes are more likely to
426 be controlled at the protein levels through mRNA degradation and translation via uORFs. Higher
427 mRNA abundance may allow TuORF mRNAs to be highly translated under certain conditions
428 that require a high level of specific proteins. For example, it has been suggested that
429 derepression of downstream translation is a general mechanism of uORF-mediated expression
430 under stress (Andreev et al., 2018). Thus, TuORFs could enable fast and dynamic regulation of
431 mORF translation. Future work monitoring the translation of TuORFs and their mORFs under
432 different conditions will help decipher the regulation and physiological roles of specific TuORFs.
433 If TuORFs trigger NMD in Arabidopsis, we would expect that TuORF mRNAs behave
434 similarly to NMD targets. In this report, we presented six observations suggesting that
435 Arabidopsis TuORF mRNAs and NMD targets are two distinct groups of transcripts: (1) NMD

436 targets, but not TuORF mRNAs, are upregulated in various NMD mutants (**Figure 3**). (2)
437 TuORF mRNAs have significantly higher transcript levels, while the NMD targets have lower
438 transcript levels than other genes (**Figure 4A and 4E**). (3) NMD and TuORFs additively repress
439 translation, suggesting they repress translation through separate mechanisms (**Figure 4B, 4F**).
440 (4) While NMD targets are efficiently degraded through both 5' to 3' and 3' to 5' decay, TuORF
441 mRNAs, like other genes, mainly rely on 5' to 3' decay (**Figure 5**). (5) The decay rates of NMD
442 targets are consistently higher than those of TuORF mRNAs (**Figure S10)**. (6) NMD targets also
443 appear to be degraded by additional pathway(s) in the absence of both *VCS* and *SOV* (**Figure**
444 **S10H**). These differences between TuORF mRNAs and NMD targets suggest that NMD is not a
445 major mechanism regulating TuORF mRNAs.
446 The fact that Arabidopsis is missing key NMD factors found in animals, including SMG1,
447 SMG5, and SMG6, suggests significant variations in the NMD mechanisms between plants and
448 animals. In particular, SMG1 is the kinase responsible for UPF1 phosphorylation required for
449 substrate recognition in animal NMD (He and Jacobson, 2015; Kurosaki et al., 2019). The lack
450 of SMG1 and the use of an alternative kinase for Arabidopsis UPF1 phosphorylation (Kerényi et
451 al., 2013) may lead to a difference in substrate recognition: while TuORF mRNAs are
452 recognized as NMD targets in animals, plant TuORF mRNAs are not.
453 Despite the difference in NMD components in plants and animals, our findings that
454 Arabidopsis NMD targets are degraded through both 5' to 3' decay and 3' to 5' decay is
455 consistent with the two-directional decay of NMD targets in animals. Note that plant SOV is the
456 homolog of DIS3L2, the exonuclease responsible for 3' to 5' decay of NMD targets in animals
457 (Chang et al., 2013; Lubas et al., 2013). Two-directional decay could be advantageous to
458 quickly remove aberrant mRNAs from the cytoplasm to prevent further translation and facilitate
459 the recycling of NMD machinery.
460 It is well established that there is an important interplay between translation efficiency
461 and mRNA stability. Two mechanisms have been proposed to explain this phenomenon: the
462 "stalled ribosome-triggered decay model," which predicts that ribosome stalling or slowing
463 promotes mRNA degradation, and the "translation factor-protected model," which predicts that
464 active translation promotes mRNA stability because the translation initiation factors protect the
465 mRNA from the decapping complex (Presnyak et al., 2015; Radhakrishnan et al., 2016;
466 Beelman and Parker, 1994; LaGrandeur and Parker, 1999; Schwartz and Parker, 2000; Chan et
467 al., 2018). Consistent with the first model, several CPuORF-containing mRNAs in Arabidopsis
468 were reported to have strong ribosome stalling patterns, and the endonucleolytic cleavages
469 upstream of the stalled ribosomes were observed (Hou et al., 2016; Yu et al., 2016). While most

470 of the TuORFs we identified did not display clear stalling patterns (e.g., **Figure 1F-G, S1B,**
471 **S2A-N**), it remains possible that ribosomes have lower translation speeds, on average, during
472 uORF translation. Because ribosomes will quickly approach termination after initiation due to the
473 short length of uORFs, and because both initiation and termination are major time-consuming
474 steps in translation (Shah et al., 2013; Lawson et al., 2021), the average translation speed per
475 codon for uORFs is likely to be slower than that for longer ORFs. Alternatively, if cells mainly
476 monitor the translation efficiency of the mORF, then the poorer mORF translation caused by
477 uORFs could explain the faster degradation of TuORF mRNAs predicted by this model. In
478 contrast, since ribosomes actively engage with TuORFs, the “translation factor-protected” model
479 does not fit TuORF mRNAs, especially because their degradation relies on decapping and 5’ to
480 3’ decay in plants. Further investigation into TuORF-mediated mRNA degradation will help
481 elucidate the role of uORFs in regulating mRNA stability.
482

483 **METHODS**

484 **Plant growth conditions and lysate preparation**

485 Arabidopsis Col-0 seeds were surface sterilized with 70% ethanol for 5 min, followed by 33%
486 bleach and 0.03% Tween 20 for 10 min, then rinsed with sterile water 5 times. The seeds were
487 imbibed at 4°C in the dark for 2 days, then grown hydroponically in sterile liquid media (2.15 g/L
488 Murashige and Skoog salt, 1% sucrose, 0.5 g/L MES, pH 5.7) while shaking at 85 rpm under a
489 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
490 7 days. At Zeitgeber time 4 (4 h after lights on), DMSO corresponding to 0.1% of the media
491 volume was added to the media (these were mock samples of our large-scale experiment). After
492 20 and 60 min, three biological replicates (~300 plants per sample) were harvested at each time
493 point and immediately flash frozen in liquid nitrogen.

494

495 Plant lysates were prepared as previously described (Hsu et al., 2016). Briefly, per 0.1 g of
496 ground tissue powder was resuspended in 400 μL of lysis buffer (100 mM Tris-HCl [pH 8], 40
497 mM KCl, 20 mM $MgCl_2$, 2% [v/v] polyoxyethylene [10] tridecyl ether [Sigma, P2393], 1% [w/v]
498 sodium deoxycholate [Sigma, D6750], 1 mM dithiothreitol, 100 μg/mL cycloheximide [Sigma,
499 C4859], 100 μg/mL chloramphenicol [Sigma R4408], and 10 units/mL DNase I [Epicenter,
500 D9905K]). The lysates were spun at 3,000 g for 3 min, and the supernatant was transferred to a
501 new tube and subsequently centrifuged at 20,000 g for 10 min. The supernatant was transferred
502 to a new tube and the RNA concentration was determined with 10x dilutions using the Qubit
503 RNA HS assay (Thermo Fisher Scientific; Q32852). Aliquots of 100 μL and 200 μL of the
504 lysates were made, and they were flash frozen in liquid nitrogen and stored at -80°C until further
505 processing.

506

507 **Ribo-seq library construction**

508 Ribosome footprints were obtained using 200 μL of the lysates described above, and
509 sequencing libraries were constructed according to our previous method (Hsu et al., 2016) with
510 the following modifications: briefly, after RNase I digestion (50 units nuclease per 40 μg of RNA;
511 the nuclease was included in TruSeq Mammalian Ribo Profile Kit, Illumina, RPHMR12126),
512 adding 15 μl of SUPERase-IN (Invitrogen, AM2696), and passing through a size exclusion
513 column (Illustra MicroSpin S-400 HR Columns; GE Healthcare; 27-5140-01), RNA > 17 nt was
514 isolated with the RNA Clean & Concentrator-25 kit (Zymo Research, R1017) and separated on
515 15% urea-TBE gels (Invitrogen, EC68852BOX). Gel slices between 28 and 30 nt were isolated,
516 and the RNAs were purified as previously described. Next, rRNA depletion was performed using

517 RiboZero Plant Leaf kit (Illumina, MRZPL1224) in one quarter of the recommended reaction
518 volume. The major change we made here was to gel-purify the ribosome footprints prior to the
519 rRNA depletion. In our previous protocol, we purified the RNA twice (first >17 nt and then <200
520 nt) to enrich ribosome footprints before the rRNA depletion and gel purification. We reasoned
521 that some footprints might be lost during the two RNA purification steps. In addition, the rRNA
522 depletion step requires a limited amount of input RNA according to the manufacturer's
523 recommendations. Therefore, we expected modifying the order of these steps could maximize
524 the input and minimize the footprint loss during the processes. Ribo-seq libraries were then
525 constructed using the TruSeq Mammalian Ribo Profile Kit (illumina, RPHMR12126) as
526 previously described with nine cycles of PCR amplification. Libraries with equal molarity were
527 pooled and sequenced on a Hi-Seq 4000 sequencer using single-end 50-bp sequencing.
528
529 **RNA-seq library construction**
530 Total RNA greater than 200 nt was purified from 100 μL of the lysates described above using
531 the RNA Clean & Concentrator-25 kit (Zymo Research, R1017) as previously described (Hsu et
532 al., 2016). RNA integrity was evaluated using a Bioanalyzer (Agilent) RNA pico chip, and RNA
533 integrity numbers (RINs) ranging from 7.2 to 7.7 were observed among the samples. A total of 4
534 μg of RNA per sample was subjected to rRNA depletion using the RiboZero Plant Leaf kit
535 (Illumina, MRZPL1224) following the manufacturer's recommendations. Then, 100 ng of rRNA-
536 depleted RNA was fragmented to around 200-nt long based on the RIN reported by the
537 Bioanalyzer, and strand-specific sequencing libraries were made using the NEBNext Ultra II
538 Directional RNA Library Prep Kit (New England Biolabs, E7760S) with eight cycles of
539 amplification. Libraries of equal molarity were pooled and sequenced on a Hi-Seq 4000 using
540 paired-end 100-bp sequencing.
541
542 **Data pre-processing and analysis**
543 Data pre-processing and analysis were performed similarly to that previously described (Hsu et
544 al., 2016), except that the Araport11 annotation was used in this study. Briefly, for Ribo-seq
545 libraries, the adaptor (AGATCGGAAGAGCACACGTCT) was clipped with fastx_clipper (FASTX
546 toolkit v0.0.14) (http://hannonlab.cshl.edu/fastx_tool- kit/). For both RNA-seq and Ribo-seq, we
547 used Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012) to remove
548 rRNA/tRNA/snRNA/snoRNA sequences. Both the RNA-seq and Ribo-seq reads were mapped
549 to the transcriptome with STAR aligner (Dobin et al., 2013) (RNA-seq parameters: --
550 outFilterMismatchNmax 2 --outFilterMultimapNmax 20 --outFilterType BySJout --

551 alignSJoverhangMin 8 --alignSJDBoverhangMin 2; all parameters used for Ribo-seq were the
552 same except --outFilterMismatchNmax 1). We then used the bam files from both RNA-seq and
553 Ribo-seq for RiboTaper (v1.3.1a) (Calviello et al., 2016) to identify translated ORFs. The
554 distribution of Ribo-seq reads in different genome features was calculated using Ribo-seQC.
555 The Ribo-seq metaplot was created using the create_metaplots.bash function in RiboTaper.
556 The Ribo-seq read lengths and offsets for RiboTaper were 24, 25, 26, 27, 28 and 8, 9, 10, 11,
557 12, respectively. Combining both 20-min and 60-min samples, the total reads mapped to the
558 genome for Ribo-seq and RNA-seq were 298.01 million and 180.67 million, respectively. The
559 discovered uORFs and mORFs were extracted from the RiboTaper output ORF_max_filt file. To
560 calculate translation efficiency, we first used STAR to map the RNA-seq and Ribo-seq reads to
561 the CDS of annotated coding genes. The resulting bam files were used to quantify the
562 transcripts per million (TPM) of each gene with RSEM (v1.3.1) (Li and Dewey, 2011). Then,
563 translation efficiency was calculated by dividing the Ribo-seq TPM by the RNA-seq TPM. The
564 root and shoot data from Hsu et al., 2016 (Hsu et al., 2016) were reanalyzed using the
565 Araport11 annotation. The P_sites_all files from RiboTaper were processed using the following
566 code: cut -f 1,3,6 P_sites_all | sort | uniq -c | sed -r 's/^(*[^]+) +/\1\t/' > output.txt to combine the
567 read counts at each P-site. Data visualization and statistical analysis were performed in R
568 (v4.0.3) (R Core Team (2013), 2017). In particular, the gene levels from the Ribo-seq and RNA-
569 seq reads were visualized with RiboPlotR (Wu and Hsu, 2021).
570
571 The mRNA half-life and mRNA decay rate data were downloaded from Szabo et al. and
572 Sorenson et al., respectively (Szabo et al., 2020; Sorenson et al., 2018). The quantitative
573 proteomics data for Arabidopsis shoots and roots were downloaded from Song et al. (Song et
574 al., 2018). The expression levels (TPM) used for investigating mRNA levels of uORF genes in
575 different tissues, developmental stages, and ecotypes in Arabidopsis and in tomato were
576 directly downloaded from the EMBO-EBI expression atlas (E-MTAB-7978: Arabidopsis tissue
577 atlas, E-GEOD-53197: 17 Arabidopsis thaliana accessions, E-MTAB-4812: tomato root, leaf,
578 flower [two stages], and fruit [six stages], and E-MTAB-4813: three longitudinal sections of six
579 stages during tomato fruit development) (Papatheodorou et al., 2020). Our previous Ribo-seq
580 and RNA-seq data for Arabidopsis roots and shoots were described in (Hsu et al., 2016); our
581 previous Ribo-seq and RNA-seq data for tomato roots were described in (Wu et al., 2019). The
582 published RNA-seq data for the NMD mutants *upf1* (*lba1*) *upf3-1* and its wild-type control (GEO:
583 GSE41432), *smg7-1 pad4* and its *pad4* control (GEO: GSE55884), *upf1-3 pad4* and its *pad4*
584 control (GEO: GSE146340), and *upf2-10* and its wild-type control (SRA: SRP056795) were

585 downloaded, and the mRNA levels were quantified by Kalisto (Bray et al., 2016) and the TPM
586 values were further analyzed to compare the steady-state mRNA levels in different categories of
587 genes.
588
589 **AVAILABILITY OF DATA AND MATERIALS**
590 All raw and processed sequencing data generated in this study have been submitted to the
591 NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession
592 number GSE183264. The genomic locations of TuORFs are available on TAIR genome browser
593 (https://www.arabidopsis.org). Codes for this study are available at
594 https://github.com/hsinyenwu/TuORF_vs_NMD_2022.
595
596 **COMPETING INTEREST STATEMENT**
597 The authors declare that they have no competing interests.
598
599 **AUTHOR CONTRIBUTIONS AND ACKNOWLEDGEMENTS**
600 HLW and PYH designed the research, PYH performed the sequencing experiments, HLW
601 analyzed the sequencing data, HLW and PYH interpreted the data, and HLW and PYH wrote
602 the paper. We thank Robert Last and Melissa Lehti-Shiu for critical reading and helpful
603 comments on the manuscript. This work used the Vincent J. Coates Genomics Sequencing
604 Laboratory at UC Berkeley, supported by an NIH S10 OD018174 Instrumentation Grant. This
605 work was supported by a National Science Foundation grant (2051885) and a Michigan State
606 University startup grant to PYH.
607
608 **REFERENCES**
609 **Andreev, D.E., Arnold, M., Kiniry, S.J., Loughran, G., Michel, A.M., Rachinskii, D., and**
610 **Baranov, P. V.** (2018). TASEP modelling provides a parsimonious explanation for the
611 ability of a single uORF to derepress translation during the integrated stress response.
612 eLife **7**: e32563.
613 **Von Arnim, A.G., Jia, Q., and Vaughn, J.N.** (2014). Regulation of plant translation by
614 upstream open reading frames. Plant Science **214**: 1–12.
615 **Baird, T.D., Cheng, K.C.C., Chen, Y.C., Buehler, E., Martin, S.E., Inglese, J., and Hogg, J.R.**
616 (2018). ICE1 promotes the link between splicing and nonsense-mediated mRNA decay.
617 eLife **7**.
618 **Bazin, J., Baerenfaller, K., Gosai, S.J., Gregory, B.D., Crespi, M., and Bailey-Serres, J.**

619 (2017). Global analysis of ribosome-associated noncoding RNAs unveils new modes of
620 translational regulation. Proceedings of the National Academy of Sciences of the United
621 States of America **114**: E10018–E10027.
622 **Beelman, C.A. and Parker, R.** (1994). Differential effects of translational inhibition in cis and in
623 trans on the decay of the unstable yeast MFA2 mRNA. Journal of Biological Chemistry
624 **269**: 9687–9692.
625 **Boudsocq, M., Barbier-Brygoo, H., and Laurière, C.** (2004). Identification of nine sucrose
626 nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in
627 Arabidopsis thaliana. Journal of Biological Chemistry **279**: 41758–41766.
628 **Brar, G.A. and Weissman, J.S.** (2015). Ribosome profiling reveals the what, when, where and
629 how of protein synthesis. Nature Reviews Molecular Cell Biology **16**: 651–664.
630 **Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L.** (2016). Near-optimal probabilistic RNA-
631 seq quantification. Nature biotechnology **34**: 525–527.
632 **Cai, Q., Fukushima, H., Yamamoto, M., Ishii, N., Sakamoto, T., Kurata, T., Motose, H., and**
633 **Takahashi, T.** (2016). The SAC51 family plays a central role in thermospermine responses
634 in arabidopsis. Plant and Cell Physiology **57**: 1583–1592.
635 **Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler,**
636 **M., Obermayer, B., and Ohler, U.** (2016). Detecting actively translated open reading
637 frames in ribosome profiling data. Nature Methods **13**: 165–170.
638 **Calvo, S.E., Pagliarini, D.J., and Mootha, V.K.** (2009). Upstream open reading frames cause
639 widespread reduction of protein expression and are polymorphic among humans.
640 Proceedings of the National Academy of Sciences of the United States of America **106**:
641 7507–7512.
642 **Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K.** (2018). Non-invasive
643 measurement of mRNA decay reveals translation initiation as the major determinant of
644 mRNA stability. eLife **7**: e32536.
645 **Chang, H.M., Triboulet, R., Thornton, J.E., and Gregory, R.I.** (2013). A role for the Perlman
646 syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature **497**: 244–248.
647 **Chew, G.L., Pauli, A., and Schier, A.F.** (2016). Conservation of uORF repressiveness and
648 sequence features in mouse, human and zebrafish. Nature Communications **7**: 11663.
649 **Chiam, N.-C., Fujimura, T., Sano, R., Akiyoshi, N., Hiroyama, R., Watanabe, Y., Motose, H.,**
650 **Demura, T., and Ohtani, M.** (2019). Nonsense-Mediated mRNA Decay Deficiency Affects
651 the Auxin Response and Shoot Regeneration in Arabidopsis. Plant and Cell Physiology **60**:
652 2000–2014.

653 **Clarke, J.T., Warnock, R.C.M., and Donoghue, P.C.J.** (2011). Establishing a time-scale for
654 plant evolution. The New Phytologist **192**: 266–301.
655 **Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D.** (2015). Young genes out of
656 the male: An insight from evolutionary age analysis of the pollen transcriptome. Molecular
657 Plant **8**: 935–945.
658 **Degtiar, E., Fridman, A., Gottlieb, D., Vexler, K., Berezin, I., Farhi, R., Golani, L., and**
659 **Shaul, O.** (2015). The feedback control of UPF3 is crucial for RNA surveillance in plants.
660 Nucleic Acids Research **43**: 4219–4235.
661 **Dever, T.E., Ivanov, I.P., and Sachs, M.S.** (2020). Conserved Upstream Open Reading Frame
662 Nascent Peptides That Control Translation. Annual Review of Genetics **54**: 237–264.
663 **Devlin, P.F., Patel, S.R., and Whitelam, G.C.** (1998). Phytochrome E Influences Internode
664 Elongation and Flowering Time in Arabidopsis. The Plant Cell **10**: 1479–1487.
665 **Deyholos, M.K., Cavaness, G.F., Hall, B., King, E., Punwani, J., Van Norman, J., and**
666 **Sieburth, L.E.** (2003). Varicose, a WD-domain protein, is required for leaf blade.
667 Development **130**: 6577–6588.
668 **Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,**
669 **Chaisson, M., and Gingeras, T.R.** (2013). STAR: ultrafast universal RNA-seq aligner.
670 Bioinformatics **29**: 15–21.
671 **Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Rätsch, G., and**
672 **Wachter, A.** (2013). Nonsense-mediated decay of alternative precursor mRNA splicing
673 variants is a major determinant of the Arabidopsis steady state transcriptome. The Plant
674 Cell **25**: 3726–3742.
675 **Ellis, C.M., Nagpal, P., Young, J.C., Hagen, G., Guilfoyle, T.J., and Reed, J.W.** (2005).
676 AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence
677 and floral organ abscission in Arabidopsis thaliana. Development **132**: 4563–4574.
678 **Filichkin, S.A., Cumbie, J.S., Dharmawardhana, P., Jaiswal, P., Chang, J.H., Palusa, S.G.,**
679 **Reddy, A.S.N., Megraw, M., and Mockler, T.C.** (2015). Environmental stresses modulate
680 abundance and timing of alternatively spliced circadian transcripts in Arabidopsis.
681 Molecular Plant **8**: 207–227.
682 **Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., Bautor,**
683 **J., Dekrout, B., Jonak, C., Jiménez-Gómez, J.M., Parker, J.E., and Riha, K.** (2014).
684 Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant
685 antibacterial defense. Cell Host and Microbe **16**: 376–390.
686 **Goeres, D.C., Van Norman, J.M., Zhang, W., Fauver, N.A., Spencer, M. Lou, and Sieburth,**

687 **L.E.** (2007). Components of the Arabidopsis mRNA decapping complex are required for
688 early seedling development. The Plant Cell **19**: 1549–1564.
689 **Harnett, D., Meerdink, E., Calviello, L., Sydow, D., and Ohler, U.** (2021). Genome-Wide
690 Analysis of Actively Translated Open Reading Frames Using RiboTaper/ORFquant.
691 Methods in Molecular Biology **2252**: 331–346.
692 **Hayden, C.A. and Jorgensen, R.A.** (2007). Identification of novel conserved peptide uORF
693 homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups
694 and preferential association with transcription factor-encoding genes. BMC Biology **5**: 32.
695 **He, F. and Jacobson, A.** (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective
696 Transcripts Is only Part of the Story. Annual Review of Genetics **49**: 339–366.
697 **Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N.** (2016). Translational control by 5′-
698 untranslated regions of eukaryotic mRNAs. Science **352**: 1413–1416.
699 **Van Der Horst, S., Snel, B., Hanson, J., and Smeekens, S.** (2019). Novel pipeline identifies
700 new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid
701 sequences in the 5′ leader of mRNAs in Arabidopsis thaliana. RNA **25**: 292–304.
702 **Hou, C.-Y., Lee, W.-C., Chou, H.-C., Chen, A.-P., Chou, S.-J., and Chen, H.-M.** (2016).
703 Global analysis of truncated RNA ends reveals new insights into ribosome stalling in
704 plants. The Plant Cell **28**: 2398–2416.
705 **Hsu, P.Y., Calviello, L., Wu, H.-Y.L., Li, F.-W., Rothfels, C.J., Ohler, U., and Benfey, P.N.**
706 (2016). Super-resolution ribosome profiling reveals unannotated translation events in
707 Arabidopsis. Proceedings of the National Academy of Sciences of the United States of
708 America **113**: E7126–E7135.
709 **Hu, Q., Merchante, C., Stepanova, A.N., Alonso, J.M., and Heber, S.** (2016). Genome-Wide
710 Search for Translated Upstream Open Reading Frames in Arabidopsis Thaliana. IEEE
711 Transactions on Nanobioscience **15**: 150–159.
712 **Huq, E. and Quail, P.H.** (2002). PIF4, a phytochrome-interacting bHLH factor, functions as a
713 negative regulator of phytochrome B signaling in Arabidopsis. EMBO Journal **21**: 2441–
714 2450.
715 **Hurt, J.A., Robertson, A.D., and Burge, C.B.** (2013). Global analyses of UPF1 binding and
716 function reveal expanded scope of nonsense-mediated mRNA decay. Genome Research
717 **23**: 1636–1650.
718 **Imai, A., Hanzawa, Y., Komura, M., Yamamoto, K.T., Komeda, Y., and Takahashi, T.** (2006).
719 The dwarf phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an
720 upstream ORF of a bHLH gene. Development **133**: 3575–3585.

721 **Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S.** (2009). Genome-
722 wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.
723 Science **324**: 218–223.
724 **Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., and Maquat, L.E.** (2008).
725 Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA
726 Decay. Cell **133**: 314–327.
727 **Jeong, H.J., Kim, Y.J., Kim, S.H., Kim, Y.H., Lee, I.J., Kim, Y.K., and Shin, J.S.** (2011).
728 Nonsense-mediated mRNA decay factors, UPF1 and UPF3, contribute to plant defense.
729 Plant and Cell Physiology **52**: 2147–2156.
730 **Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J.** (2016). Upstream ORFs are prevalent
731 translational repressors in vertebrates. The EMBO Journal **35**: 706–723.
732 **Jorgensen, R.A. and Dorantes-Acosta, A.E.** (2012). Conserved peptide upstream open
733 reading frames are associated with regulatory genes in angiosperms. Frontiers in Plant
734 Science **3**: 191.
735 **Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J.** (2014). Translational dynamics
736 revealed by genome-wide profiling of ribosome footprints in Arabidopsis. Proceedings of
737 the National Academy of Sciences of the United States of America **111**: E203–E212.
738 **Kalyna, M. et al.** (2012). Alternative splicing and nonsense-mediated decay modulate
739 expression of important regulatory genes in Arabidopsis. Nucleic Acids Research **40**:
740 2454–2469.
741 **Kerényi, F., Wawer, I., Sikorski, P.J., Kufel, J., and Silhavy, D.** (2013). Phosphorylation of
742 the N- and C-terminal UPF1 domains plays a critical role in plant nonsense-mediated
743 mRNA decay. Plant Journal **76**: 836–848.
744 **Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E., and Silhavy, D.** (2006).
745 Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-
746 mediated decay in plants. Nucleic Acids Research **34**: 6147–6157.
747 **Kim, B.H., Cai, X., Vaughn, J.N., and Von Arnim, A.G.** (2007). On the functions of the h
748 subunit of eukaryotic initiation factor 3 in late stages of translation initiation. Genome
749 Biology **8**: R60.
750 **Kozak, M.** (1999). Initiation of translation in prokaryotes and eukaryotes. Gene **234**: 187–208.
751 **Kurihara, Y. et al.** (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs
752 by NMD in Arabidopsis. Proceedings of the National Academy of Sciences of the United
753 States of America **106**: 2453–2458.
754 **Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., and Matsui, M.** (2018).

755 Transcripts from downstream alternative transcription start sites evade uORF-mediated
756 inhibition of gene expression in Arabidopsis. Proceedings of the National Academy of
757 Sciences of the United States of America **115**: 7831–7836.
758 **Kurihara, Y., Makita, Y., Shimohira, H., Fujita, T., Iwasaki, S., and Matsui, M.** (2020).
759 Translational landscape of protein-coding and non-protein-coding RNAs upon light
760 exposure in Arabidopsis. Plant and Cell Physiology **61**: 536–545.
761 **Kurosaki, T., Li, W., Hoque, M., Popp, M.W.L., Ermolenko, D.N., Tian, B., and Maquat, L.E.**
762 (2014). A Post-Translational regulatory switch on UPF1 controls targeted mRNA
763 degradation. Genes and Development **28**: 1900–1916.
764 **Kurosaki, T., Popp, M.W., and Maquat, L.E.** (2019). Quality and quantity control of gene
765 expression by nonsense-mediated mRNA decay. Nature Reviews Molecular Cell Biology
766 2019 20:7 **20**: 406–420.
767 **LaGrandeur, T. and Parker, R.** (1999). The cis acting sequences responsible for the differential
768 decay of the unstable MFA2 and stable PGK1 transcipts in yeast include the context of the
769 translational start codon. RNA **5**: 420–433.
770 **Langmead, B. and Salzberg, S.L.** (2012). Fast gapped-read alignment with Bowtie 2. Nature
771 Methods **9**: 357–359.
772 **Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G.,**
773 **Hahn, M.G., Feldmann, K.A., and Benfey, P.N.** (1996). The SCARECROW gene
774 regulates an asymmetric cell division that is essential for generating the radial organization
775 of the Arabidopsis root. Cell **86**: 423–433.
776 **Lawson, M.R., Lessen, L.N., Wang, J., Prabhakar, A., Corsepius, N.C., Green, R., and**
777 **Puglisi, J.D.** (2021). Mechanisms that ensure speed and fidelity in eukaryotic translation
778 termination. Science **373**: 876–882.
779 **Lee, D.S.M., Park, J., Kromer, A., Baras, A., Rader, D.J., Ritchie, M.D., Ghanem, L.R., and**
780 **Barash, Y.** (2021). Disrupting upstream translation in mRNAs is associated with human
781 disease. Nature Communications 2021 12:1 **12**: 1–14.
782 **Leivar, P., Monte, E., Al-Sady, B., Carle, C., Storer, A., Alonso, J.M., Ecker, J.R., and Quail,**
783 **P.H.** (2008). The Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and
784 PIF4, regulates responses to prolonged red light by modulating phyB levels. The Plant Cell
785 **20**: 337–352.
786 **Li, B. and Dewey, C.N.** (2011). RSEM: accurate transcript quantification from RNA-Seq data
787 with or without a reference genome. BMC Bioinformatics **12**: 323.
788 **Li, Y.R. and Liu, M.J.** (2020). Prevalence of alternative AUG and non-AUG translation initiators

789 and their regulatory effects across plants. Genome Research **30**: 1418–1433.
790 **Lin, Y., May, G.E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., and Joel**
791 **McManus, C.** (2019). Impacts of uORF codon identity and position on translation
792 regulation. Nucleic Acids Research **47**: 9358–9367.
793 **Liu, M.J., Wu, S.H., Wu, J.F., Lin, W.D., Wu, Y.C., Tsai, T.Y., Tsai, H.L., and Wu, S.H.** (2013).
794 Translational landscape of photomorphogenic Arabidopsis. Plant Cell **25**: 3699–3710.
795 **Lloyd, J.P.B. and Davies, B.** (2013). SMG1 is an ancient nonsense-mediated mRNA decay
796 effector. The Plant journal : for cell and molecular biology **76**: 800–810.
797 **Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H., and Dziembowski, A.**
798 (2013). Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation
799 pathway of human cytoplasmic mRNA. The EMBO journal **32**: 1855–1868.
800 **Merchante, C., Brumos, J., Yun, J., Hu, Q., Spencer, K.R., Enríquez, P., Binder, B.M.,**
801 **Heber, S., Stepanova, A.N., and Alonso, J.M.** (2015). Gene-specific translation
802 regulation mediated by the hormone-signaling molecule EIN2. Cell **163**: 684–697.
803 **Mizoguchi, T., Hayashida, N., Yamaguchi-Shinozaki, K., Kamada, H., and Shinozaki, K.**
804 (1995). Two genes that encode ribosomal-protein S6 kinase homologs are induced by cold
805 or salinity stress in Arabidopsis thaliana. FEBS Letters **358**: 199–204.
806 **Nagarajan, V.K., Kukulich, P.M., Von Hagel, B., and Green, P.J.** (2019). RNA degradomes
807 reveal substrates and importance for dark and nitrogen stress responses of Arabidopsis
808 XRN4. Nucleic Acids Research **47**: 9216–9230.
809 **Nagpal, P., Ellis, C.M., Weber, H., Ploense, S.E., Barkawi, L.S., Guilfoyle, T.J., Hagen, G.,**
810 **Alonso, J.M., Cohen, J.D., Farmer, E.E., Ecker, J.R., and Reed, J.W.** (2005). Auxin
811 response factors ARF6 and ARF8 promote jasmonic acid production and flower
812 maturation. Development **132**: 4107–4118.
813 **Ni, M., Tepperman, J.M., and Quail, P.H.** (1998). PIF3, a phytochrome-interacting factor
814 necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix
815 protein. Cell **95**: 657–667.
816 **Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D.** (2009). Plant upstream
817 ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant
818 Molecular Biology **71**: 367–378.
819 **Ohtani, M. and Wachter, A.** (2019). NMD-Based Gene Regulation - A Strategy for Fitness
820 Enhancement in Plants? Plant and Cell Physiology **60**: 1953–1960.
821 **Papatheodorou, I. et al.** (2020). Expression Atlas update: From tissues to single cells. Nucleic
822 Acids Research **48**: D77–D83.

823 **Park, Y.S., Hong, S.W., Oh, S.A., Kwak, J.M., Lee, H.H., and Nam, H.G.** (1993). Two putative
824 protein kinases from Arabidopsis thaliana contain highly acidic domains. Plant Molecular
825 Biology **22**: 615–624.
826 **Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S.,**
827 **Weinberg, D., Baker, K.E., Graveley, B.R., and Coller, J.** (2015). Codon optimality is a
828 major determinant of mRNA stability. Cell **160**: 1111–1124.
829 **R Core Team (2013)** (2017). R: A language and environment for statistical computing. R
830 Foundation for Statistical Computing, Vienna, Austria.
831 **Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Coller, J.** (2016).
832 The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring
833 Codon Optimality. Cell **167**: 122-132.e9.
834 **Raxwal, V.K. and Riha, K.** (2016). Nonsense mediated RNA decay and evolutionary
835 capacitance. Biochimica et Biophysica Acta **1859**: 1538–1543.
836 **Raxwal, V.K., Simpson, C.G., Gloggnitzer, J., Entinze, J.C., Guo, W., Zhang, R., Brown,**
837 **J.W.S., and Riha, K.** (2020). Nonsense-Mediated RNA Decay Factor UPF1 Is Critical for
838 Posttranscriptional and Translational Gene Regulation in Arabidopsis. The Plant Cell **32**:
839 2725–2741.
840 **Rayson, S., Arciga-Reyes, L., Wootton, L., de Torres Zabala, M., Truman, W., Graham, N.,**
841 **Grant, M., and Davies, B.** (2012). A role for nonsense-mediated mrna decay in plants:
842 Pathogen responses are induced in arabidopsis thaliana nmd mutants. PLoS ONE **7**:
843 e31917.
844 **Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C., and Riha, K.** (2012). Aberrant
845 growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is
846 caused by autoimmune-like response. Nucleic Acids Research **40**: 5615–5624.
847 **Saul, H. et al.** (2009). The upstream open reading frame of the Arabidopsis AtMHX gene has a
848 strong impact on transcript accumulation through the nonsense-mediated mRNA decay
849 pathway. The Plant Journal **60**: 1031–1042.
850 **Schwartz, D.C. and Parker, R.** (2000). mRNA Decapping in Yeast Requires Dissociation of the
851 Cap Binding Protein, Eukaryotic Translation Initiation Factor 4E. Molecular and Cellular
852 Biology **20**: 7933–7942.
853 **Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B.** (2013). Rate-Limiting Steps in
854 Yeast Protein Translation. Cell **153**: 1589–1601.
855 **Sharrock, R.A. and Quail, P.H.** (1989). Novel phytochrome sequences in Arabidopsis thaliana:
856 structure, evolution, and differential expression of a plant regulatory photoreceptor family.

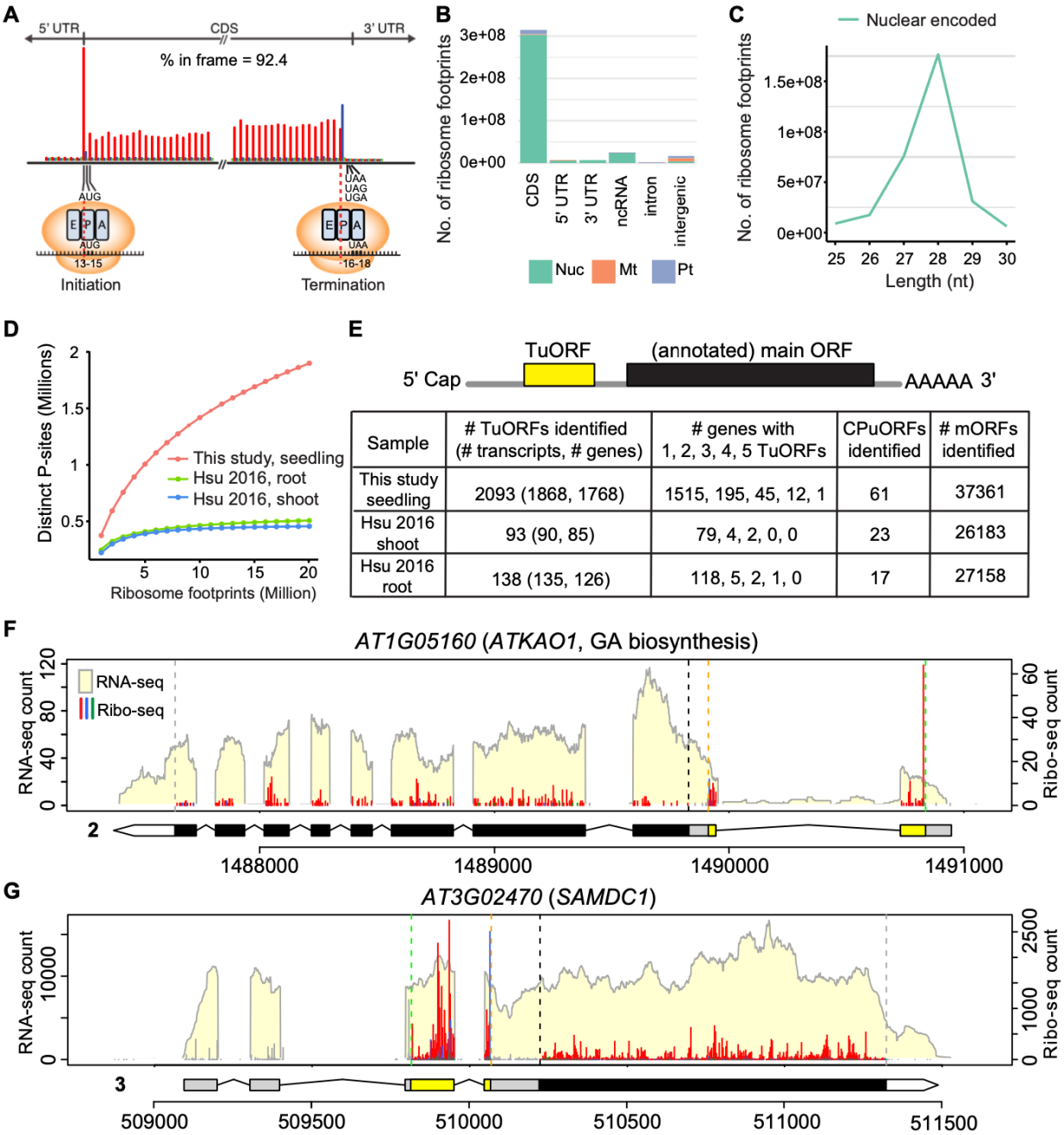
857 Genes & Development **3**: 1745–1757.
858 **Sieburth, L.E. and Vincent, J.N.** (2018). Beyond transcription factors: Roles of mrna decay in
859 regulating gene expression in plants [version 1; referees: 3 approved]. F1000Research **7**:
860 1940.
861 **Silverstone, A.L., Ciampaglio, C.N., and Sun, T.P.** (1998). The Arabidopsis RGA gene
862 encodes a transcriptional regulator repressing the gibberellin signal transduction pathway.
863 The Plant Cell **10**: 155–169.
864 **Song, G., Hsu, P.Y., and Walley, J.W.** (2018). Assessment and Refinement of Sample
865 Preparation Methods for Deep and Quantitative Plant Proteome Profiling. Proteomics **18**:
866 e1800220.
867 **Sorenson, R.S., Deshotel, M.J., Johnson, K., Adler, F.R., and Sieburth, L.E.** (2018).
868 Arabidopsis mRNA decay landscape arises from specialized RNA decay substrates,
869 decapping-mediated feedback, and redundancy. Proceedings of the National Academy of
870 Sciences of the United States of America **115**: E1485–E1494.
871 **Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps,**
872 **J.A., and Kay, S.A.** (2000). Cloning of the Arabidopsis clock gene TOC1, an
873 autoregulatory response regulator homolog. Science **289**: 768–771.
874 **Szabo, E.X., Reichert, P., Lehniger, M.K., Ohmer, M., de Francisco Amorim, M., Gowik, U.,**
875 **Schmitz-Linneweber, C., and Laubinger, S.** (2020). Metabolic labeling of RNAs uncovers
876 hidden features and dynamics of the Arabidopsis transcriptome. The Plant Cell **32**: 871–
877 887.
878 **Takahashi, H., Takahashi, A., Naito, S., and Onouchi, H.** (2012). BAIUCAS: A novel BLAST-
879 based algorithm for the identification of upstream open reading frames with conserved
880 amino acid sequences and its application to the Arabidopsis thaliana genome.
881 Bioinformatics **28**: 2231–2241.
882 **Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., Chiba, Y., Hirai,**
883 **M.Y., Akiyama, T., Onouchi, H., Naito, S., and Fujiwara, T.** (2016). The minimum open
884 reading frame, AUG-Stop, induces boron-dependent ribosome stalling and mRNA
885 degradation. The Plant Cell **28**: 2830–2849.
886 **Uchiyama-Kadokura, N., Murakami, K., Takemoto, M., Koyanagi, N., Murota, K., Naito, S.,**
887 **and Onouchi, H.** (2014). Polyamine-responsive ribosomal arrest at the stop codon of an
888 upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA
889 decay in Arabidopsis thaliana. Plant and Cell Physiology **55**: 1556–1567.
890 **Vaughn, J.N., Ellingson, S.R., Mignone, F., and Arnim, A. von** (2012). Known and novel

891 post-transcriptional regulatory sequences are conserved across plant families. RNA **18**:
892 368–84.
893 **Wolfa, Y.I., Novichkovb, P.S., Kareva, G.P., Koonina, E. V., and Lipmana, D.J.** (2009). The
894 universal distribution of evolutionary rates of genes and distinct characteristics of
895 eukaryotic genes of different apparent ages. Proceedings of the National Academy of
896 Sciences of the United States of America **106**: 7273–7280.
897 **Wu, H.-Y.L. and Hsu, P.Y.** (2021). RiboPlotR: a visualization tool for periodic Ribo-seq reads.
898 Plant Methods **17**: 1–7.
899 **Wu, H.Y.L., Song, G., Walley, J.W., and Hsu, P.Y.** (2019). The tomato translational landscape
900 revealed by transcriptome assembly and ribosome profiling. Plant Physiology **181**: 367–
901 380.
902 **Xu, G., Yuan, M., Ai, C., Liu, L., Zhuang, E., Karapetyan, S., Wang, S., and Dong, X.** (2017).
903 uORF-mediated translation allows engineered plant disease resistance without fitness
904 costs. Nature **545**: 491–494.
905 **Xu, J., Yang, J.Y., Niu, Q.W., and Chua, N.H.** (2006). Arabidopsis DCP2, DCP1, and
906 VARICOSE form a decapping complex required for postembryonic development. The Plant
907 Cell **18**: 3386–3398.
908 **Xu, Z., Hu, L., Shi, B., Geng, S., Xu, L., Wang, D., and Lu, Z.J.** (2018). Ribosome elongating
909 footprints denoised by wavelet transform comprehensively characterize dynamic cellular
910 translation events. Nucleic Acids Research **46**: e109.
911 **Yoine, M., Ohto, M.A., Onai, K., Mita, S., and Nakamura, K.** (2006). The lba1 mutation of
912 UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic
913 phenotypic changes and altered sugar signalling in Arabidopsis. The Plant Journal **47**: 49–
914 62.
915 **Yu, X., Willmann, M.R., Anderson, S.J., and Gregory, B.D.** (2016). Genome-wide mapping of
916 uncapped and cleaved transcripts reveals a role for the nuclear mrna cap-binding complex
917 in cotranslational rna decay in Arabidopsis. The Plant Cell **28**: 2385–2397.
918 **Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C.** (2018). Genome
919 editing of upstream open reading frames enables translational control in plants. Nature
920 Biotechnology **36**: 894–898.
921 **Zhang, H., Wang, Y., Wu, X., Tang, X., Wu, C., and Lu, J.** (2021). Determinants of genome-
922 wide distribution and evolution of uORFs in eukaryotes. Nature Communications **12**: 1–17.
923 **Zhang, S.H., Lawton, M.A., Hunter, T., and Lamb, C.J.** (1994). atpk1, a novel ribosomal
924 protein kinase gene from Arabidopsis. I. Isolation, characterization, and expression.

925 Journal of Biological Chemistry **269**: 17586–17592.
926 **Zhang, W., Murphy, C., and Sieburth, L.E.** (2010). Conserved RNaseII domain protein
927 functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant
928 phenotypes. Proceedings of the National Academy of Sciences of the United States of
929 America **107**: 15981–15985.
930 **Zinshteyn, B., Sinha, N.K., Enam, S.U., Koleske, B., and Green, R.** (2021). Translational
931 repression of NMD targets by GIGYF2 and EIF4E2. PLoS Genetics **17**: e1009813.
932

933 **FIGURES 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

934

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

936 (A) Metagene analysis of 28-nt ribosome footprints mapped to regions near the start and
937 stop codons of annotated ORFs. The reads are presented with their first nt at the P-site,
938 which is the 13th nt for 28-nt footprints. The reads are colored in red, blue, and green to
939 indicate they are in the first (expected), second, and third reading frames, respectively.

940 The majority of footprints were mapped to the CDS in the expected reading frame
941 (92.4% in frame).
942 (B) Genomic features mapped by ribosome footprints. Reads that mapped to nuclear (Nuc)-,
943 mitochondrial (Mt)-, and plastid (Pt)-encoded genes are shown.
944 (C) Length distribution of ribosome footprints. Reads that mapped to nuclear-encoded genes
945 are presented.
946 (D) Distinct P-sites detected in 1 to 20 million randomly selected ribosome footprints from
947 our current and previous datasets.
948 (E) Numbers of TuORFs, CPuORFs, and mORFs identified in our current and previous
949 datasets.
950 (F-G) RNA-seq and Ribo-seq profiles of *ATKAO1 and SAMDC1* from our current data. RNA-
951 seq coverage is shown with a light-yellow background. Ribo-seq reads are presented
952 with their first nt at the P-site, and they are colored in red, blue, and green to indicate
953 they are in the first (expected), second, and third reading frames, respectively. Within the
954 gene models, yellow boxes represent the TuORFs, black boxes represent the annotated
955 mORFs, and gray and white regions indicate 5' UTRs and 3' UTRs, respectively. The
956 specific isoform being considered is indicated to the left of the gene model. Within the
957 profiles, green and orange vertical dashed lines represent translation start and stop,
958 respectively, for the TuORF. Black and gray vertical dashed lines represent translation
959 start and stop, respectively, for the annotated mORF.
960
961
962
963
964
965

Figure 2

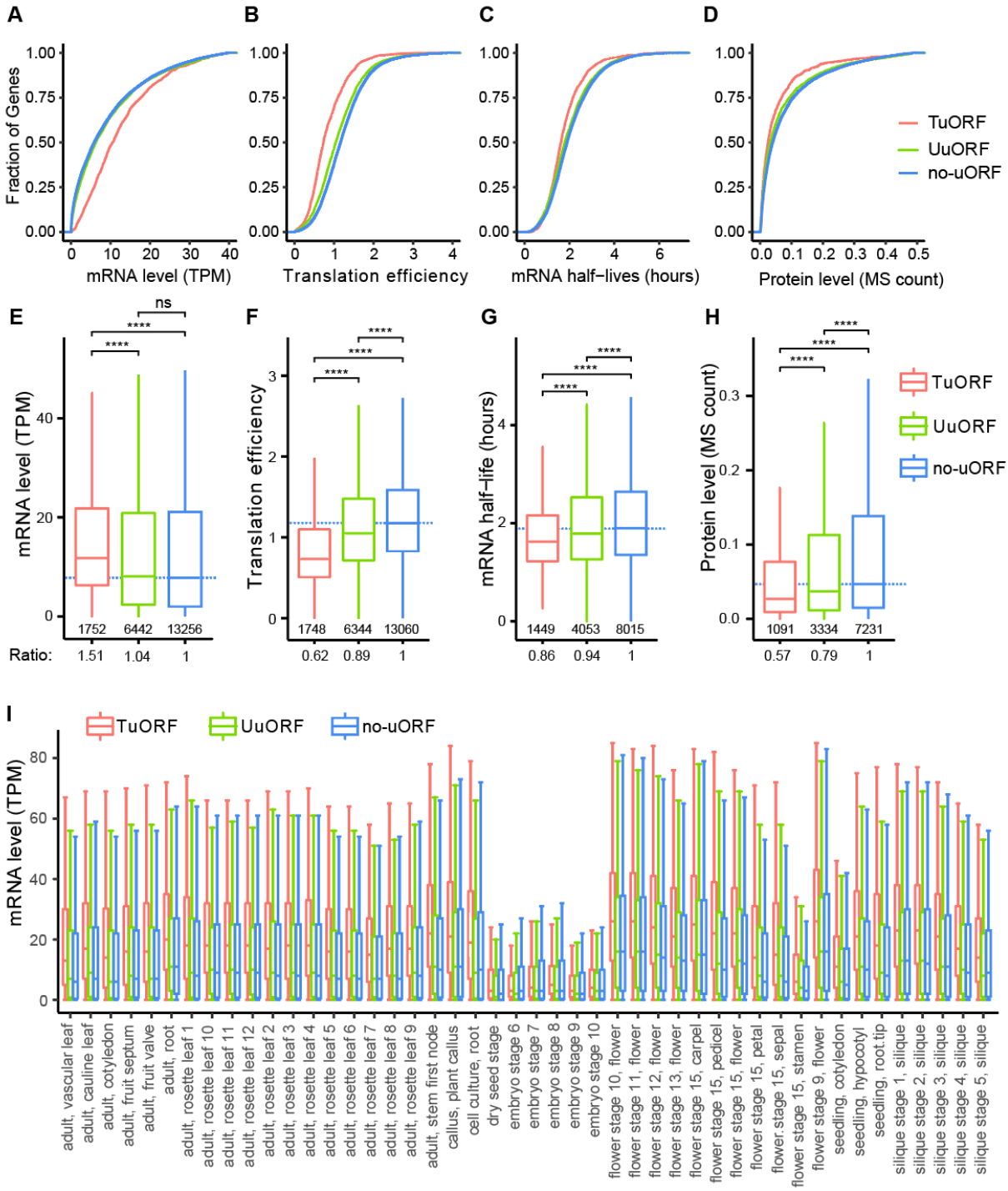


Figure 2. TuORF-containing genes have higher mRNA abundance but are associated with lower translation efficiency, shorter mRNA half-lives, and lower protein abundance

970 (A-D) Cumulative plots and (E-H) boxplots for the mRNA levels, translation efficiency, mRNA
971 half-lives, and protein levels of TuORF, UuORF, and no-uORF genes in Arabidopsis seedlings.
972 The mRNA half-life data, which were measured with metabolic labeling in Arabidopsis
973 seedlings, were extracted from (Szabo et al., 2020), and the protein abundance as determined
974 by quantitative proteomics in Arabidopsis seedling shoots were extracted from (Song et al.,
975 2018). Within the boxplots (E-H), the dashed lines mark the median level of the no-uORF
976 genes; the number of genes in each category is listed below the lower whisker. The ratios
977 underneath the plots indicate the median of each group normalized to that of the no-uORF
978 genes. The statistical significance was determined by Wilcoxon signed-rank test. The adjusted
979 p-value was determined using the Benjamini and Yekutieli procedure to control for the false
980 discovery rate in multiple testing (* 0.05 > p > 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-4,
981 **** 1e-4 > p > 0). These statistical analyses and graphic layouts were applied to all boxplots in
982 the study.
983 (I) The mRNA levels of TuORF, UuORF, and no-uORF genes in 46 Arabidopsis tissues and
984 developmental stages from independent studies. The TPM values from RNA-seq quantification
985 were downloaded from the EMBO-EBI expression atlas (Papatheodorou et al., 2020). Additional
986 82 Arabidopsis and tomato RNA-seq datasets were presented in Figure S5.
987
988
989

Figure 3

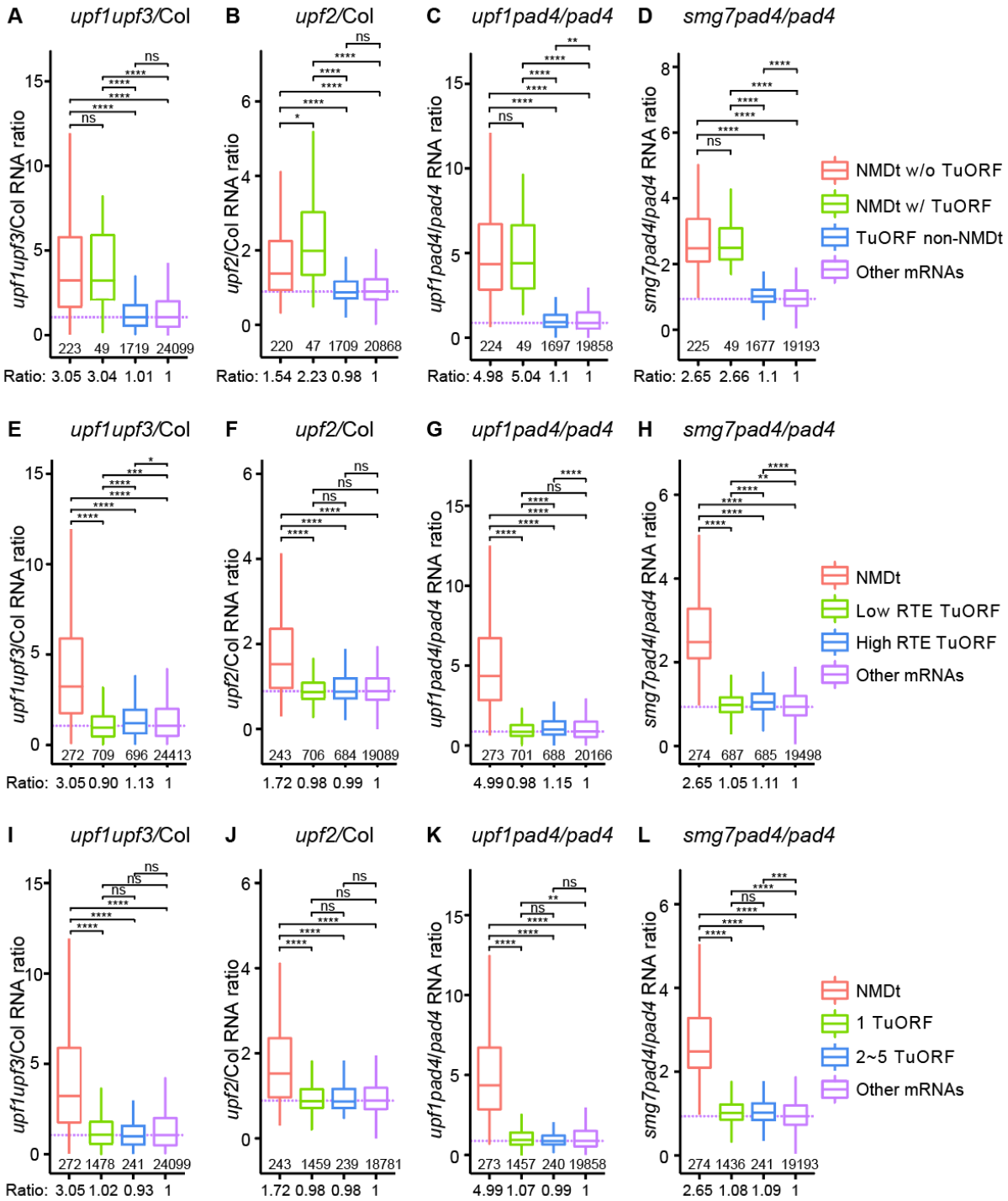


Figure 3: TuORF mRNAs are not upregulated in the NMD mutants

The ratios of mRNA levels between NMD mutants and their control plants from four RNA-seq datasets (Drechsel et al., 2013; Merchante et al., 2015; Raxwal et al., 2020; Gloggnitzer et al., 2014) were used to examine the effects of the NMD pathway on four gene groups: NMD targets

995 without TuORF (NMDt w/o TuORF), NMD targets with TuORF (NMDt w/ TuORF), TuORF
996 mRNAs not identified as NMD targets (TuORF non-NMDt) and other mRNAs. (A) *upf1* (*lba1*
997 allele) *upf3-1* double mutant compared with wild type (Col-0). (B) *upf2-10* compared with wild
998 type (Col-0). (C) *upf1-3 pad4* compared with *pad4*. (D) *smg7-1 pad4* compared with *pad4*. See
999 Table S3 for specific growth conditions for each dataset. (E-H) Same datasets as 4A-D, but the
1000 TuORF mRNAs were grouped based on low or high relative TuORF translation efficiency (RTE)
1001 (i.e., TE_{TuORF} / TE_{mORF}). (I-L) Same datasets as 4A-D, but the TuORF mRNAs were grouped
1002 based on the number of TuORF(s). Within the boxplots, the dashed lines mark the median level
1003 of other mRNAs; the number of genes in each category is listed below the lower whisker. The
1004 ratios underneath the plots indicate the median of each group normalized to that of other
1005 mRNAs. The statistical significance was determined as described in the Figure 2 legend.

Figure 4

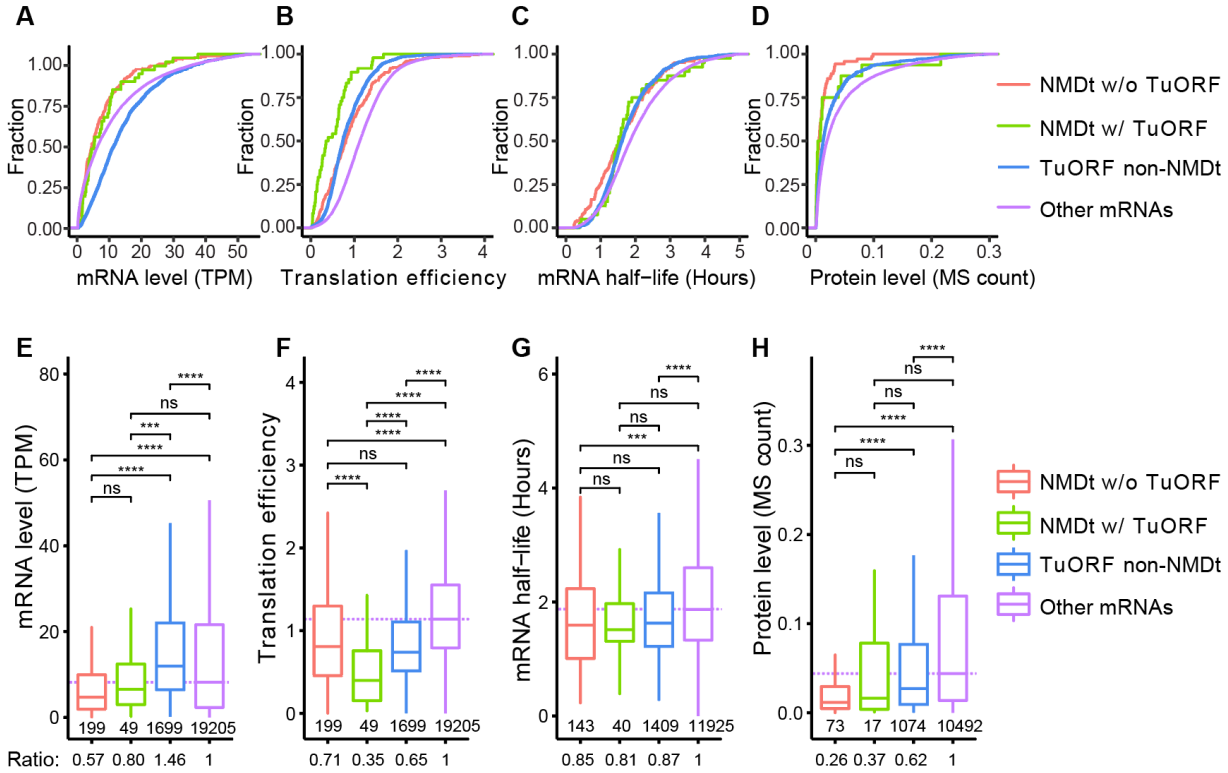


Figure 4. Expression patterns of NMD targets and TuORF mRNAs.

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 not identified as NMD targets 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 5

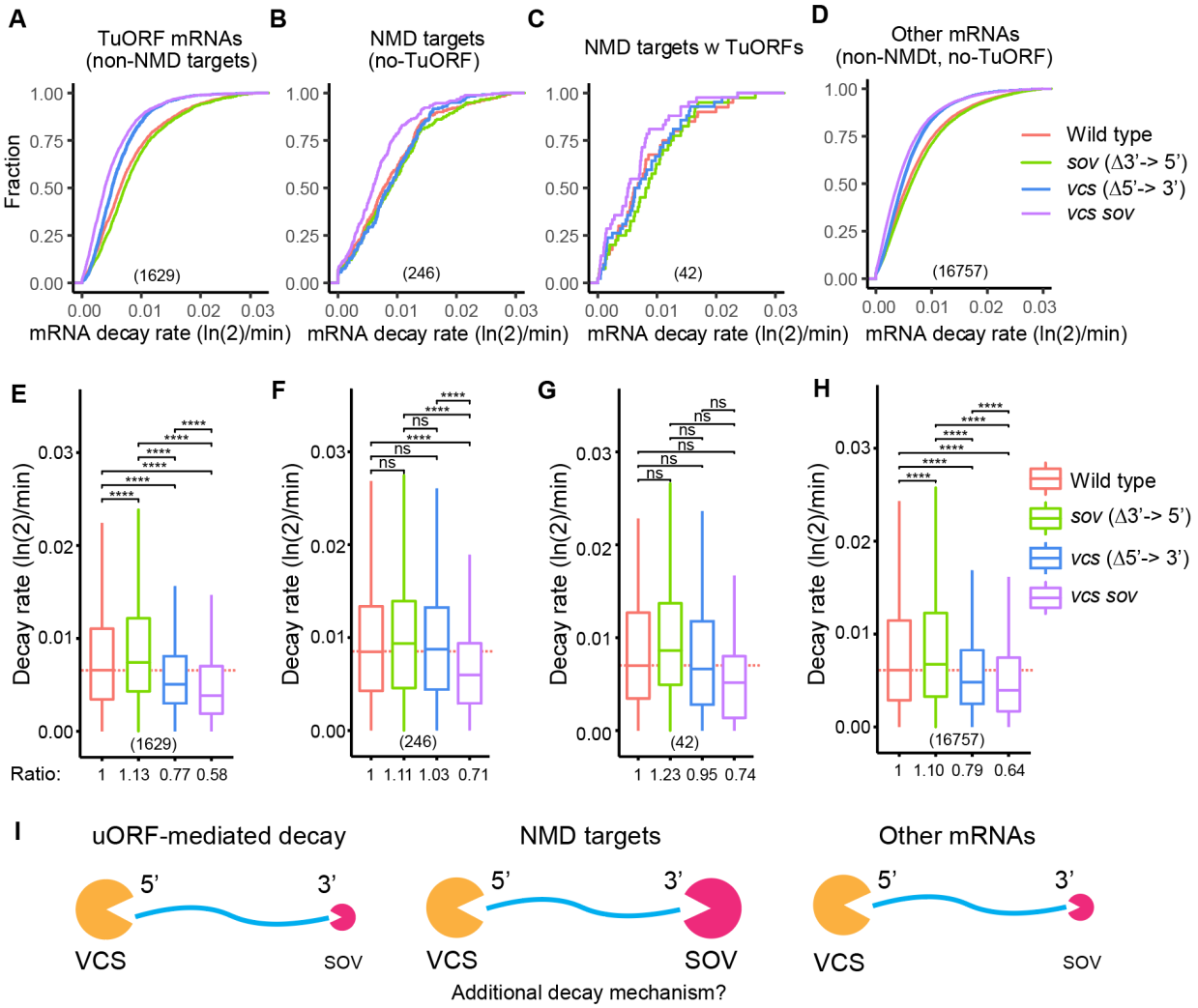


Figure 5. TuORF mRNAs and NMD targets are degraded through different mRNA decay mechanisms

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 (Sorenson et al., 2018). Because Col-0 is naturally an *sov* mutant, the wild type here is Col-0 carrying the functional *SOV* allele from Landsberg *erecta* driven by its native promoter (Zhang et al., 2010; Goeres et al., 2007). The *sov* mutant here is Col-0, and *vcs* is Col-0 with the *vsc-7* allele and the L*er SOV* transgene. The *vcs sov* double mutant is Col-0 carrying the *vcs-7* mutation (Sorenson et al., 2018). Within the boxplots, the dashed lines mark the median level of wild type; 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 wild type. The statistical significance

1030 was determined as described in the Figure 2 legend. (I) Model of mRNA decay for NMD targets,
1031 TuORF mRNAs, and other mRNAs. The orange packman indicates 5'-to-3' decay machinery,
1032 and the hot pink packman indicates 3'-to-5' decay machinery.
1033

1034 **SUPPLEMENTARY DATA**

1035 **Figure S1.** Correlations among samples and comparisons between our previous and current
1036 datasets

1037 **Figure S2.** Examples of TuORFs in important regulatory genes.

1038 **Figure S3.** Protein abundance of the uORF-containing genes in Arabidopsis root

1039 **Figure S4.** The mRNA levels and translation efficiency of TuORF mRNAs in tomato

1040 **Figure S5.** mRNA levels of uORF-containing genes in different ecotypes, growth stages, and
1041 tissues in Arabidopsis and tomato

1042 **Figure S6.** mRNA levels of CPuORF-containing genes compared to those of no-uORF genes

1043 **Figure S7.** TuORF genes are larger than other genes

1044 **Figure S8.** Lengths of the uORF peptides for NMD target and non-NMD target TuORFs

1045 **Figure S9.** Lengths of UTRs and CDS for NMD targets and TuORF genes

1046 **Figure S10.** The mRNA decay rates of NMD targets and TuORF-containing mRNAs in different
1047 genetic backgrounds

1048 **Table S1**. Translated ORFs identified by RiboTaper

1049 - **Table S1A-C.** Translated uORFs identified in the current dataset and our previous
1050 Arabidopsis shoot and root datasets
1051 - **Table S1D-F.** Translated mORFs identified in the current dataset and our previous
1052 Arabidopsis shoot and root datasets

1053 **Table S2.** GO-term analysis of TuORF genes

1054 **Table S3.** Summary of the large-scale datasets used in this study

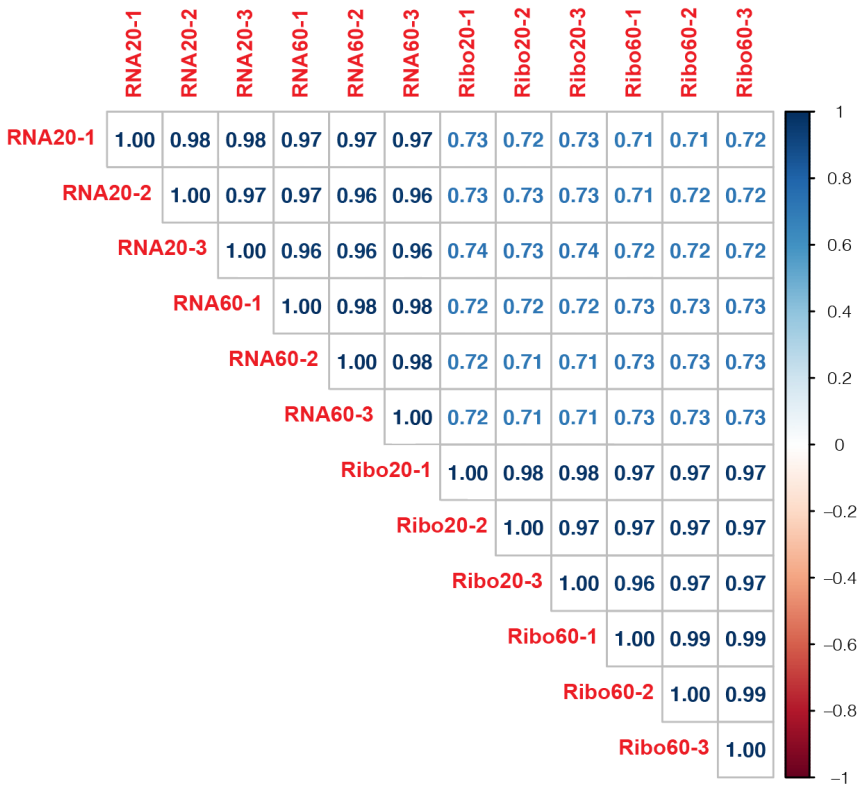
1055 **Table S4.** NMD target gene list

1056 - **Table S4A.** High-confidence NMD targets expressed in our data
1057 - **Table S4B.** 49 genes identified as high-confidence NMD target TuORF genes

1058

SUPPLEMENTARY FIGURES

A



B

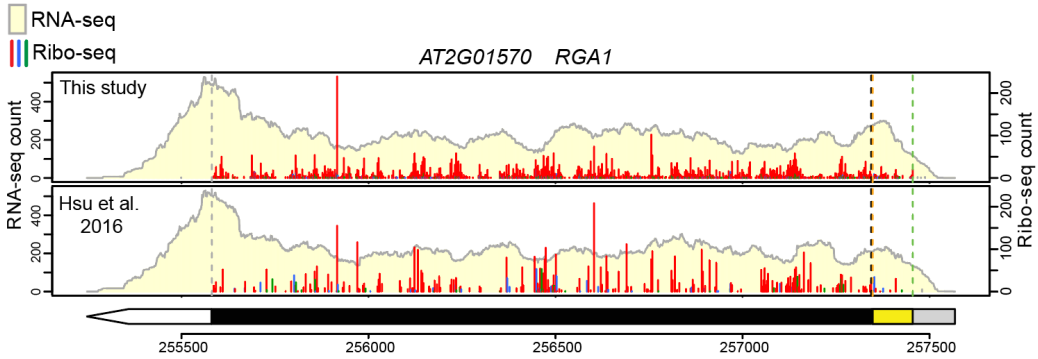
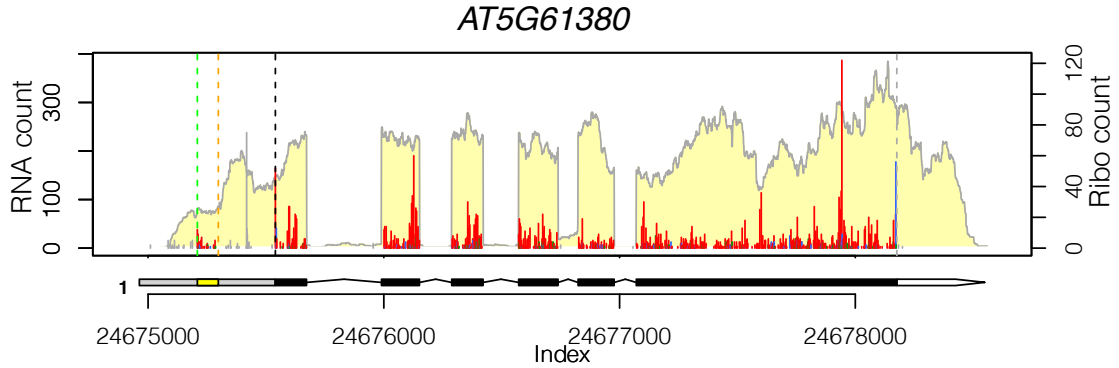


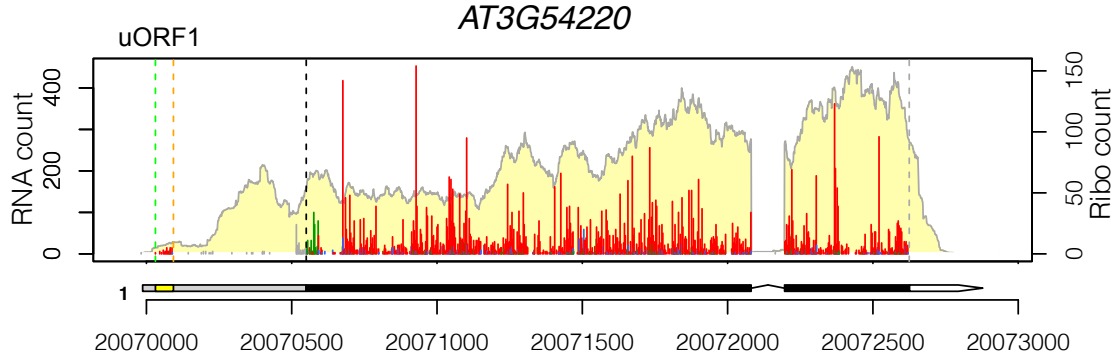
Figure S1. Correlations among samples and comparisons between previous and current RNA-seq and Ribo-seq datasets. (A) Pearson correlations among three biological replicates of our current RNA-seq and Ribo-seq analyses of samples harvested after 20 or 60 min of DMSO treatment (these were control samples 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), shoot) and the current data. Data representations are the same as those described in the Figure 1F legend.

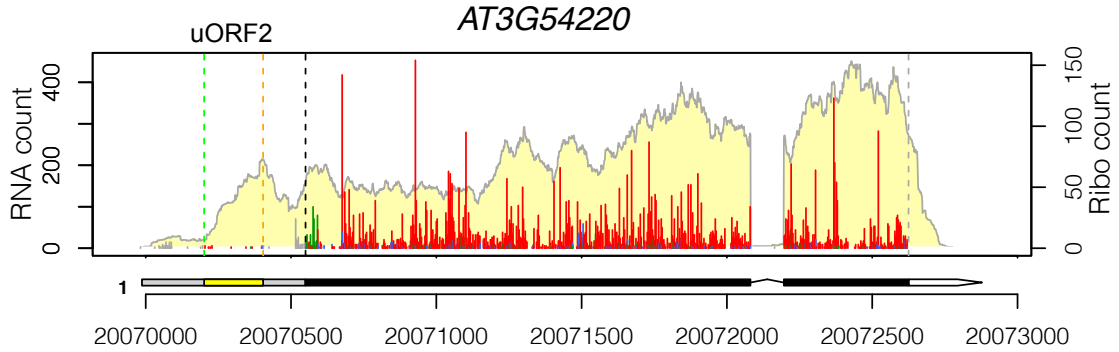
Figure S2. TuORFs in known important regulatory genes

A. *TOC1*

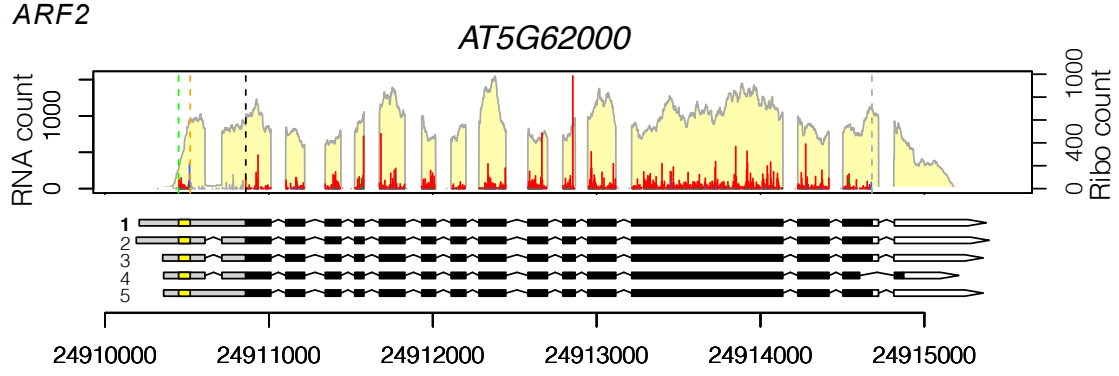


B. *SCARECROW*

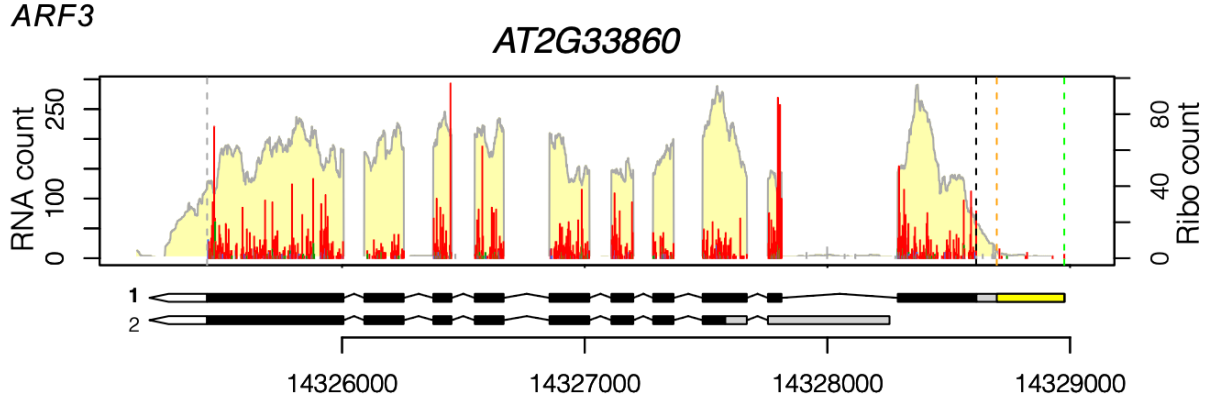




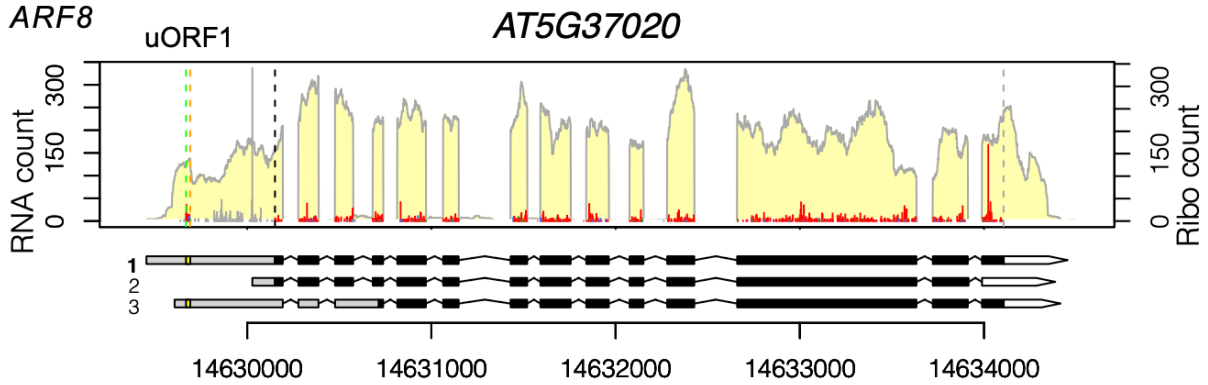
C. *ARF2*

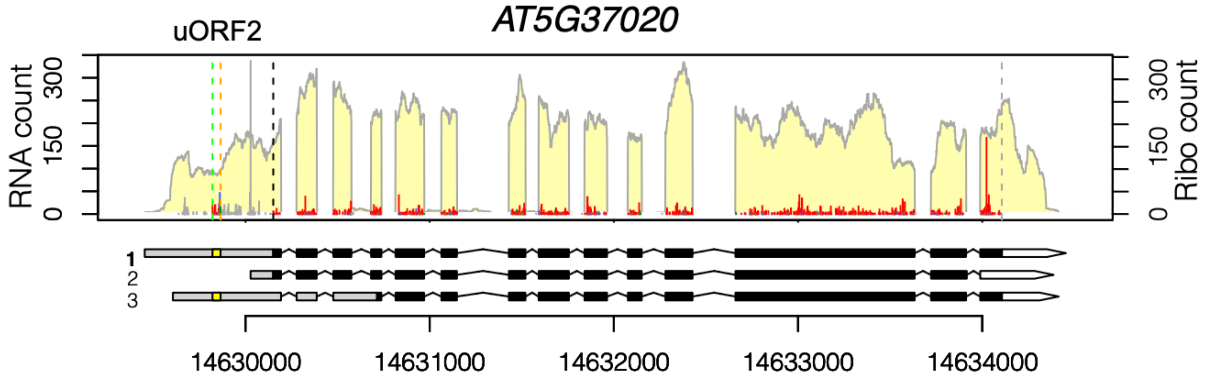


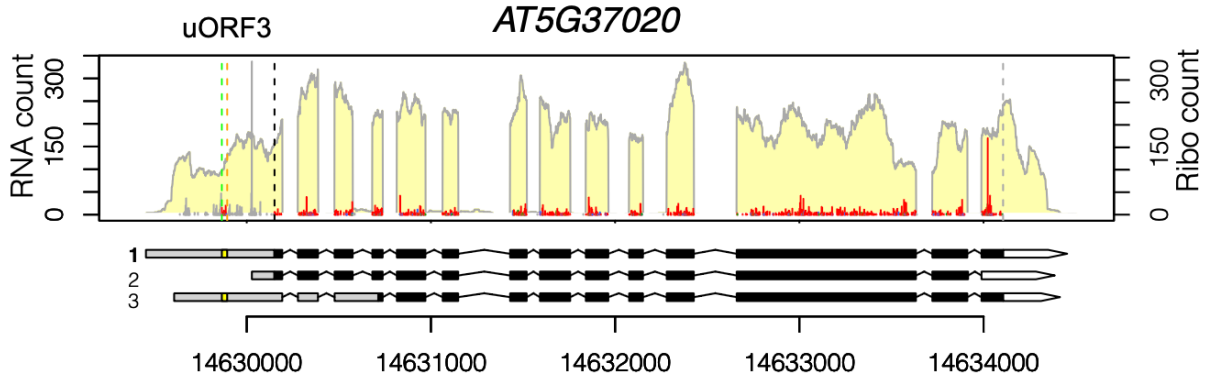
D. *ARF3*


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

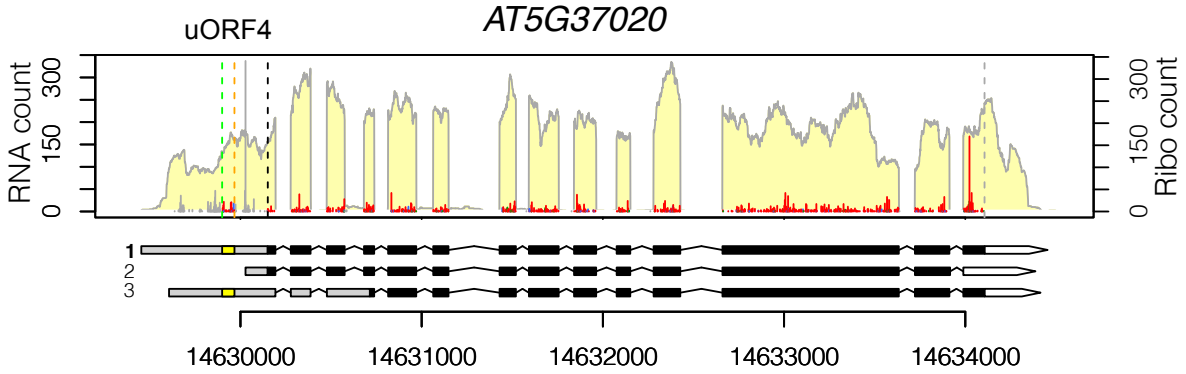
E. *ARF8*


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

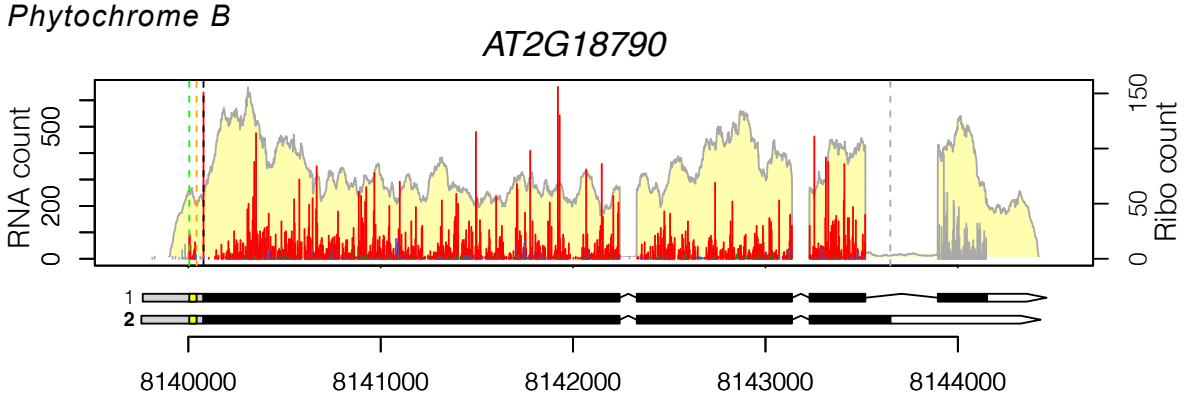

uORF2
AT5G37020
RNA count
0
150
300
Ribo count
0
150
300
1
2
3
14630000
14631000
14632000
14633000
14634000


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

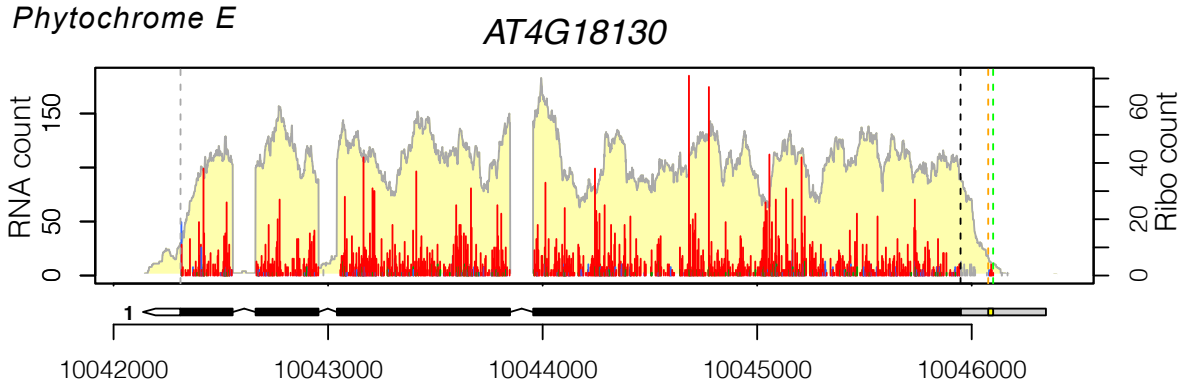
E. *ARF8* (continued)



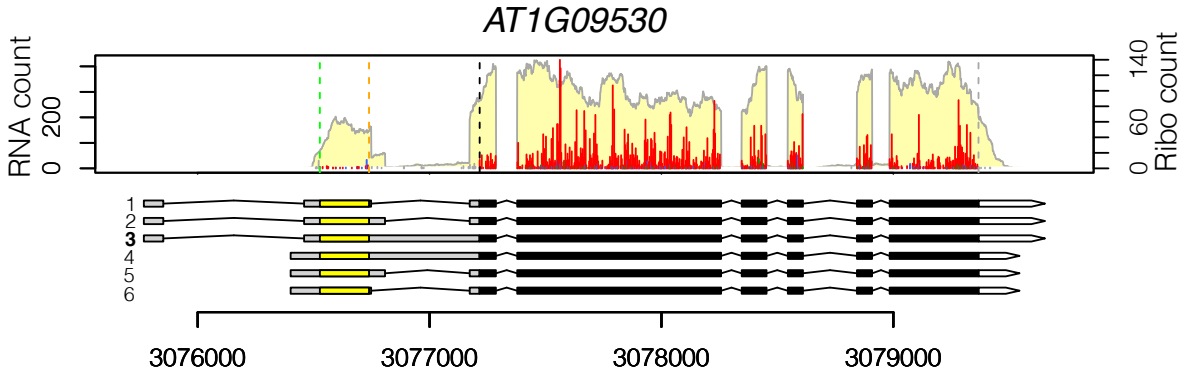
F. *Phytochrome B*



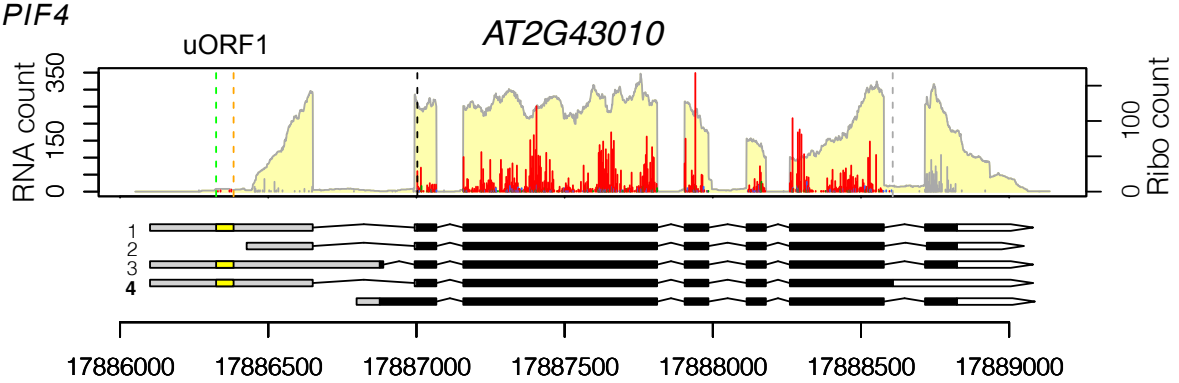
G. *Phytochrome E*

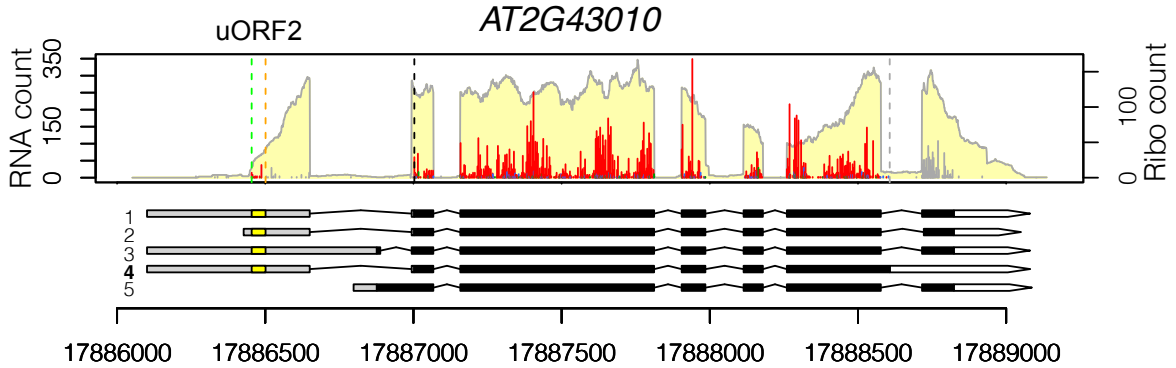


H. *PIF3*

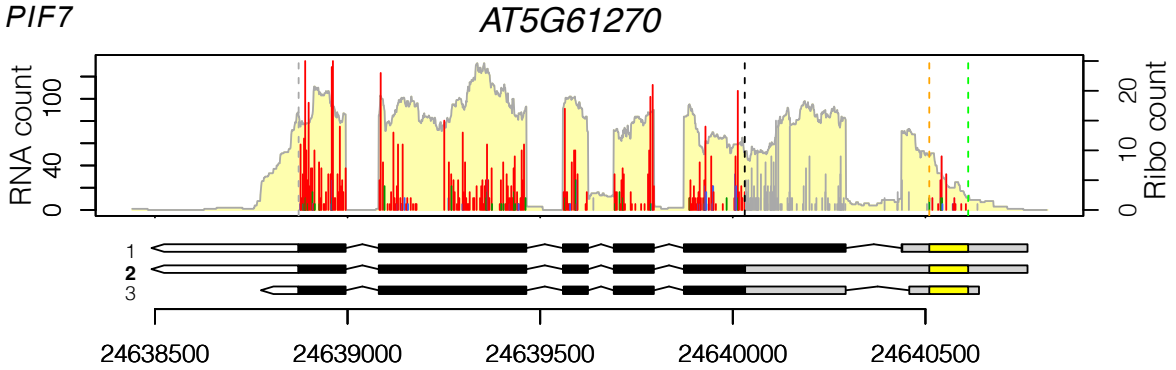


I. *PIF4*

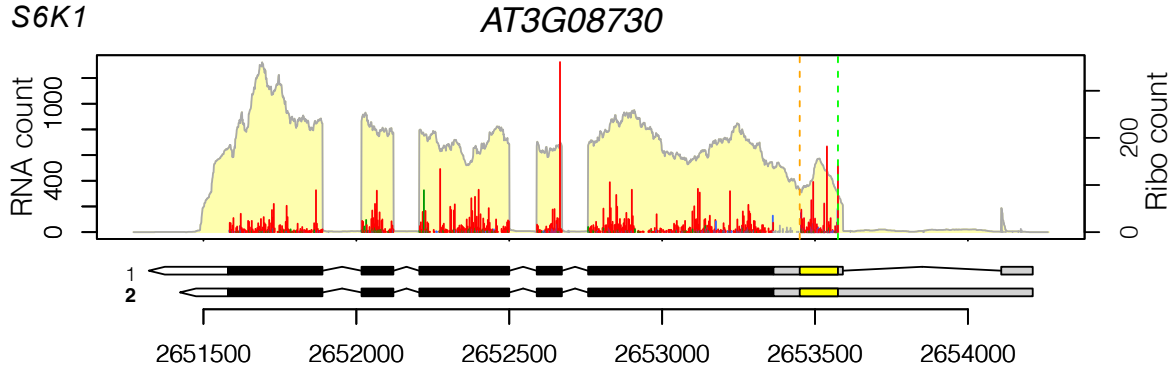




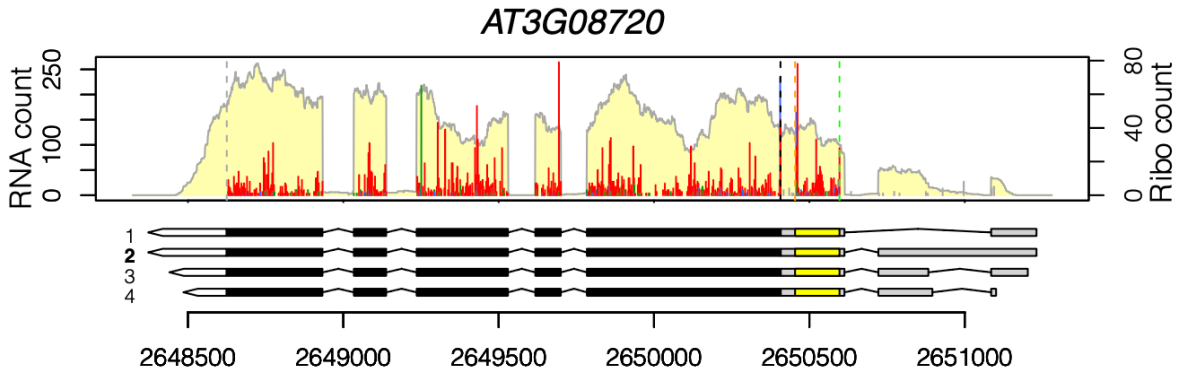
J. *PIF7*



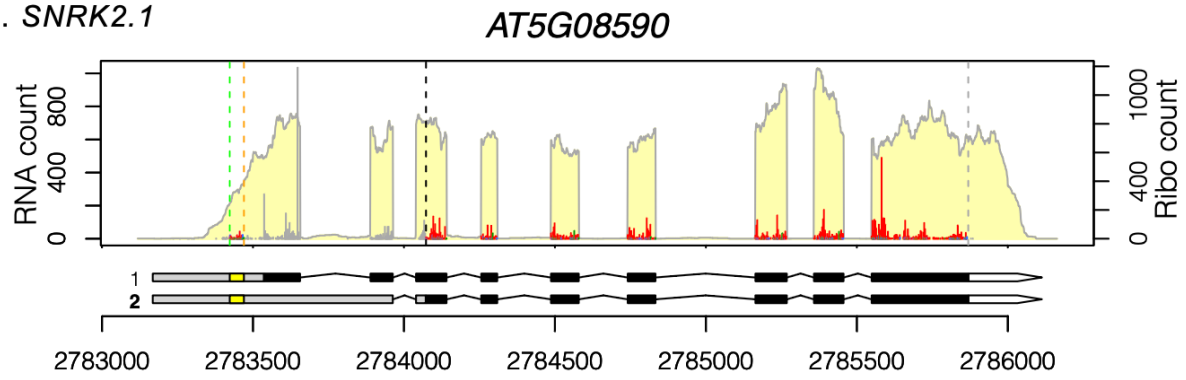
K. *S6K1*



L. *S6K2*



M. *SNRK2.1*



N. *SNRK2.5*

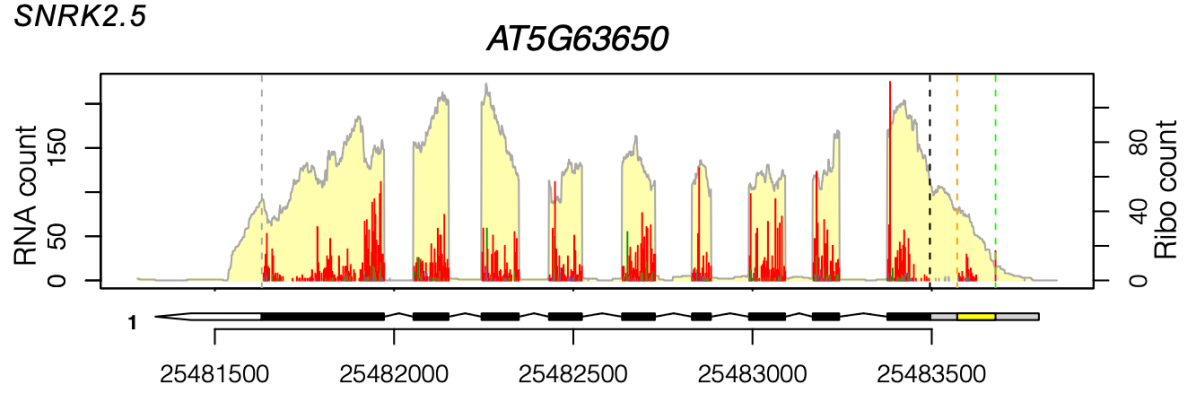


Figure S2. Examples of TuORFs in important regulatory genes. Data are presented as described in the Figure 1F legend. Isoform numbers are shown to the left of the gene models, and the isoform being considered is bolded. For genes that have multiple TuORFs identified (B, E, and I), the 3-nt periodicity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

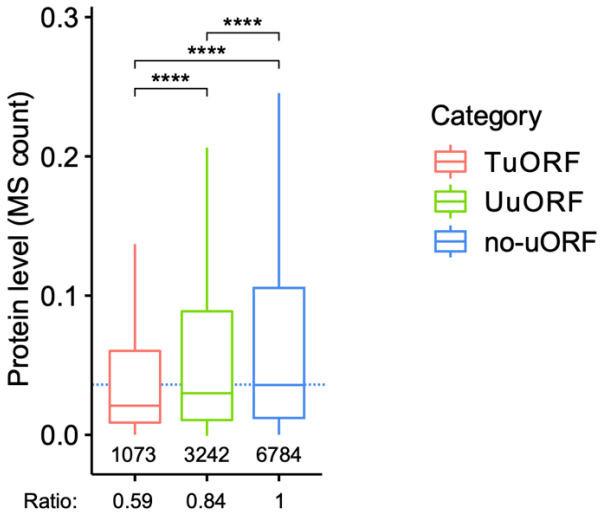


Figure S3. Protein abundance of the uORF-containing genes in Arabidopsis roots. The protein quantification data from Arabidopsis seedling roots were extracted from (Song et al., 2018). Within the boxplot, the dashed line marks the median level of no-uORF genes; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized to that of no-uORF genes. The statistical significance was determined as described in the Figure 2 legend.

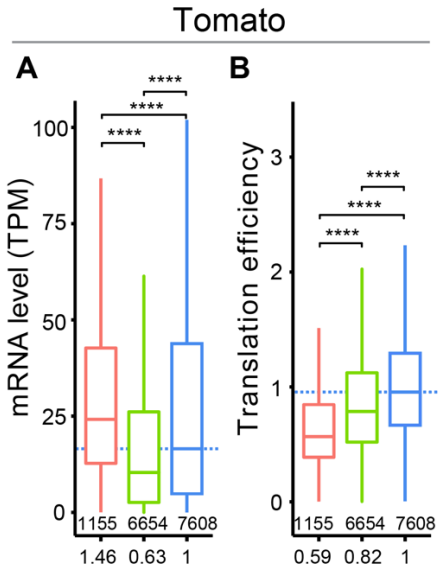


Figure S4. Tomato TuORF mRNAs are also associated with higher mRNA abundance and lower translation efficiency. The mRNA levels and translation efficiency of TuORF, UuORF, and no-uORF genes from a reanalysis of our previous tomato root data (Wu et al., 2019) were presented. The dashed lines mark the median level of the no-uORF genes.

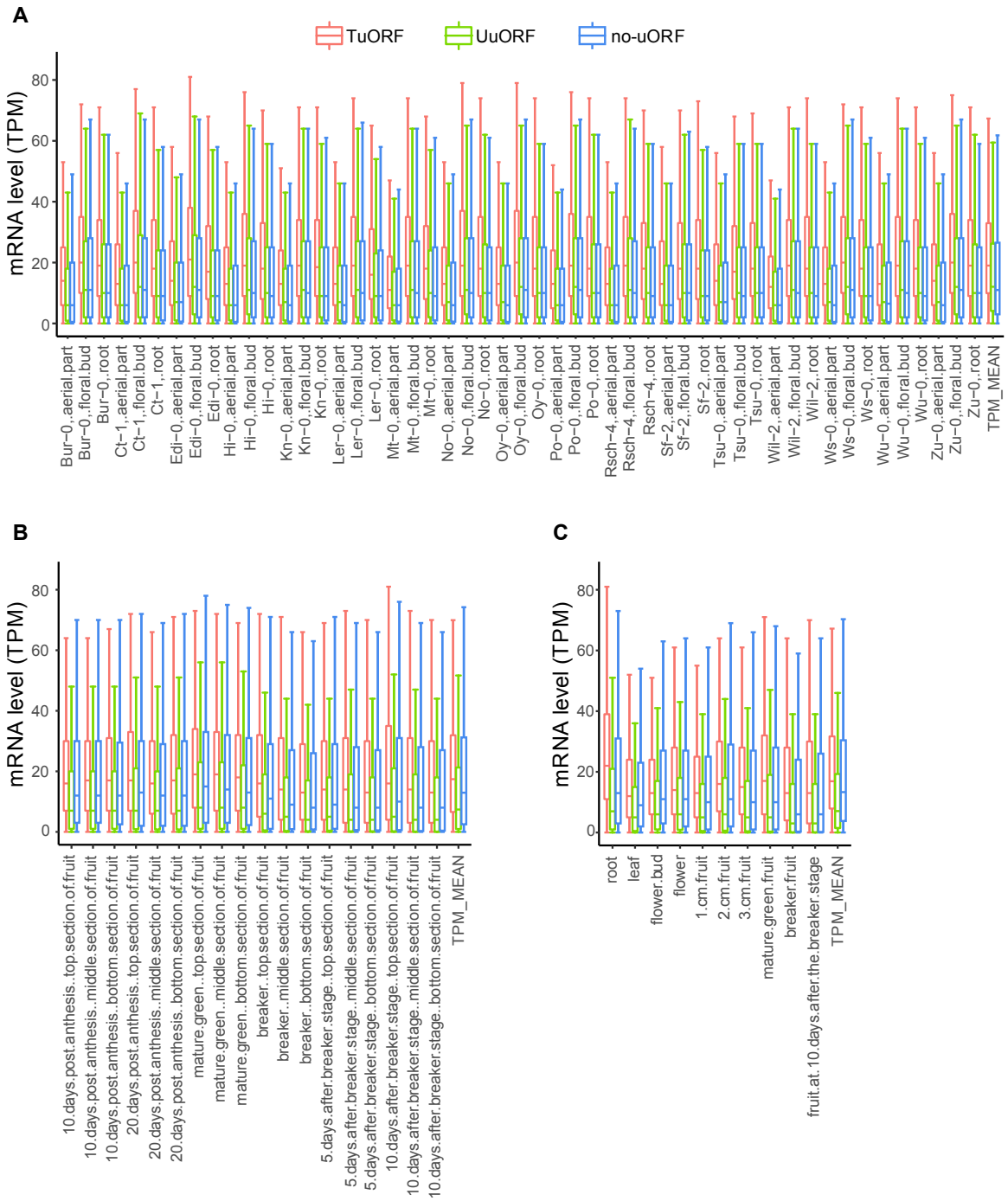


Figure S5. mRNA levels of uORF-containing genes in different tissues, developmental stages and ecotypes in Arabidopsis and tomato.

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 (Papatheodorou et al., 2020) (see METHODS).

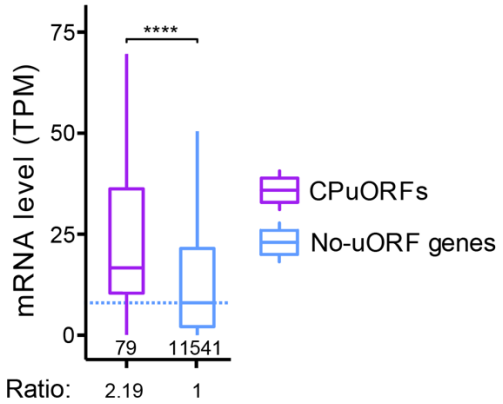


Figure S6. mRNA levels of CPuORF-containing genes compared with no-uORF genes. Within the boxplot, the dashed line marks the median level of no-uORF genes; the number of genes in each category is listed below the lower whisker. The ratios underneath the plots indicate the median of each group normalized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

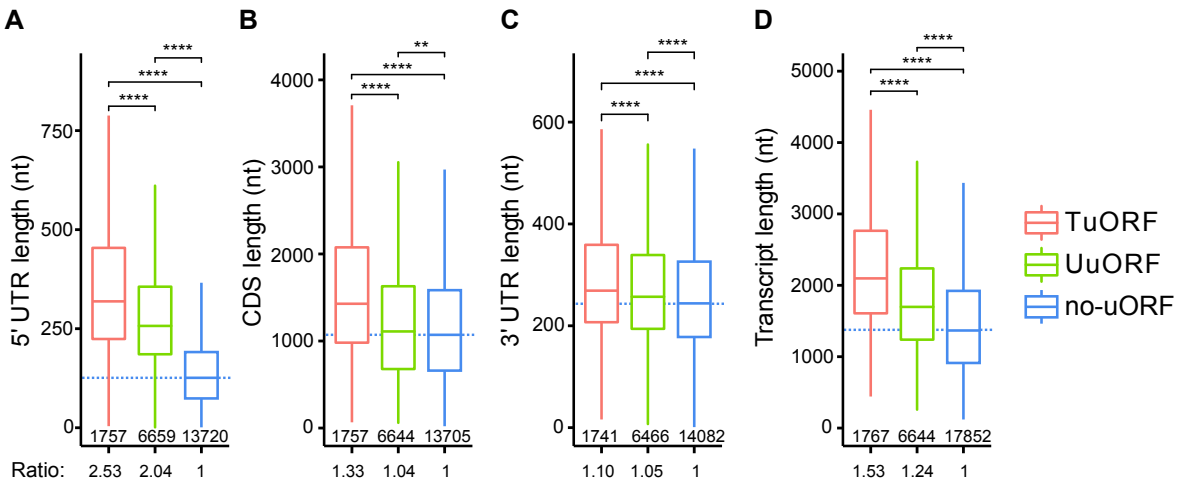


Figure S7. TuORF genes are larger than other genes

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.

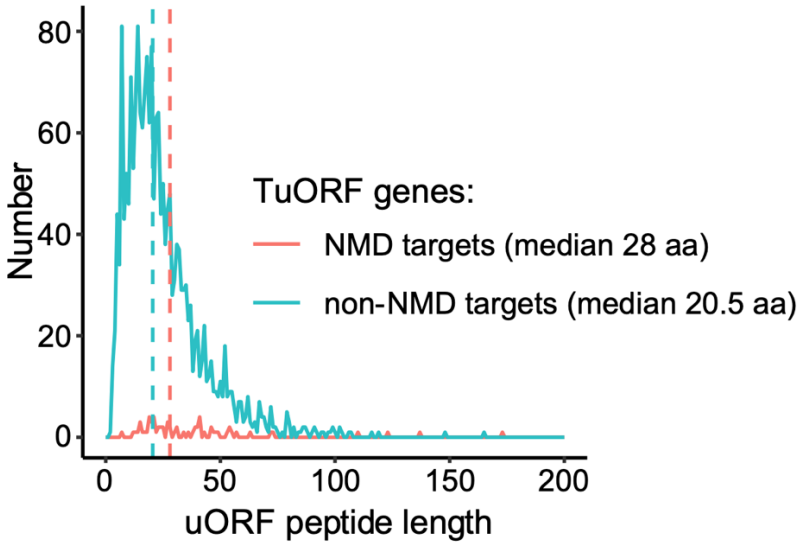


Figure S8. Distributions of uORF peptide lengths in NMD target and non-NMD target TuORF genes. Vertical dashed lines indicate the median value of each group.

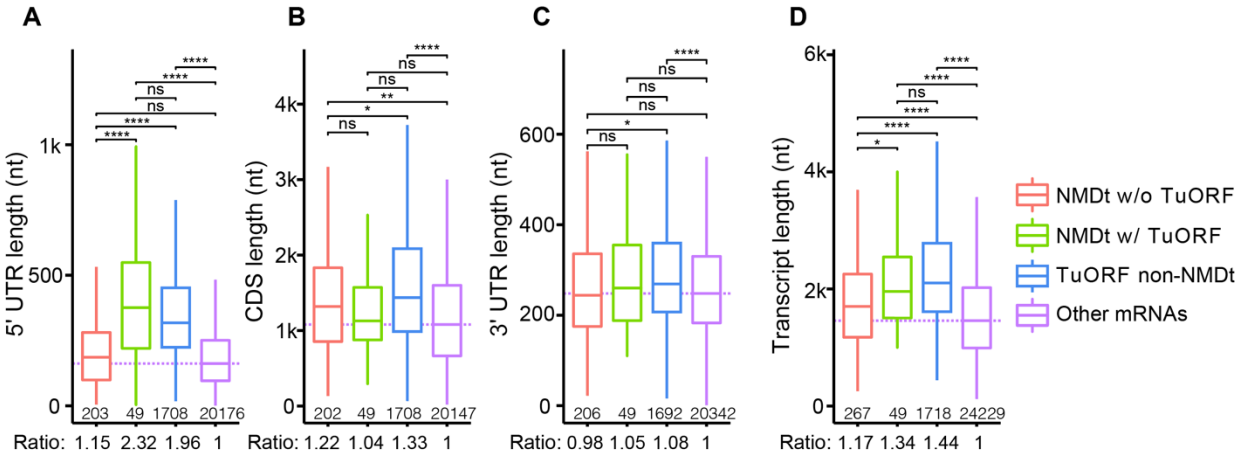


Figure S9. Lengths of UTRs and CDSs for NMD targets and TuORF mRNAs.
Lengths of the 5' UTRs (A), CDSs (B), 3' UTRs (C), and entire transcripts (D) of NMD targets (NMDt) without a TuORF, NMD targets with a TuORF, non-NMD TuORF mRNAs, and other mRNAs. Within the boxplots, the dashed lines mark the median 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.

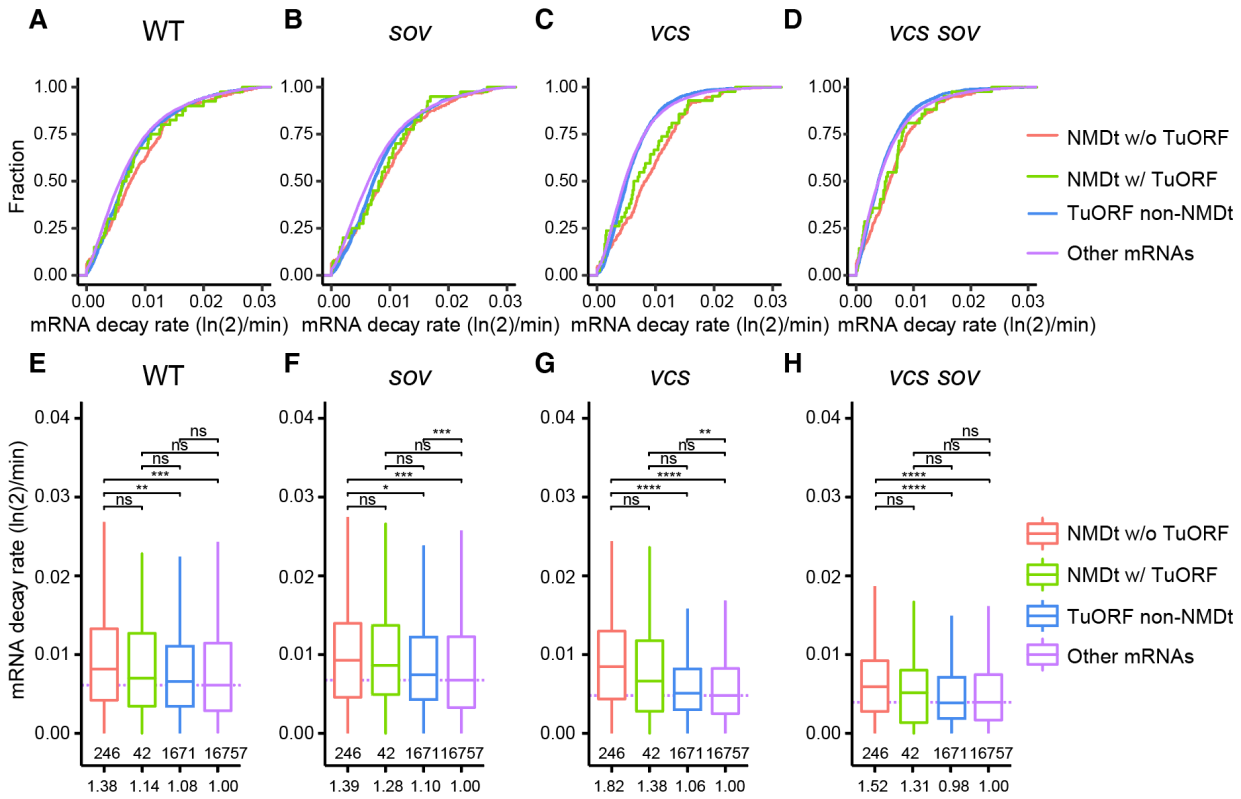


Figure S10. The mRNA decay rates of NMD targets and TuORF-containing mRNAs in different genetic backgrounds. Cumulative plots (A-D) and boxplots (E-H) showing the mRNA decay rates for NMD targets and TuORF mRNAs in wild type and mutants defective in either 5' to 3' decay (*vcs-7*), 3' to 5' decay (*sov*), or both (*vcs sov).* The mRNA decay rates were extracted from (Sorenson et al., 2018). 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:

Hsu, P.Y., Calviello, L., Wu, H.-Y.L., Li, F.-W., Rothfels, C.J., Ohler, U., and Benfey, P.N. (2016). Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **113**: E7126–E7135.

Papatheodorou, I. et al. (2020). Expression Atlas update: From tissues to single cells. Nucleic Acids Research **48**: D77–D83.

Song, G., Hsu, P.Y., and Walley, J.W. (2018). Assessment and Refinement of Sample Preparation Methods for Deep and Quantitative Plant Proteome Profiling. Proteomics **18**: e1800220.

Sorenson, R.S., Deshotel, M.J., Johnson, K., Adler, F.R., and Sieburth, L.E. (2018). 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 **115**: E1485–E1494.

Wu, H.-Y.L., Song, G., Walley, J.W., and Hsu, P.Y. (2019). Translational landscape in tomato revealed by transcriptome assembly and ribosome profiling. bioRxiv **181**: 367–380.