Actively translated uORFs reduce translation and mRNA stability independent of NMD in Arabidopsis

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

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

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

ORCIDs:

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

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

ABSTRACT

Upstream ORFs (uORFs) are widespread cis-regulatory elements in the 5' untranslated regions of eukaryotic genes. Translation of uORFs could negatively regulate protein synthesis by repressing main ORF (mORF) translation and by reducing mRNA stability presumably through nonsense-mediated decay (NMD). While the above expectations were supported in animals, they have not been extensively tested in plants. Using ribosome profiling, we systematically identified 2093 Actively Translated uORFs (ATuORFs) in Arabidopsis seedlings and examined their roles in gene expression regulation by integrating multiple genome-wide datasets. Compared with genes without uORFs, we found ATuORFs result in 38%, 14%, and 43% reductions in translation efficiency, mRNA stability, and protein levels, respectively. The effects of predicted but not actively translated uORFs are much weaker than those of ATuORFs. Interestingly, ATuORF-containing genes are also expressed at higher levels and encode longer proteins with conserved domains, features that are common in evolutionarily older genes.

Moreover, we provide evidence that uORF translation in plants, unlike in vertebrates, generally does not trigger NMD. We found ATuORF-containing transcripts are degraded through 5' to 3' decay, while NMD targets are degraded through both 5' to 3' and 3' to 5' decay, suggesting uORF-associated mRNA decay and NMD have distinct genetic requirements. Furthermore, we showed ATuORFs and NMD repress translation through separate mechanisms. Our results reveal that the potent inhibition of uORFs on mORF translation and mRNA stability in plants are independent of NMD, highlighting a fundamental difference in gene expression regulation by uORFs in the plant and animal kingdoms.

INTRODUCTION

Upstream open reading frames (uORFs) are common translational repressors that act in cis of eukaryotic protein-coding transcripts. Considering AUG as a translation start, 30-70% of genes contain at least one potential uORF in moss, Arabidopsis, fly, mouse and human [1–3]. According to the scanning model of mRNA translation, the 43S preinitiation complex binds the 5' cap and scans the 5' untranslated region (UTR) to search for an optimal initiation sequence [4–6]. 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 main ORF (mORF) [7]. The mORFs downstream of uORFs may get translated through leaky scanning or reinitiation [8,9]. In many organisms, uORF translation generally reduces mORF translation [2,10–13], and it is estimated to reduce protein production by 30-80% [10,14].

In addition, uORF translation is expected to trigger mRNA degradation, presumably through nonsense-mediated decay (NMD). NMD is an evolutionarily conserved regulatory mechanism to eliminate aberrant mRNAs, thus preventing the production of abnormal proteins [15–17]. Normally, during the first round of translation, the ribosome evicts the exon junction complexes (EJCs), which are deposited on mRNAs during splicing, on its way to the termination codon. Unevicted

EJCs caused by a premature termination codon (PTC) or a long 3' UTR are predicted to activate NMD. There are two phases in NMD: substrate recognition and substrate degradation. In plants, the first phase centers on UP-FRAMESHIFT SUPRESSOR1 (UPF1) and the associated proteins UPF2 and UPF3; the second phase involves SUPPRESSOR WITH MORPHOLOGICAL EFFECT ON GENITALIA7 (SMG7) [18]. Conventionally, putative NMD targets have been identified based on transcripts that overly accumulate in the NMD mutants [19–22]. However, NMD mutants affect numerous aspects of plant physiology, and some alleles display severe growth retardation and lethality due to constitutive immune responses [23–28]. Moreover, UPF1 has been reported to have multiple functions in plants, such as alternative splicing regulation, translational control and NMD [29,30]. Thus, some transcripts that are upregulated in NMD mutants might be caused by secondary effects of NMD. To identify direct NMD targets, *upf1* and *smg7* were introduced into *pad4*, in which the immune response is inactivated by the *pad4* mutation [28,30]. The 333 consensus upregulated transcripts in *upf1 pad4* and *smg7 pad4* were defined as high-confidence NMD targets [30]. Moreover, it was shown these NMD targets are translationally regulated by UPF1 and their turnover involves decapping and 5' to 3' decay [30].

As uORF translation mimics a transcript with an unusually long 3' UTR, uORF-containing mRNAs are predicted to be NMD targets [17,31]. Consistent with this expectation, genome-wide studies in zebrafish, mouse, and human have found that transcripts with translated uORFs have lower mRNA levels globally [2,10]. Moreover, transcripts contain translated uORFs in mouse embryonic stem cells have increased mRNA levels in *upf1* mutant, supporting these transcripts are targets of NMD [32]. In plants, transcripts upregulated in *upf1* and *upf3* are enriched for those possess predicted uORFs [19,21]. However, examples exist that demonstrate some uORFs do not trigger NMD in yeast, animals and plants [20,33–36]. It has been suggested that NMD might only control a fraction of uORF-containing mRNAs in plants [37–39], such as those containing uORFs

with > 35 codons [37]. It is unknown whether some uORF-containing mRNAs escape NMD or whether uORF-containing mRNAs are generally not targets of NMD. Overall, it remains unclear how translated uORFs affect mRNA stability and the role of NMD in this process in plants.

Although many uORFs have been predicted based on the nucleotide sequences in eukaryotic genomes [3], an outstanding question remains what fraction of uORFs are translated to suppress mORF translation and mRNA stability. Recently, ribosome profiling, i.e., deep sequencing of ribosome footprints, has been exploited to study genome-wide mRNA translation [40,41]. Several studies in Arabidopsis have deployed various strategies to identify translated uORFs [11,42–46]. For example, over 3000 uORFs contain at least one ribosome footprint in etiolated seedlings [11]. As ribosome footprints alone do not guarantee active translation, a more stringent criterion, 3-nucleotide (3-nt) periodicity, which corresponds to translating ribosomes deciphering 3 nts per codon, has been exploited to identify Actively Translated uORFs (ATuORFs) [40,47–49]. Our previous work and a recent study using RiboTaper [47], which computes the statistical significance of 3-nt periodicity, have uncovered 187-1378 ATuORFs in Arabidopsis [43,45].

Here we further improved our methodology and identified 2093 ATuORFs in Arabidopsis seedlings. We systematically examined the assumed roles of ATuORFs in mORF translation and mRNA decay by integrating multiple genome-wide datasets, and we explored which transcript properties are associated with ATuORFs. Consistent with the expectations, we found ATuORF mRNAs have lower translation efficiency (TE), shorter mRNA half-lives and result in lower protein levels. However, we found ATuORF mRNAs in plants, unlike in animals, have higher steady-state mRNA levels on average. The faster decay and higher mRNA abundance suggest these genes are highly transcribed. Moreover, we showed that the accelerated decay of ATuORF mRNAs is genetically independent of the NMD pathway. Our results reveal

unexpected features and regulation of ATuORF genes in Arabidopsis and suggest that plants and animals have evolved different uORF-dependent mechanisms of gene regulation.

RESULTS

Actively translated uORFs identified via an improved ribosome profiling method

Previously we optimized ribosome profiling to reach high 3-nt periodicity, and we exploited this periodic feature of mRNA translation to identify actively translated ORFs in Arabidopsis [43]. Although the data revealed many unannotated translation events, the number of ATuORFs uncovered was low, presumably due to an insufficient Ribo-seq coverage within uORFs. As uORFs are relatively short, low Ribo-seq coverage limits the statistical power to identify translated ORFs based on significant 3-nt periodicity. Here we further improved the coverage of Ribo-seq by minimizing the potential loss of ribosome footprints during the purification process (see Materials and Methods for details).

In addition to high correlation between samples (Fig. S1A), our new Ribo-seq data displayed expected high-quality characteristics, including strong 3-nt periodicity (92% in-frame reads), high enrichment for coding sequences (CDSs), and characteristic ribosome footprint lengths (Fig. 1A-C). The improved coverage is evident by examining the number of unique ribosome footprints detected globally. At a sequencing depth of 20 million reads, the number of unique ribosome footprints in our new data was 4.01-fold of our previous dataset (Fig. 1D). The improvement in coverage is also evident in individual transcript profiles within both uORFs and mORFs. For example, an ATuORF in *REPRESSOR OF GA1* (*RGA1*) was identified in our current data but missed in our previous study despite having similar mRNA levels (Fig. S1B).

The high-coverage data improved the number of ATuORFs uncovered. In total, 2093 ATuORFs were identified among 1768 genes in the current dataset based on strong 3-nt periodicity (Fig. 1E-G and Table S1A). This result is in contrast to the 93 and 138 ATuORFs identified in our previous Arabidopsis shoot and root data using the same statistical analysis and

the annotation (Fig. 1E and Table S1B-C). The new data also detected translation from more Conserved Peptide uORFs (CPuORFs) and annotated coding sequences (mORFs of protein-coding genes)(Fig. 1E and Table S1D-F). These results showed that our enhanced-coverage Ribo-seq data dramatically improved the identification of ATuORFs.

We found that many ATuORF-containing genes encode important regulators of plant growth and environmental responses. For example, this list included well-characterized transcription factors in GA signaling (RGA1), the circadian clock (TIMING OF CAB EXPRESSION 1), root development (SCARECROW), auxin signaling (AUXIN RESPONSE FACTOR2/3/8) and light signaling (PHYTOCHROME INTERACTING FACTOR3/4/7) [50–57]. The list also included extensively studied protein kinases, such as Phytochrome B and E, ribosomal-protein S6 kinase (S6K1/2), and SNF1-related protein kinase (SNRK2.1/2.5) (Fig. S1B, S2A-N) [58–63]. Previous analyses of predicted uORFs found transcription factors and protein kinases are overrepresented among uORF-containing genes [1,64]. Consistent with the predictions, functions related to transcription factors and protein kinases/phosphatases were enriched in our Gene Ontology (GO) term analysis of ATuORF-containing genes (Table S2). Similar results were observed in our tomato ATuORF gene list [12], supporting that ATuORFs control important cellular regulators in plants.

ATuORF-containing genes were highly expressed but associated with lower TE, shorter mRNA half-lives, and lower protein abundance

To understand how uORF translation affects gene expression, we compared mRNA abundance and TE in ATuORF-containing genes (hereafter, ATuORF genes), genes containing sequence predicted but Poorly Translated uORFs (hereafter, PTuORF genes), and genes containing an annotated 5' UTR but no uORFs (hereafter "no-uORF genes"). Note that PTuORFs may be translated at a low frequency but fail to pass the threshold to be considered actively translated based on significant 3-nt periodicity. We also compared mRNA half-lives and

protein levels, determined by metabolic labeling and quantitative proteomics, respectively, in the three groups of genes using published datasets [65,66] to study the global effects of uORFs on mRNA stability and protein abundance in Arabidopsis (Fig. 2A-D).

While uORF translation is associated with lower steady-state mRNA abundance in vertebrates [2,10], surprisingly, we found ATuORF genes have higher steady-state mRNA levels than the other two groups in Arabidopsis (Fig. 2A). However, ATuORF genes also have shorter mRNA half-lives (Fig. 2C). The higher steady-state mRNA levels and shorter mRNA half-lives of ATuORF genes suggest that they are highly transcribed. Higher steady-state mRNA levels in ATuORF genes were also observed in our previous tomato root data [12] (Fig. 2E), suggesting that this phenomenon could be prevalent in plants. In addition, we found the ATuORF genes have higher expression levels in 128 other RNA-seq datasets in diverse tissues, developmental stages, and ecotypes in Arabidopsis and/or tomato (Fig. 2G, Fig. S3). Consistent with our observations, a recent report showed no reductions in mRNA levels in transcripts containing putative translated uORFs, in which both ATuORFs and PTuORFs might be included, in Arabidopsis and tomato [39]. Collectively, these unexpected results reveal plants and animals implement ATuORFs differently in gene regulation.

One uninspiring explanation for the higher mRNA abundance of ATuORF genes is that ATuORFs are more likely to be detected in transcripts that are highly abundant. To explore this possibility, we first examined CPuORFs, which were identified based on the evolutionary conservation of the uORF peptide sequences and independent of transcript abundance. We found the mRNA abundance of CPuORF genes was also higher than that in no-uORF genes (Fig. S4), consistent with the observation that genes with ATuORFs are expressed at higher levels. Second, many PTuORF genes were also highly expressed (Fig. 2H). In fact, the distributions of mRNA levels of ATuORF and PTuORF genes completely overlapped, indicating that mRNA abundance is not a major determinant of whether a uORF is detected as translated (Fig. 2H). Finally, ATuORFs were associated with lower TE, shorter mRNA half-lives, and less

protein than PTuORF genes, indicating that ATuORFs are functionally associated with potent inhibitory effects (Fig. 2A-F). Together, our results support that the ATuORFs were not identified simply due to higher mRNA abundance; rather, they are selectively translated to regulate the translation and mRNA stability of ATuORF genes.

Interestingly, even though ATuORF genes in Arabidopsis start with 51% more mRNA than no-uORF genes, they produce 41-43% less protein, on average (Fig. 2A, 2D, and S5), likely due to the combined effects of 38% and 14% reduction in translational efficiency and mRNA half-lives, respectively. These observations indicate multilayer post-transcriptional regulation is involved in repressing protein synthesis from ATuORF genes at a global level. Notably, in the four parameters examined in Arabidopsis, PTuORF genes always have the levels between ATuORF genes and no-uORF genes (i.e., 4% increase in mRNA levels, 11%, 6%, and 21% reduction in translation efficiency, mRNA half-lives, and protein abundance). These results imply that at least some PTuORFs might be translated at low levels and have less impact on gene expression regulation than ATuORFs.

ATuORF genes are larger on average

One interesting question is whether there are structural differences among ATuORF genes, PTuORF genes and no-uORF genes. One possibility is that a gene with a longer 5' UTR would have a higher chance of containing one or more ATuORFs. Strikingly, we found ATuORF genes have significantly longer transcripts, as well as longer 5' UTRs, CDSs and 3'UTRs than PTuORF genes; and no-uORF genes have shortest lengths in all categories examined (Fig. 3A-D). The difference was most dramatic in the 5' UTR length (Fig. 3A). These results support ATuORFs are more likely to be found on longer 5' UTRs and reveal ATuORF genes tend to encode larger proteins (Fig. 3B). The enrichment of ATuORFs genes for transcription factors, protein kinases, and related functions in GO term functions (Table S2) also implies ATuORF genes tend to encode proteins with conserved domains. Evolutionarily, older genes are

generally longer, expressed at higher levels, and have protein sequences that are under stronger purifying selection [67,68]. Our observations that ATuORFs have higher expression levels, longer transcript lengths and enrichment for transcription factors and protein kinases (Table S2) inspire an intriguing hypothesis that perhaps older genes are more likely to implement uORFs to regulate their mRNA translation and degradation.

ATuORF genes have distinct expression patterns from those of NMD targets

Conventionally, uORF translation is expected to trigger NMD [15–17,31]. In vertebrates, transcripts containing translated uORFs are associated with a lower steady-state mRNA abundance [2,10] and have increased mRNA levels in a NMD mutant [32], consistent with being potential targets of NMD. However, there are also examples suggesting that certain uORFs do not trigger NMD in plants and animals [20,33,34,36]. Despite having higher mRNA abundance (Fig. 2A), ATuORF genes in Arabidopsis also have shorter mRNA half-lives (Fig. 2C); therefore, we investigated whether ATuORF-containing transcripts are still subject to NMD.

If ATuORFs trigger NMD, we expect ATuORF mRNAs would be upregulated in NMD mutants and their expression patterns would follow those of NMD targets. Among the 333 high-confidence NMD targets that were upregulated in both *upf1 pad4* and *smg7 pad4* [30], 248 genes with an annotated 5' UTR were expressed in our data (Table S3A). However, of 1748 ATuORF genes, only 49 were identified as high-confidence NMD targets (Table S3B) (corresponding to 2.8% of ATuORF-containing genes and 20% of NMD targets). These results indicate that ATuORFs generally do not trigger NMD; rather, a good fraction of NMD targets are regulated by ATuORFs.

To elucidate the role of ATuORFs and NMD in gene expression regulation, we examined the expression patterns of these genes (Fig. 4). While both NMD targets and ATuORF genes are associated with shorter mRNA half-lives than other genes (Fig. 4C), we found NMD targets have overall lower mRNA abundance and protein levels than ATuORF genes (Fig. 4A and 4D).

Importantly, ATuORF-containing NMD targets have significantly lower TE than either NMD targets without ATuORFs or ATuORF-containing non-NMD target genes (Fig. 4B). This additive repression of TE for ATuORF-containing NMD targets suggests that NMD and ATuORFs suppress translation through separate mechanisms. This result also supports that ATuORFs generally do not trigger NMD.

Genes controlled by both NMD and ATuORFs might encode critical regulators, and their protein levels require precise regulation. This list included key enzymes in gibberellin and polyamine biosynthesis, and many transcription factors (Fig. 1F-G, and Table S3B). For example, one of the three conserved uORFs in *SUPPRESSOR OF ACAULIS 51* (*SAC51*) was discovered in a genetic screen looking for suppressors of the dwarf phenotype of *acaulis5* mutant [69]. When the particular uORF in *SAC51* was mutated, it increased the transcript level of *SAC51*, which encodes a bHLH transcription factor that was proposed to activate the genes required for stem elongation.

Relationship between ATuORF length and NMD

A prior study suggested uORFs that encode a peptide longer than 35 amino acids could trigger NMD in Arabidopsis [37]. Interestingly, the uORF whose mutation increased the transcript level of *SAC51* [69] encodes 53 amino acids, and this gene was identified as a high-confidence NMD target [30]. As 2.8% of ATuORF genes in our data are also NMD targets, we examined whether ATuORF length is associated with NMD targetability. We found that on average, the ATuORF-containing NMD targets have longer uORF peptides than those of non-NMD target ATuORF genes (median length 28 a.a. vs. 20.5 a.a., p= 9.874e-07, Wilcoxon signed rank test). However, the distributions of the uORF peptide lengths in NMD targets and non-NMD targets largely overlapped (Fig. 4E), indicating that ATuORF length alone is insufficient to determine whether the ATuORFs can trigger NMD in these transcripts.

Different genetic requirements for the mRNA decay of ATuORF genes and NMD targets

mRNA decay can occur in the 5' to 3' or 3' to 5' directions, which are regulated by distinct pathways [70]. A genome-wide survey of mRNA decay rates in wild-type plants, single mutants defective in either 5' to 3' decay or 3' to 5' decay, as well as their double mutant plants, has been conducted [71]. In this experiment, the *vcs-7* mutant, a null mutant of *VARICOSE*, which encodes a scaffold protein for the mRNA decapping complex, was used to examine 5' to 3' mRNA decay; the *sov* mutant, defective in *SUPPRESSOR OF VARICOSE*, which encodes a ribonuclease II, was used to examine 3' to 5' decay; and *vcs/sov* double mutant plants were defective for mRNA decay activities in both directions [72–75]. We exploited this published dataset to study the genetic requirements for the mRNA decay of ATuORF genes, NMD targets, ATuORF-containing NMD targets and other genes (Fig. 5). Similar to genes that are non-NMD targets and no-ATuORF (Fig. 5A and E), the mRNA decay rates for ATuORF genes were dramatically reduced in the *vcs* mutant and the *vcs/sov* double mutants (Fig. 5B and F), indicating that decapping and 5' to 3' decay are critical for degrading these transcripts.

In contrast, the mRNA decay of NMD targets without ATuORFs was only significantly reduced in the *vcs/sov* double mutant, suggesting that NMD targets are efficiently degraded by both the 5' to 3' and 3' to 5' pathways (Fig. 5C and G). Although the small sample size of NMD targets containing ATuORFs resulted in statistical insignificance (Fig. 5D and H), the decay rates in this group in the four genetic backgrounds shared similar patterns as those of NMD targets without ATuORFs (Fig. 5C and G). These results imply that the NMD feature is epistatic to the ATuORF feature to rule the mRNA decay of the ATuORF-containing NMD targets. Most importantly, our results reveal that ATuORF genes and NMD targets have distinct genetic requirements for mRNA decay and suggest existence of an ATuORF-mediated mRNA decay pathway that is dependent on mRNA decapping in Arabidopsis.

DISCUSSION

In this study, we identified ATuORFs based on improved Ribo-seq and systematically examined their presumed roles in gene expression. Our results support that uORFs negatively regulate mORF translation and mRNA stability. Consistent with this inhibition, ATuORF mRNAs result in lower protein levels. However, our study reveal several unexpected observations: 1) the ATuORF mRNAs in plants have higher steady-state mRNA levels and faster degradation, 2) ATuORF mRNAs are longer and encode larger proteins, 3) the degradation of ATuORF mRNAs is independent of NMD, and 4) ATuORFs and NMD repress translation through different mechanisms.

The higher steady-state mRNA abundance in ATuORF genes in plants is different from the observations in animals [2,10], suggesting plants and animals have different design logics for implementing uORFs in gene expression regulation. Moreover, the overall higher mRNA abundance and faster mRNA decay rates of these genes in plants indicate that ATuORF genes have higher transcription rates. To our knowledge, the transcription rates of ATuORF genes have not been globally examined in plants, and it remains to be determined whether these high transcription rates are conserved in other species.

In addition to higher expression levels, we found ATuORF genes are generally longer in transcript length and the subregions, and they encode larger proteins (Fig. 3). Moreover, the enrichment of ATuORF genes for transcription factors and protein kinases (Table S2) suggests that they contain highly conserved protein domains. These properties coincide with the observations that older genes, on average, are longer, expressed at higher levels, and subject to stronger purifying selection [67,68]. This raised an interesting question as to whether older genes are more likely to use uORFs to control their protein levels through mRNA degradation and translation. Higher mRNA abundance may allow ATuORF mRNAs to be highly translated under certain conditions that require a high level of specific proteins. For example, it was concluded that derepression of downstream translation is a general mechanism of uORF-mediated

expression under stress [76]; normally these transcripts would be quickly degraded to avoid over-accumulation of their proteins. Thus, ATuORFs could provide fast and dynamic regulation of mORF translation. Future work monitoring the translation of ATuORFs and their mORFs under different conditions will help decipher the regulation and physiological roles of specific ATuORFs.

In this study, we show four pieces of evidence that generally ATuORF mRNAs are not NMD targets. First, less than 3% of ATuORF genes are high-confidence NMD targets [30]. Second, the steady-state mRNA levels of NMD targets are much lower than those of ATuORF mRNAs (Fig. 4A). Moreover, the NMD targets with ATuORF(s) further reduce translation efficiency compared to the NMD targets without ATuORF(s) or the non-NMD target ATuORF mRNAs (Fig. 4B). Finally, while NMD targets are efficiently degraded through both 5' to 3' and 3' to 5' decay, ATuORF mRNAs rely on 5' to 3' decay (Fig. 5). It is unlikely that such a large number of transcripts escape NMD altogether; rather, our results suggest ATuORFs generally do no trigger NMD in plants.

It is known there is an important interplay between translation efficiency and mRNA stability. Two models have been proposed: the "stalled ribosome-triggered decay model" predicts ribosome stalling or slowing promotes mRNA degradation, while the "translation factor-protected model" predicts that active translation promotes mRNA stability as the translation initiation factors protect the mRNAs from the decapping complex [77–82]. Consistent with the first model, several CPuORF-containing mRNAs in Arabidopsis have been reported to have strong ribosome stalling, and the predicted endogenous cleavages upstream of the stalled ribosomes were observed [83,84]. While most of the ATuORFs we identified did not display clear stalling patterns (e.g., Fig. 1F and G, S1B, S2A-N), it remains possible ribosomes may have lower translation speed, on average, during uORF translation. As ribosomes will quickly approach termination after initiation due to the short length of uORFs, and since both initiation and termination are major time-consuming steps in translation [85,86], the average translation speed

per codon for uORFs is likely to be slower than that of longer ORFs. Alternatively, if cells mainly monitor the translation efficiency of the mORF, then the poorer mORF translation caused by uORFs could explain the faster degradation of ATuORF mRNAs predicted by this model. On the other hand, since ribosomes actively engage with ATuORFs, the “translation factor-protected” model does not fit ATuORF mRNAs, especially because their degradation relies on decapping and 5’ to 3’ decay in plants. Further investigation into ATuORF mRNA degradation will help elucidate the role of uORFs in regulating mRNA stability.

MATERIALS AND METHODS

Plant growth conditions and lysate preparation

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

Plant lysates were prepared as previously described. Briefly, per 0.1 g of grounded tissue power was resuspended in 400 μL of lysis buffer (100 mM Tris-HCl [pH 8], 40 mM KCl, 20 mM MgCl2, 2% [v/v] polyoxyethylene [10] tridecyl ether [Sigma, P2393], 1% [w/v] sodium deoxycholate [Sigma, D6750], 1 mM dithiothreitol, 100 μg/mL cycloheximide [Sigma, C4859], 100 μg/mL chloramphenicol [Sigma R4408], and 10 units/mL DNase I [Epicenter, D9905K]). The lysates were spun at 3,000 g for 3 min, and the supernatant was transferred to a new tube and

subsequently centrifuged at 20,000 g for 10 min. The supernatant was transferred to a new tube and the RNA concentration was determined with 10x dilutions using the Qubit RNA HS assay (Thermo Fisher Scientific; Q32852). Aliquots of 100 µL and 200 µL of the lysates were made, and they were flash frozen in liquid nitrogen and stored at -80°C until further processing.

Ribo-seq library construction

Ribosome footprints were obtained using 200 µL of the lysates described above, and sequencing libraries were constructed according to our previous method with the following modifications: briefly, after RNase I digestion (50 units nuclease per 40 µg of RNA; the nuclease was included in TruSeq Mammalian Ribo Profile Kit, illumina, RPHMR12126) and passing through a size exclusion column (illustra MicroSpin S-400 HR Columns; GE Healthcare; 27-5140-01), RNA > 17 nt was isolated with RNA Clean & Concentrator-25 kit (Zymo Research, R1017) and separated on 15% urea-TBE gels (Invitrogen, EC68852BOX). Gel slices between 28 and 30 nt were isolated, and the RNAs were purified as previously described. Next, rRNA depletion was performed using RiboZero Plant Leaf kit (Illumina, MRZPL1224) in one quarter of the recommended reaction volume. Ribo-seq libraries were then constructed using the TruSeq Mammalian Ribo Profile Kit (illumina, RPHMR12126) as previously described with 9 cycles of PCR amplification. Libraries with equal molarity were pooled and sequenced on Hi-Seq 4000 using single-end 50-bp sequencing.

RNA-seq library construction

Total RNA greater than 200 nt was purified using 100 µL of the lysates above as previously described with RNA Clean & Concentrator-25 kit (Zymo Research, R1017). RNA integrity was evaluated using a Bioanalyzer (Agilent) RNA pico chip and RNA integrity numbers (RINs) ranging from 7.2 to 7.7 were observed among the samples. A total of 4 µg of RNA per sample was subjected to rRNA depletion using RiboZero Plant Leaf kit (Illumina, MRZPL1224) following

the manufacturer's recommendations. Then, 100 ng of rRNA-depleted RNA was fragmented to around 200-nt long based on the RIN reported by the Bioanalyzer, and strand-specific sequencing libraries were made using NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs, E7760S) with 8 cycles of amplification. Libraries of equal molarity were pooled and sequenced on Hi-Seq 4000 using paired-end 100-bp sequencing.

Data pre-processing and analysis

Data pre-processing and analysis were performed similarly to that previously described [43], except the Araport11 annotation was used in this study. Briefly, for Ribo-seq libraries, the adaptor (AGATCGGAAGAGCACACGTCT) was clipped with fastx_clipper (FASTX toolkit v0.0.14) (http://hannonlab.cshl.edu/fastx_tool- kit/). For both RNA-seq and Ribo-seq, we used Bowtie2 (v2.3.4.1) [87] to remove rRNA/tRNA/snRNA/snoRNA sequences. Both the RNA-seq and Ribo-seq reads were mapped to the transcriptome with STAR aligner [88] (RNA-seq parameters: --outFilterMismatchNmax 2 --outFilterMultimapNmax 20 --outFilterType BySJout --alignSJoverhangMin 8 --alignSJDBoverhangMin 2; Ribo-seq used the same parameter except --outFilterMismatchNmax 1). We then used the bam files (combining both 20-minute and 60-minute samples) from both RNA-seq and Ribo-seq for RiboTaper [47] to identify translated ORFs. The Ribo-seq metaplot was created using the create_metaplots.bash function in RiboTaper (v1.3.1a). The distribution of Ribo-seq reads in different genome features was calculated using Ribo-seQC. The Ribo-seq read lengths and offsets for RiboTaper were 24, 25, 26, 27, 28 and 8, 9, 10, 11, 12, respectively. The discovered uORFs and main ORFs were extracted from the RiboTaper output ORF_max_filt file. To calculate translation efficiency, we first used STAR to map the RNA-seq and Ribo-seq reads to the CDS of annotated coding genes. The resulting bam files were used to quantify the transcripts per million (TPM) of each gene with RSEM (v1.3.1) [89]. Then, translation efficiency was calculated, dividing the Ribo-seq TPM by the RNA-seq TPM. The root and shoot data from Hsu *et al.*, 2016 [43] were reanalyzed using the Araport11

annotation. The P_sites_all files from RiboTaper were processed using the following code: cut -f 1,3,6 P_sites_all | sort | uniq -c | sed -r 's/^(*[^]+) +/\1\t/' > output.txt to combine the read counts at each P-site. Data visualization and statistical analysis were performed in R (v4.0.3) [90]. In particular, the gene levels from the Ribo-seq and RNA-seq reads were visualized with RiboPlotR [91].

The mRNA half-life and mRNA-decay rate data were downloaded from Szabo et al. and Sorenson et al., respectively [65,71]. The quantitative proteomics data for Arabidopsis shoots and roots were downloaded from Song et al [66]. The Arabidopsis gene expression levels (TPM) used for investigating mRNA levels for uORF genes in different Arabidopsis organs, growth stages and ecotypes were directly downloaded from the EMBO-EBI expression atlas (E-MTAB-7978: Arabidopsis tissue atlas, E-GEOD-53197: 17 Arabidopsis thaliana accessions, E-MTAB-4812: tomato root, leaf, flower [two stages] and fruit [six stages], and E-MTAB-4813: three longitudinal sections of six stages during tomato fruit development) [92]. The raw RNA-seq and Ribo-seq data in this study were deposited in the Gene Expression Omnibus (GEO) under accession number GSE183264.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by an NIH S10 OD018174 Instrumentation Grant. This work was supported by a

National Science Foundation grant (2051885) and a Michigan State University startup grant to PYH.

REFERENCES

1 Kim B-H, Cai X, Vaughn JN, von Arnim AG. On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome biology* 2007; **8**: R60.

2 Johnstone TG, Bazzini AA, Giraldez AJ. Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO Journal* 2016; **35**: 706–723.

3 Zhang H, Wang Y, Wu X, Tang X, Wu C, Lu J. Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nature Communications* 2021; **12**: 1–17.

4 Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science* 2016; **352**: 1413–1416.

5 Jackson RJ, Hellen CUT, Pestova T V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* 2010; **11**: 113–127.

6 Hinnebusch AG. The scanning mechanism of eukaryotic translation initiation. *Annual Review of Biochemistry* 2014; **83**: 779–812.

7 Dever TE, Ivanov IP, Sachs MS. Conserved Upstream Open Reading Frame Nascent Peptides That Control Translation. *Annual Review of Genetics* 2020; **54**: 237–264.

8 Kozak M. Constraints on reinitiation of translation in mammals. *Nucleic Acids Research* 2001; **29**: 5226–5232.

9 Kozak M. Initiation of translation in prokaryotes and eukaryotes. *Gene* 1999; **234**: 187–208.

10 Chew G-L, Pauli A, Schier AF. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications* 2016; **7**: 11663.

11 Liu M-J, Wu S-H, Wu J-F, Lin W-D, Wu Y-C, Tsai T-Y, Tsai H-L, Wu S-H. Translational landscape of photomorphogenic Arabidopsis. *The Plant Cell* 2013; **25**: 3699–710.

12 Wu H-YL, Song G, Walley JW, Hsu PY. The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiology* 2019; **181**: 367–380.

13 Lin Y, May GE, Kready H, Nazzaro L, Mao M, Spealman P, Creeger Y, Joel McManus C. Impacts of uORF codon identity and position on translation regulation. *Nucleic Acids Research* 2019; **47**: 9358–9367.

14 Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread

reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**: 7507–7512.

15 He F, Jacobson A. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is only Part of the Story. *Annual Review of Genetics* 2015; **49**: 339–366.

16 Hug N, Longman D, Cáceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* 2016; **44**: 1483–1495.

17 Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nature Reviews Molecular Cell Biology 2019 20:7* 2019; **20**: 406–420.

18 Raxwal VK, Riha K. Nonsense mediated RNA decay and evolutionary capacitance. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 2016; **1859**: 1538–1543.

19 Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A, Toyoda T, Shinozaki K, Seki M. Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**: 2453–2458.

20 Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, Marshall J, Fuller J, Cardle L, McNicol J, Dinh HQ, Barta A, Brown JWS. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Research* 2012; **40**: 2454–2469.

21 Drechsel G, Kahles A, Kesarwani AK, Stauffer E, Behr J, Drewe P, Rätsch G, Wachter A. Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *The Plant Cell* 2013; **25**: 3726–3742.

22 Degtiar E, Fridman A, Gottlieb D, Vexler K, Berezin I, Farhi R, Golani L, Shaul O. The feedback control of UPF3 is crucial for RNA surveillance in plants. *Nucleic Acids Research* 2015; **43**: 4219–4235.

23 Filichkin SA, Cumbie JS, Dharmawardhana P, Jaiswal P, Chang JH, Palusa SG, Reddy ASN, Megraw M, Mockler TC. Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Molecular Plant* 2015; **8**: 207–227.

24 Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, Binder BM, Heber S, Stepanova AN, Alonso JM. Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. *Cell* 2015; **163**: 684–697.

25 Chiam N-C, Fujimura T, Sano R, Akiyoshi N, Hiroyama R, Watanabe Y, Motose H, Demura T, Ohtani M. Nonsense-Mediated mRNA Decay Deficiency Affects the Auxin Response and Shoot Regeneration in Arabidopsis. *Plant and Cell Physiology* 2019; **60**: 2000–2014.

26 Jeong HJ, Kim YJ, Kim SH, Kim YH, Lee IJ, Kim YK, Shin JS. Nonsense-mediated mRNA decay factors, UPF1 and UPF3, contribute to plant defense. *Plant and Cell Physiology* 2011; **52**: 2147–2156.

27 Rayson S, Arciga-Reyes L, Wootton L, de Torres Zabala M, Truman W, Graham N, Grant M, Davies B. A role for nonsense-mediated mrna decay in plants: Pathogen responses are induced in Arabidopsis thaliana nmd mutants. *PLoS ONE* 2012; **7**. doi:10.1371/journal.pone.0031917.

28 Riehs-Kearnan N, Gloggnitzer J, Dekrout B, Jonak C, Riha K. Aberrant growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Research* 2012; **40**: 5615–5624.

29 Yoine M, Nishii T, Nakamura K. Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. *Plant and Cell Physiology* 2006; **47**: 572–580.

30 Raxwal VK, Simpson CG, Gloggnitzer J, Entinze JC, Guo W, Zhang R, Brown JWS, Riha K. Nonsense-Mediated RNA Decay Factor UPF1 Is Critical for Posttranscriptional and Translational Gene Regulation in Arabidopsis. *The Plant Cell* 2020; **32**: 2725–2741.

31 Kertész S, Kerényi Z, Mérai Z, Bartos I, Pálfy T, Barta E, Silhavy D. Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Research* 2006; **34**: 6147–6157.

32 Hurt JA, Robertson AD, Burge CB. Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Research* 2013; **23**: 1636–1650.

33 Tanaka M, Sotta N, Yamazumi Y, Yamashita Y, Miwa K, Murota K, Chiba Y, Hirai MY, Akiyama T, Onouchi H, Naito S, Fujiwara T. The minimum open reading frame, AUG-Stop, induces boron-dependent ribosome stalling and mRNA degradation. *The Plant Cell* 2016; **28**: 2830–2849.

34 Dyle MC, Kolakada D, Cortazar MA, Jagannathan S. How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. *Wiley Interdisciplinary Reviews: RNA* 2020; **11**. doi:10.1002/wrna.1560.

35 Decourty L, Doyen A, Malabat C, Frachon E, Rispal D, Séraphin B, Feuerbach F,

Jacquier A, Saveanu C. Long Open Reading Frame Transcripts Escape Nonsense-Mediated mRNA Decay in Yeast. *Cell Reports* 2014; **6**: 593–598.

36 Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. *Journal of Cell Science* 2016; **129**: 461–467.

37 Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D. Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Molecular Biology* 2009; **71**: 367–378.

38 Kurihara Y, Makita Y, Kawashima M, Fujita T, Iwasaki S, Matsui M. Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 2018; **115**: 7831–7836.

39 Li YR, Liu MJ. Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants. *Genome Research* 2020; **30**: 1418–1433.

40 Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009; **324**: 218–23.

41 Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology* 2015; **16**: 651–664.

42 Juntawong P, Girke T, Bazin J, Bailey-Serres J. Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 2014; **111**: E203-12.

43 Hsu PY, Calviello L, Wu H-YL, Li F-W, Rothfels CJ, Ohler U, Benfey PN. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 2016; **113**: E7126–E7135.

44 Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J. Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proceedings of the National Academy of Sciences of the United States of America* 2017; **114**: E10018–E10027.

45 Kurihara Y, Makita Y, Shimohira H, Fujita T, Iwasaki S, Matsui M. Translational landscape of protein-coding and non-protein-coding RNAs upon light exposure in Arabidopsis. *Plant and Cell Physiology* 2020; **61**: 536–545.

46 Hu Q, Merchante C, Stepanova AN, Alonso JM, Heber S. Genome-Wide Search for Translated Upstream Open Reading Frames in Arabidopsis Thaliana. *IEEE Transactions*

on Nanobioscience 2016; **15**: 150–159.

47 Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U. Detecting actively translated open reading frames in ribosome profiling data. *Nature Methods* 2016; **13**: 165–170.

48 Xu Z, Hu L, Shi B, Geng S, Xu L, Wang D, Lu ZJ. Ribosome elongating footprints denoised by wavelet transform comprehensively characterize dynamic cellular translation events. *Nucleic Acids Research* 2018; **46**. doi:10.1093/nar/gky533.

49 Harnett D, Meerdink E, Calviello L, Sydow D, Ohler U. Genome-Wide Analysis of Actively Translated Open Reading Frames Using RiboTaper/ORFquant. *Methods in Molecular Biology* 2021; **2252**: 331–346.

50 Silverstone AL, Ciampaglio CN, Sun TP. The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *The Plant Cell* 1998; **10**: 155–169.

51 Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, Panda S, Kreps JA, Kay SA. Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* 2000; **289**: 768–771.

52 Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. *Cell* 1996; **86**: 423–433.

53 Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW. AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in Arabidopsis thaliana. *Development* 2005; **132**: 4563–4574.

54 Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, Guilfoyle TJ, Hagen G, Alonso JM, Cohen JD, Farmer EE, Ecker JR, Reed JW. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development* 2005; **132**: 4107–4118.

55 Ni M, Tepperman JM, Quail PH. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* 1998; **95**: 657–667.

56 Huq E, Quail PH. PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in Arabidopsis. *EMBO Journal* 2002; **21**: 2441–2450.

57 Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH. The

Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to prolonged red light by modulating phyB levels. *The Plant Cell* 2008; **20**: 337–352.

58 Sharrock RA, Quail PH. Novel phytochrome sequences in Arabidopsis thaliana: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes & Development* 1989; **3**: 1745–1757.

59 Devlin PF, Patel SR, Whitelam GC. Phytochrome E Influences Internode Elongation and Flowering Time in Arabidopsis. *The Plant Cell* 1998; **10**: 1479.

60 Zhang SH, Lawton MA, Hunter T, Lamb CJ. atpk1, a novel ribosomal protein kinase gene from Arabidopsis. I. Isolation, characterization, and expression. *Journal of Biological Chemistry* 1994; **269**: 17586–17592.

61 Mizoguchi T, Hayashida N, Yamaguchi-Shinozaki K, Kamada H, Shinozaki K. Two genes that encode ribosomal-protein S6 kinase homologs are induced by cold or salinity stress in Arabidopsis thaliana. *FEBS Letters* 1995; **358**: 199–204.

62 Park YS, Hong SW, Oh SA, Kwak JM, Lee HH, Nam HG. Two putative protein kinases from Arabidopsis thaliana contain highly acidic domains. *Plant Molecular Biology* 1993; **22**: 615–624.

63 Boudsocq M, Barbier-Brygoo H, Laurière C. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana. *Journal of Biological Chemistry* 2004; **279**: 41758–41766.

64 Von Arnim AG, Jia Q, Vaughn JN. Regulation of plant translation by upstream open reading frames. *Plant Science* 2014; **214**: 1–12.

65 Szabo EX, Reichert P, Lehniger MK, Ohmer M, de Francisco Amorim M, Gowik U, Schmitz-Linneweber C, Laubinger S. Metabolic labeling of RNAs uncovers hidden features and dynamics of the Arabidopsis transcriptome. *The Plant Cell* 2020; **32**: 871–887.

66 Song G, Hsu PY, Walley JW. Assessment and refinement of sample preparation methods for deep and quantitative plant proteome profiling. *Proteomics* 2018; **18**: 1800220.

67 Wolfa YI, Novichkovb PS, Kareva GP, Koonina E V., Lipmana DJ. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**: 7273–7280.

68 Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, Zhang D. Young genes out of the male: An insight from evolutionary age analysis of the pollen transcriptome. *Molecular Plant* 2015;

8: 935–945.

69 Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T. The dwarf phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. *Development* 2006; **133**: 3575–85.

70 Sieburth LE, Vincent JN. Beyond transcription factors: Roles of mrna decay in regulating gene expression in plants. *F1000Research* 2018; **7**: 1940.

71 Sorenson RS, Deshotel MJ, Johnson K, Adler FR, Sieburth LE. Arabidopsis mRNA decay landscape arises from specialized RNA decay substrates, decapping-mediated feedback, and redundancy. *Proceedings of the National Academy of Sciences of the United States of America* 2018; **115**: E1485–E1494.

72 Deyholos MK, Cavaness GF, Hall B, King E, Punwani J, Van Norman J, Sieburth LE. Varicose, a WD-domain protein, is required for leaf blade. *Development* 2003; **130**: 6577–6588.

73 Goeres DC, Van Norman JM, Zhang W, Fauver NA, Spencer M Lou, Sieburth LE. Components of the Arabidopsis mRNA decapping complex are required for early seedling development. *The Plant Cell* 2007; **19**: 1549–1564.

74 Xu J, Yang JY, Niu QW, Chua NH. Arabidopsis DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. *The Plant Cell* 2006; **18**: 3386–3398.

75 Zhang W, Murphy C, Sieburth LE. Conserved RNaseII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**: 15981–15985.

76 Andreev DE, Arnold M, Kiniry SJ, Loughran G, Michel AM, Rachinskii D, Baranov P V. TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response. *eLife* 2018; **7**. doi:10.7554/eLife.32563.

77 Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Coller J. Codon optimality is a major determinant of mRNA stability. *Cell* 2015; **160**: 1111.

78 Radhakrishnan A, Chen YH, Martin S, Alhusaini N, Green R, Coller J. The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell* 2016; **167**: 122-132.e9.

79 Beelman CA, Parker R. Differential effects of translational inhibition in cis and in trans on

the decay of the unstable yeast MFA2 mRNA. *Journal of Biological Chemistry* 1994; **269**: 9687–9692.

80 Schwartz DC, Parker R. mRNA Decapping in Yeast Requires Dissociation of the Cap Binding Protein, Eukaryotic Translation Initiation Factor 4E. *Molecular and Cellular Biology* 2000; **20**: 7933–7942.

81 LaGrandeur T, Parker R. The cis acting sequences responsible for the differential decay of the unstable MFA2 and stable PGK1 transcipts in yeast include the context of the translational start codon. *RNA* 1999; **5**: 420–433.

82 Chan LY, Mugler CF, Heinrich S, Vallotton P, Weis K. Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife* 2018; **7**. doi:10.7554/eLife.32536.

83 Hou C-Y, Lee W-C, Chou H-C, Chen A-P, Chou S-J, Chen H-M. Global analysis of truncated RNA ends reveals new insights into ribosome stalling in plants. *The Plant Cell* 2016; **28**: 2398–2416.

84 Yu X, Willmann MR, Anderson SJ, Gregory BD. Genome-wide mapping of uncapped and cleaved transcripts reveals a role for the nuclear mrna cap-binding complex in cotranslational rna decay in Arabidopsis. *The Plant Cell* 2016; **28**: 2385–2397.

85 Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-Limiting Steps in Yeast Protein Translation. *Cell* 2013; **153**: 1589–1601.

86 Lawson MR, Lessen LN, Wang J, Prabhakar A, Corsepius NC, Green R, Puglisi JD. Mechanisms that ensure speed and fidelity in eukaryotic translation termination. *Science* 2021; **373**: 876–882.

87 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012; **9**: 357–9.

88 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**: 15–21.

89 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011; **12**: 323.

90 R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2017. doi:/S0103-64402004000300015.

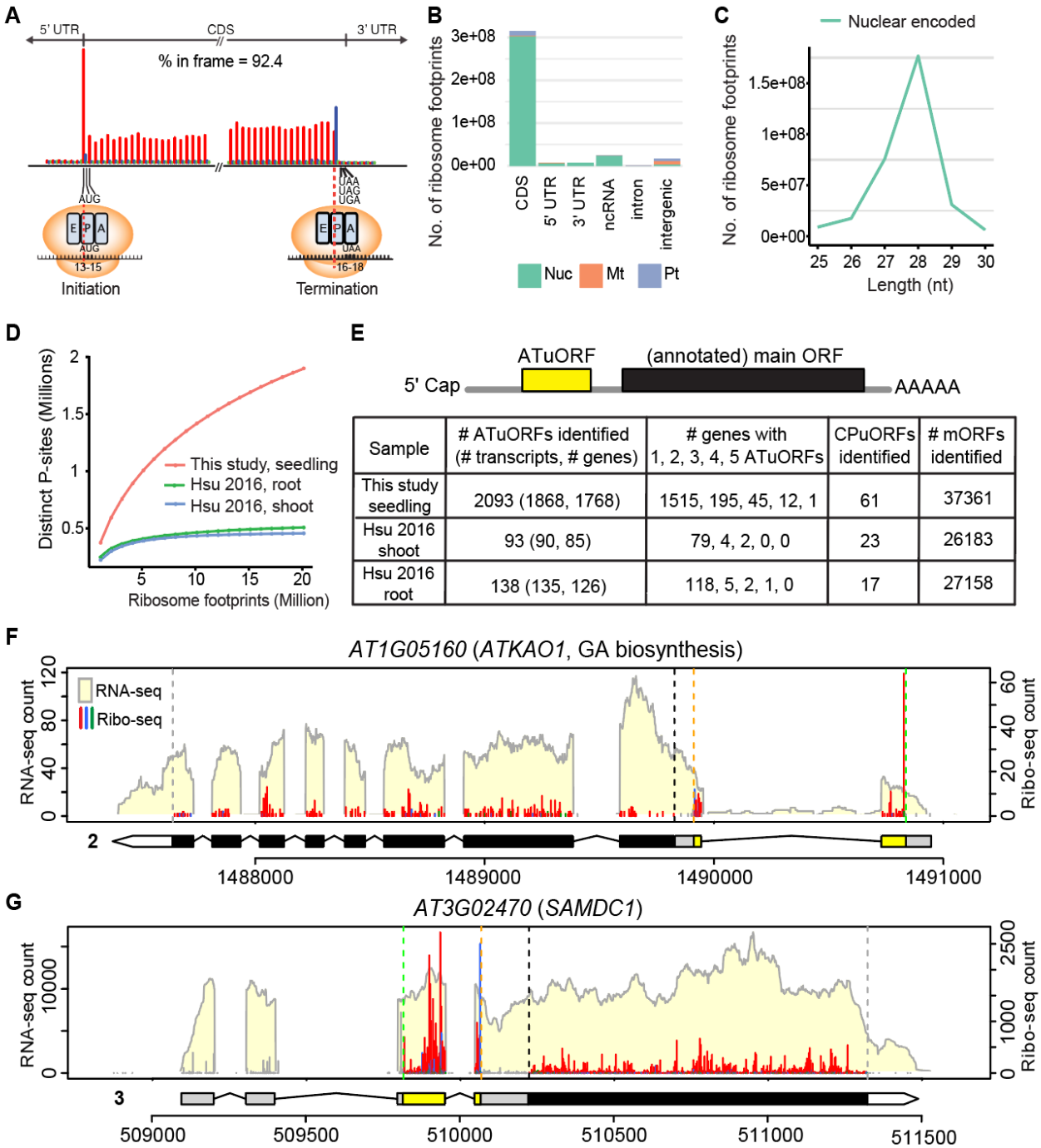
91 Wu H-YL, Hsu PY. Visualizing the periodic Ribo-seq reads with RiboPlotR. *bioRxiv* 2019. doi:10.1101/694646.

92 Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S, Fonseca

NA, Füllgrabe A, Green M, Huang N, Huerta L, Iqbal H, Jianu M, Mohammed S, Zhao L, Jarnuczak AF, Jupp S, Marioni J, Meyer K, Petryszak R, Prada Medina CA, Talavera-López C, Teichmann S, Vizcaino JA, Brazma A. Expression Atlas update: From tissues to single cells. *Nucleic Acids Research* 2020; **48**: D77–D83.

FIGURE LEGENDS

Figure 1. Enhanced ribosome profiling improves uORF identification



Sample	# ATuORFs identified (# transcripts, # genes)	# genes with 1, 2, 3, 4, 5 ATuORFs	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

Fig.1. Enhanced ribosome profiling coverage improves uORF identification

(A) Metagene analysis of 28-nt ribosome footprints mapped to regions near the start and stop codons of annotated ORFs. The reads are presented with their first nt at the P-site,

which is the 13th nt for 28-nt footprints. The reads are colored in red, blue, and green to indicate they are in the first (expected), second, or third reading frames, respectively. The majority of footprints were mapped to the CDS in the expected reading frame (92.4% in frame).

(B) Genomic features mapped based on ribosome footprints. Reads that mapped to nuclear (Nuc), mitochondrial (Mt), and plastid (Pt)-encoded genes are shown.

(C) Length distribution of ribosome footprints. Reads that mapped to nuclear-encoded genes are presented.

(D) Distinct P-sites detected in 1 to 20 million randomly selected ribosome footprints from our current and previous datasets.

(E) Numbers of ATuORFs, CPuORFs and mORFs identified in our current and previous datasets.

(F-G) RNA-seq and Ribo-seq profiles of *ATKAO1 and SAMDC1* in our current data. RNA-seq coverage is shown with a light yellow background. Ribo-seq reads are presented with their first nt at the P-site, and they are colored in red, blue, and green to indicate they are in the first (expected), second, and third reading frames, respectively. Within the gene models, yellow boxes represent the ATuORFs, black boxes represent the annotated mORFs, and gray and white regions indicate 5' UTRs and 3' UTRs, respectively. On the left of the gene model, the specific isoform being considered is indicated. Within the profiles, green and orange vertical dashed lines represent translation start and stop, respectively, for the ATuORF. Black and gray vertical dashed lines represent translation start and stop, respectively, for the annotated mORF.

Figure 2. Expression levels of ATuORF genes, PTuORF genes, and no-uORF genes

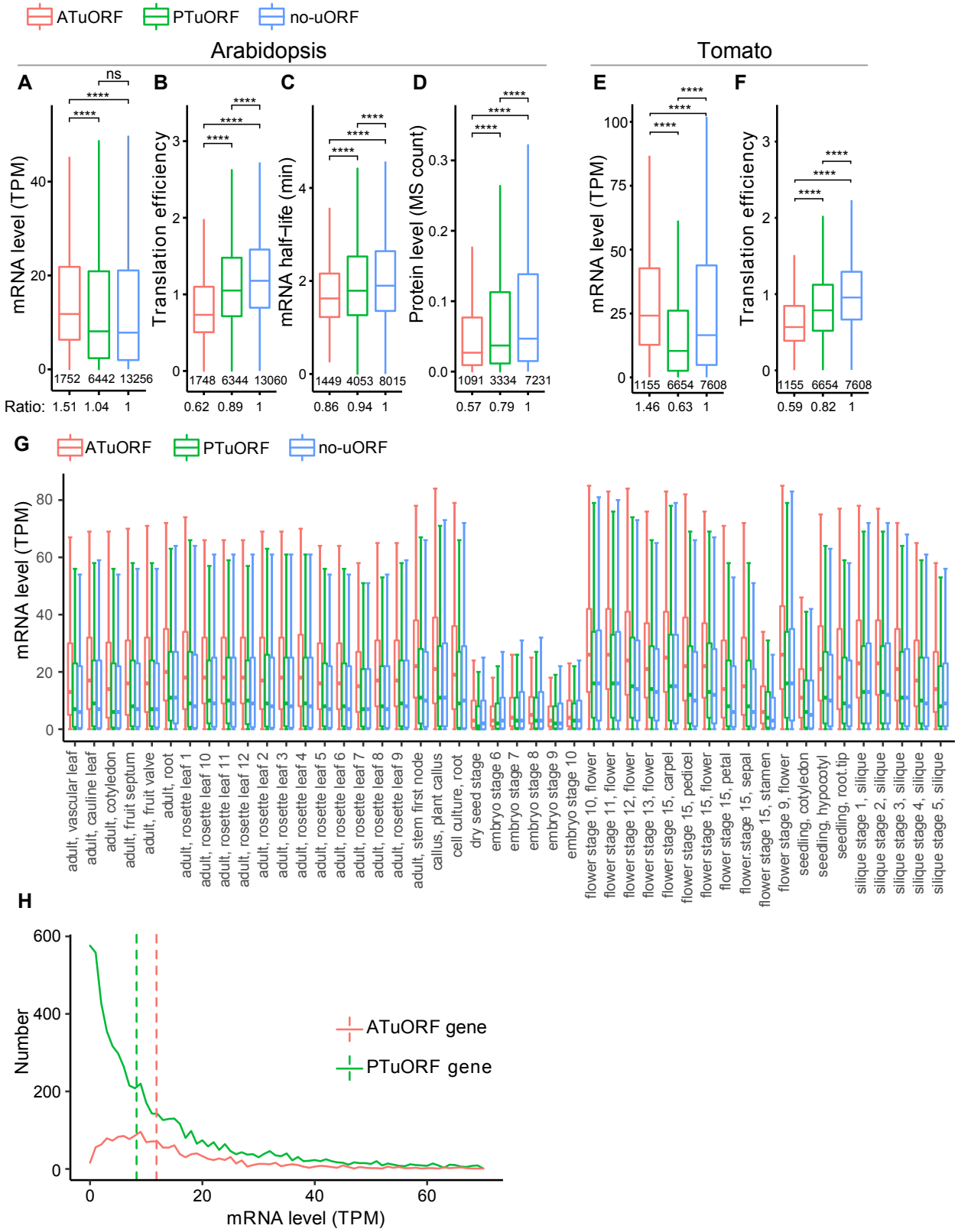


Fig. 2. Expression levels of ATuORF, PTuORF, and no-uORF genes.

(A-D) The mRNA levels, translation efficiency, mRNA half-lives, and protein levels of ATuORF, PTuORF, and no-uORF genes in Arabidopsis seedlings. The mRNA half-life data, which were measured with metabolic labeling in Arabidopsis seedlings, were extracted from [65], and the protein abundance as determined by quantitative proteomics in Arabidopsis seedling shoots were extracted from [66]. Within the boxplots, the number of genes in each category is listed below the lower whisker. Underneath the plots, the ratios indicate the median of each group normalized to that of the no-uORF genes. The statistical significance was determined via Wilcoxon signed-rank test. The adjusted p-value was determined using the Benjamini & Yekutieli procedure to control for the false discovery rate in multiple testing (*: 0.05 > p > 0.01, ** 0.01 > p > 0.001, *** 0.001 > p > 1e-4, **** 1e-4 > p > 0). These statistical analyses and graphic layouts are used for all boxplots throughout the study.

(E-F) The mRNA levels and translation efficiency of ATuORF, PTuORF, and no-uORF genes in our previous tomato root data extracted from [12].

(G) The mRNA levels of ATuORF, PTuORF, and no-uORF genes in 46 Arabidopsis tissues and developmental stages from independent studies. The TPM values for the RNA-seq quantification were downloaded from the EMBO-EBI expression atlas [92] (see Materials and Methods).

(H) The distributions of the mRNA levels of ATuORF and PTuORF genes in Arabidopsis seedlings.

Figure 3. Sizes of ATuORF, PTuORF, and no-uORF genes

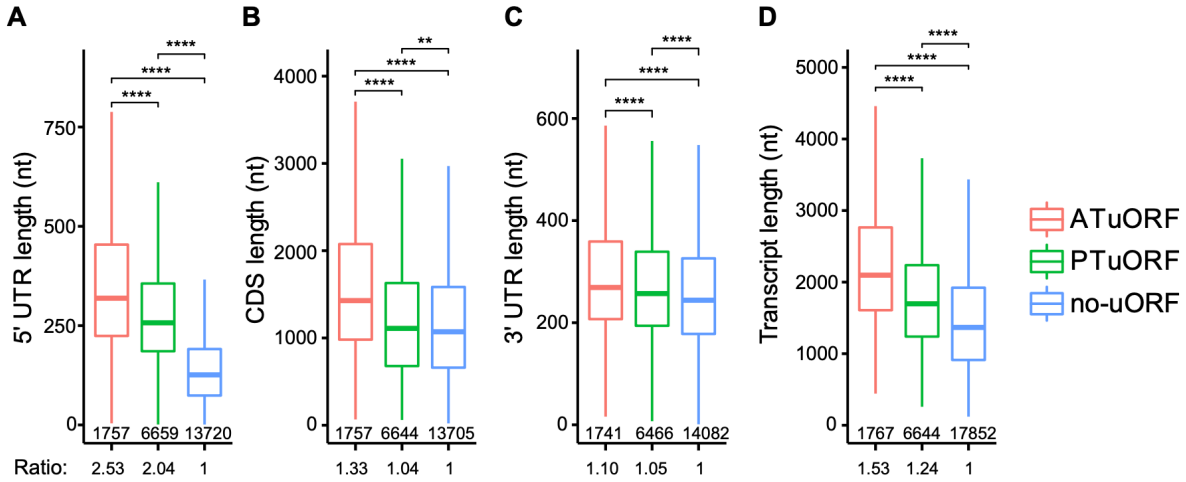


Fig. 3. Sizes of ATuORF, PTuORF, and no-uORF genes.

Lengths of the 5' UTRs (A), CDSs (B), 3' UTRs (C), and entire transcripts (D) of ATuORF, PTuORF, and no-uORF genes.

Figure 4: Expression levels of NMD targets and ATuORF genes

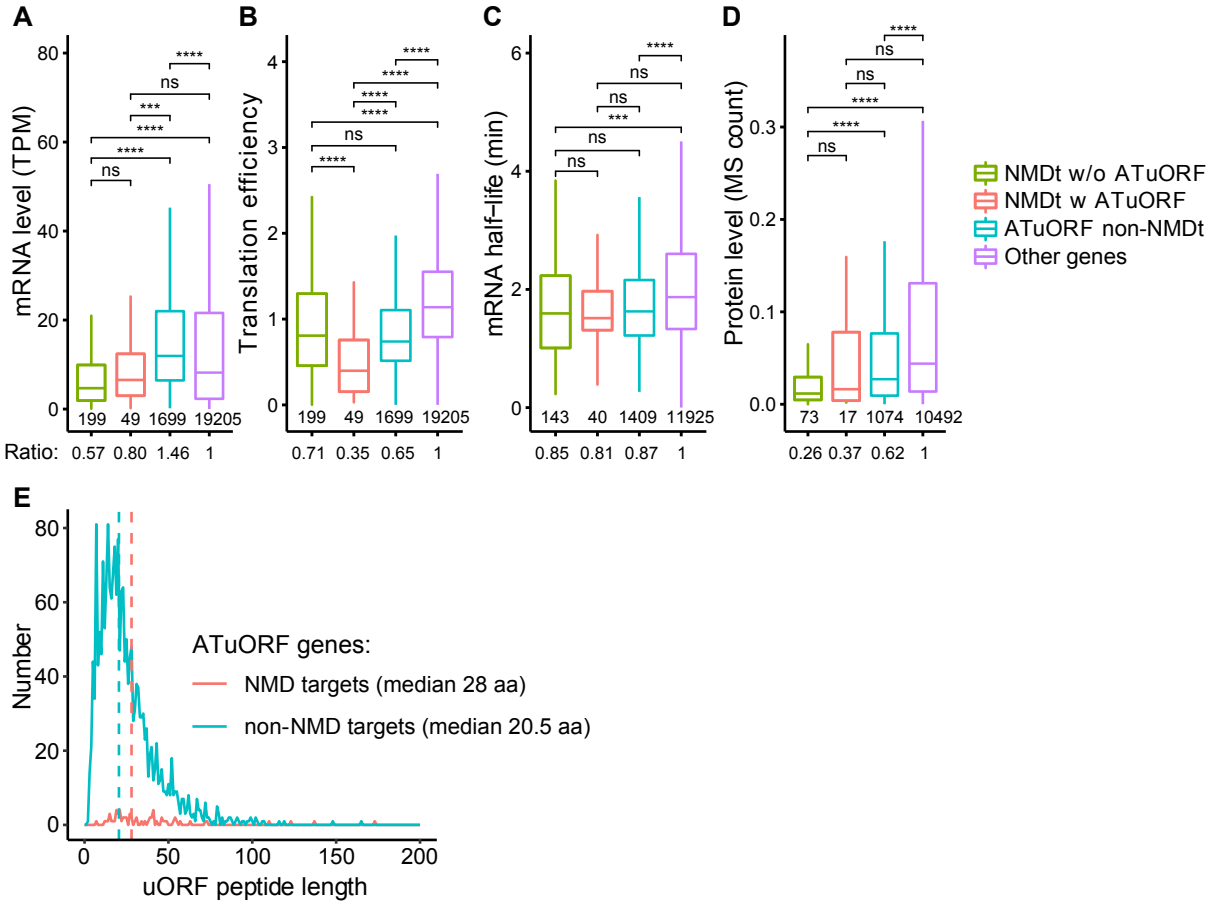


Fig. 4. Expression levels of NMD targets and ATuORF genes.

(A-D) The mRNA levels, translation efficiency, mRNA half-lives, and protein levels of NMD targets (NMDt) without or with ATuORFs compared to those of ATuORF genes and other genes.

(E) The distributions of uORF peptide lengths in NMD target and non-NMD target ATuORF genes.

Figure 5. mRNA decay rates of NMD targets and ATuORF genes

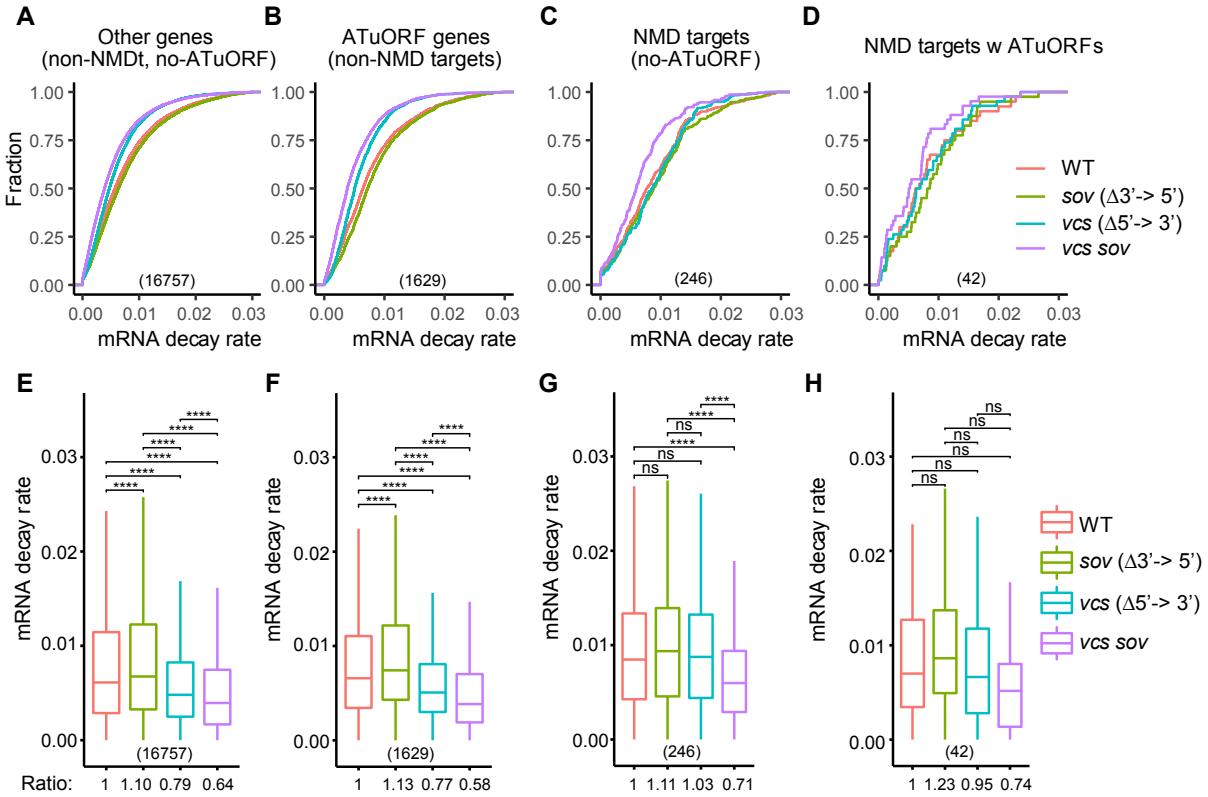


Fig. 5. The mRNA decay rates of NMD targets and ATuORF genes.

Cumulative plots (A-D) and boxplots (E-H) showing the mRNA decay rates for NMD targets and ATuORF genes in wild type and mutants defective in either 5'-to-3' decay (*vcs-7*), 3'-to-5' decay (*sov*), or both (*vcs sov*). The mRNA decay rates were extracted from [71]. The number of genes in each category is listed above the x-axis. Because Col-0 is naturally a *sov* mutant, the wild type here is Col-0 carrying the functional *SOV* allele from Landsberg *erecta* driven by its native promoter [73,75]. The *sov* here is Col-0, and *vcs* is Col-0 with *vsc-7* and the L*er SOV* transgene. The *vcs sov* double mutant is Col-0 carrying the *vcs-7* mutation [71].

Supplementary Figures

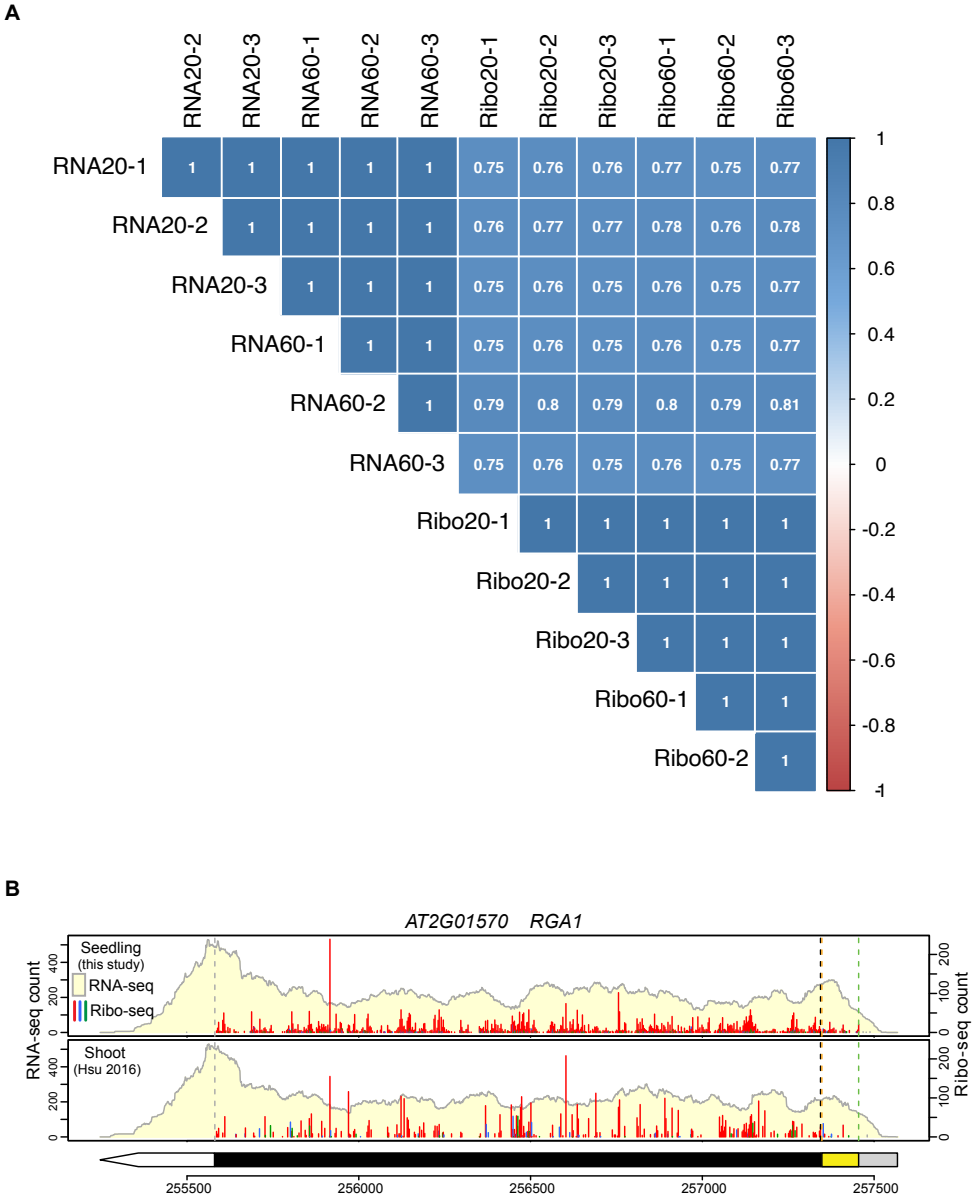
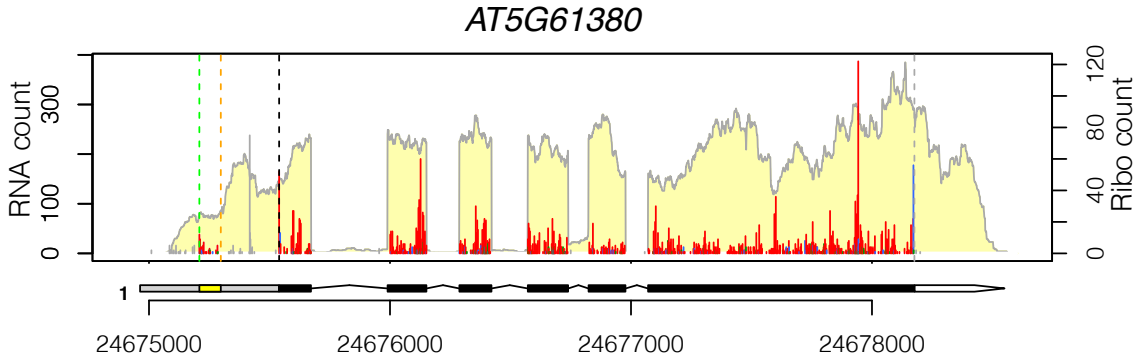


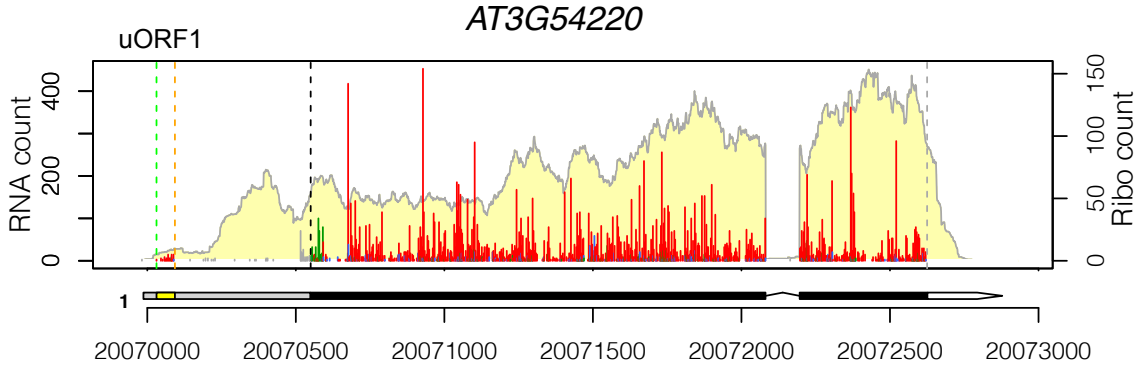
Fig. S1. Correlations among samples and comparisons between our previous and current datasets. (A) Pearson correlation among three biological replicates of our current RNA-seq and Ribo-seq samples. (B) Higher Ribo-seq read coverage in both the uORF and mORF regions in our current dataset using *RGA1* as an example. Note the mRNA levels are similar between our previous and current datasets. Data representations are the same as those described in the Fig. 1E legend.

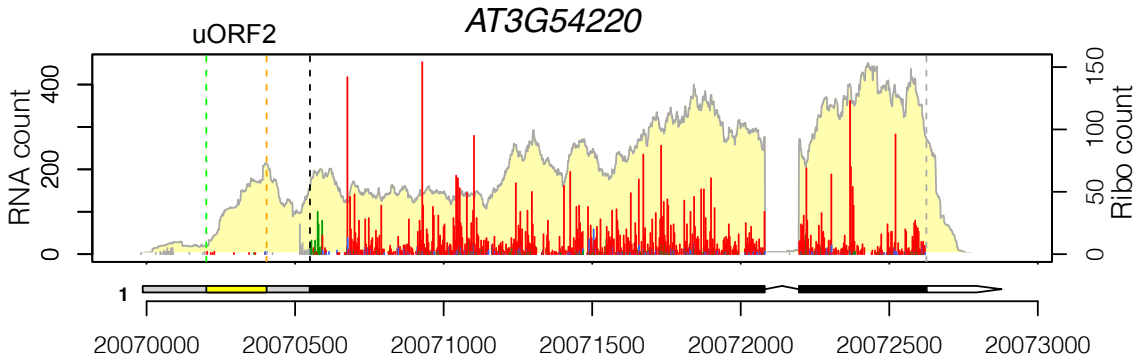
Figure S2. ATuORFs in known important regulatory genes

A. *TOC1*

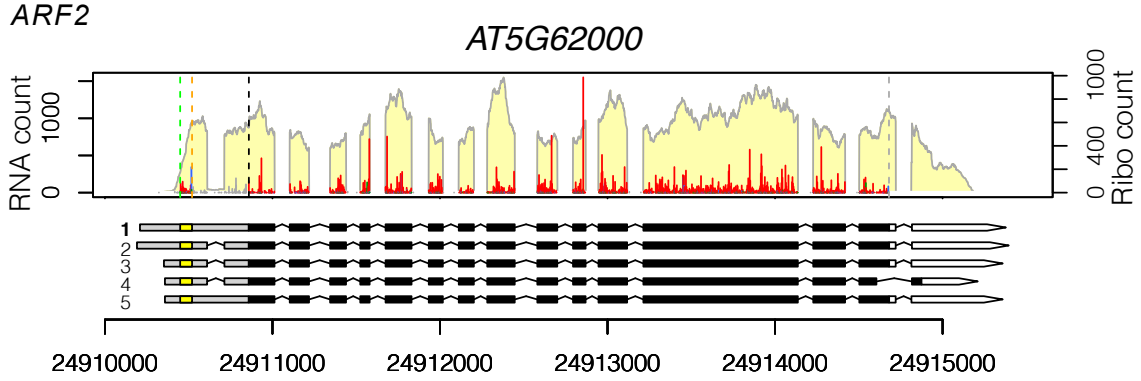


B. *SCARECROW*

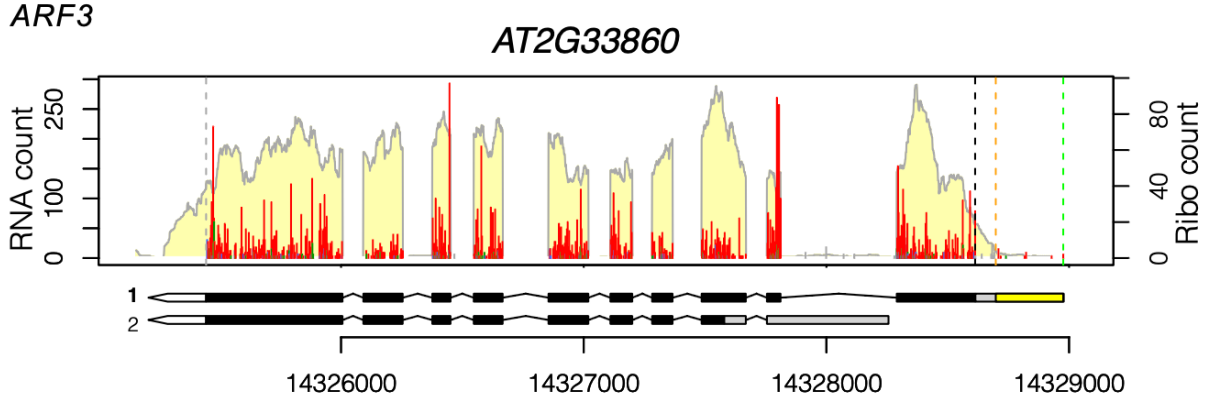




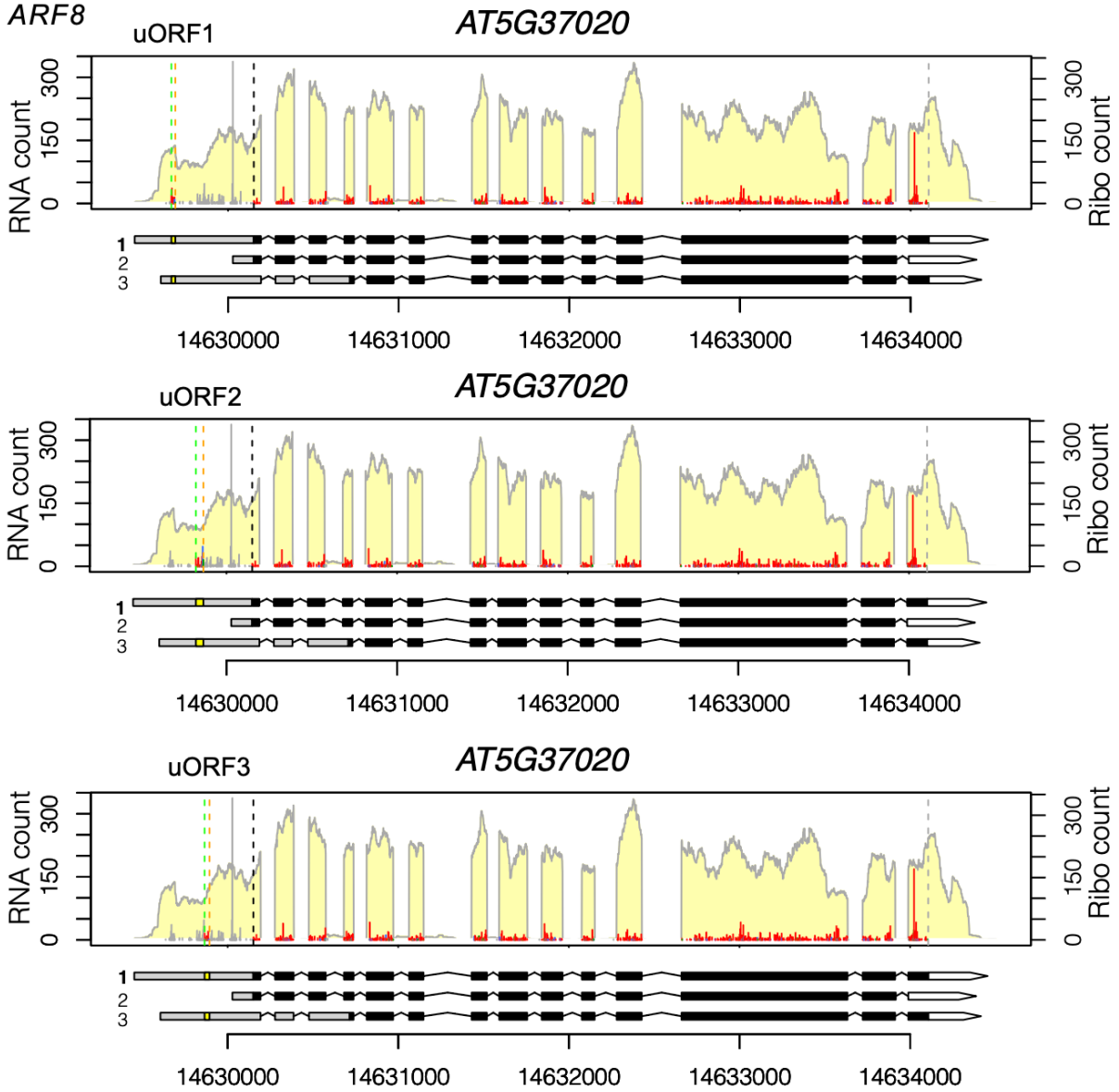
C. *ARF2*



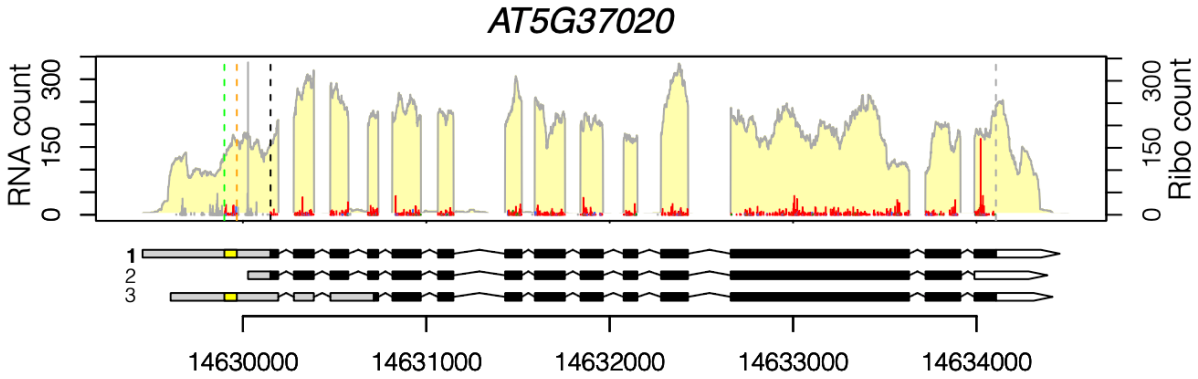
D. *ARF3*


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

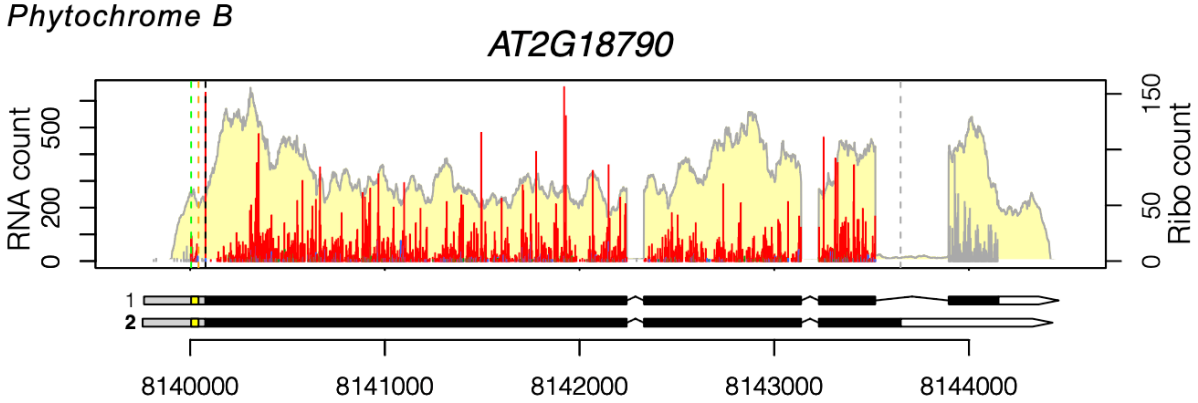
E. *ARF8*


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

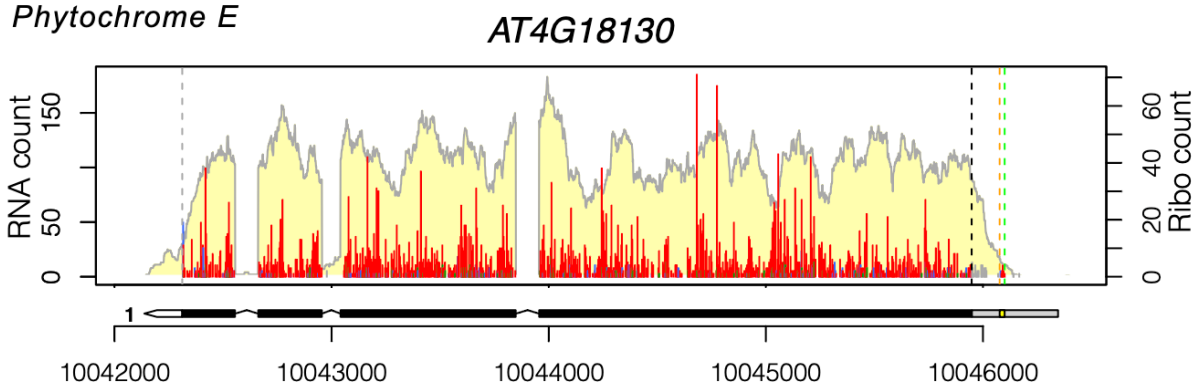
E. *ARF8*



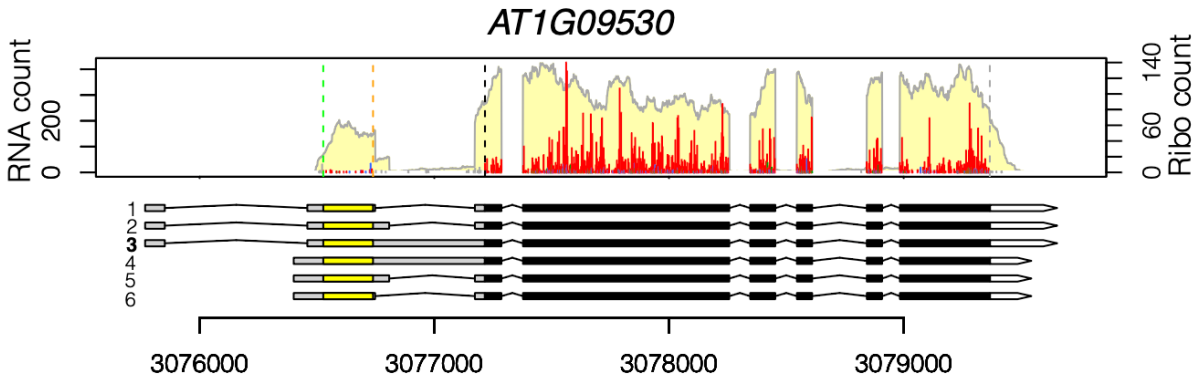
F. *Phytochrome B*



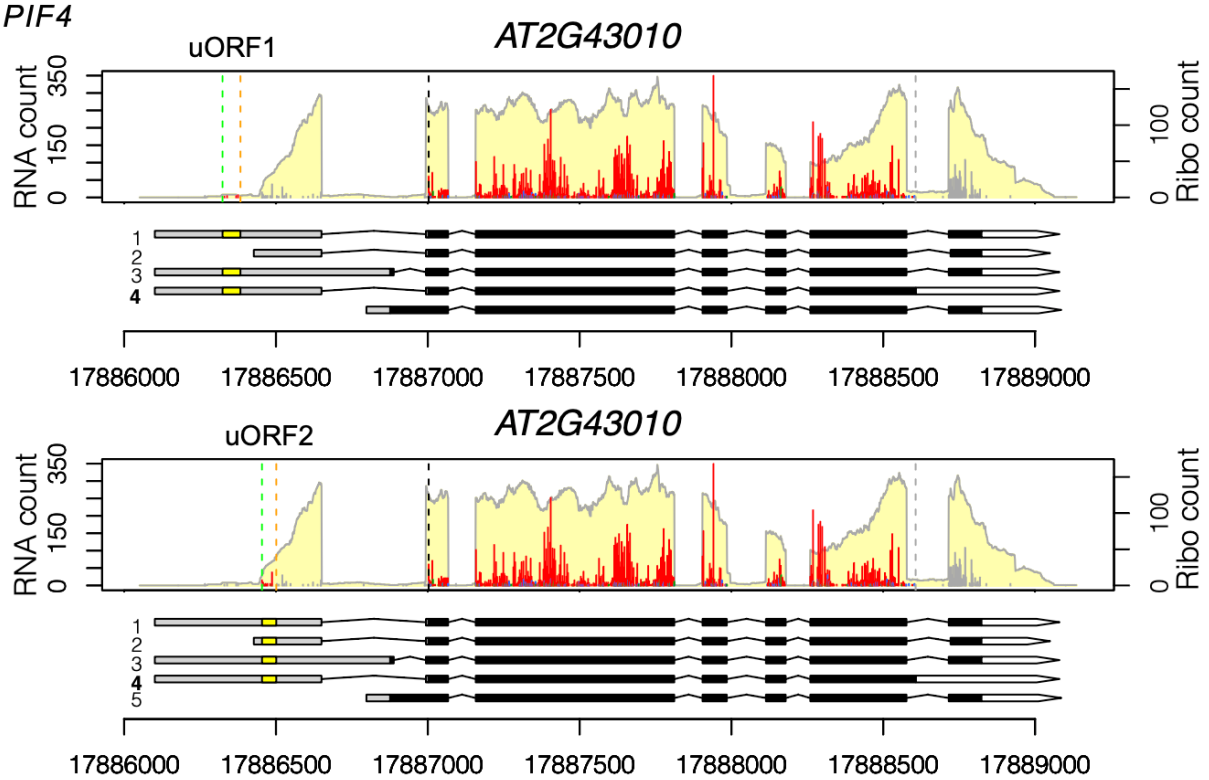
G. *Phytochrome E*



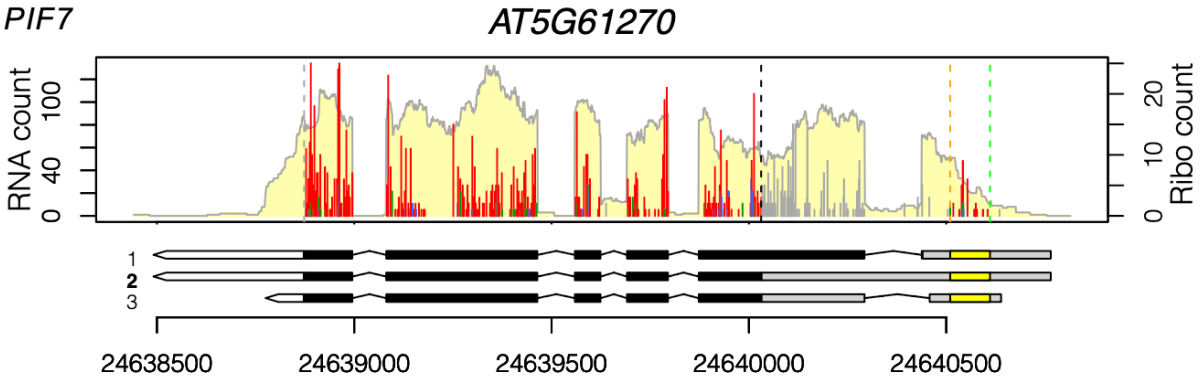
H. *PIF3*



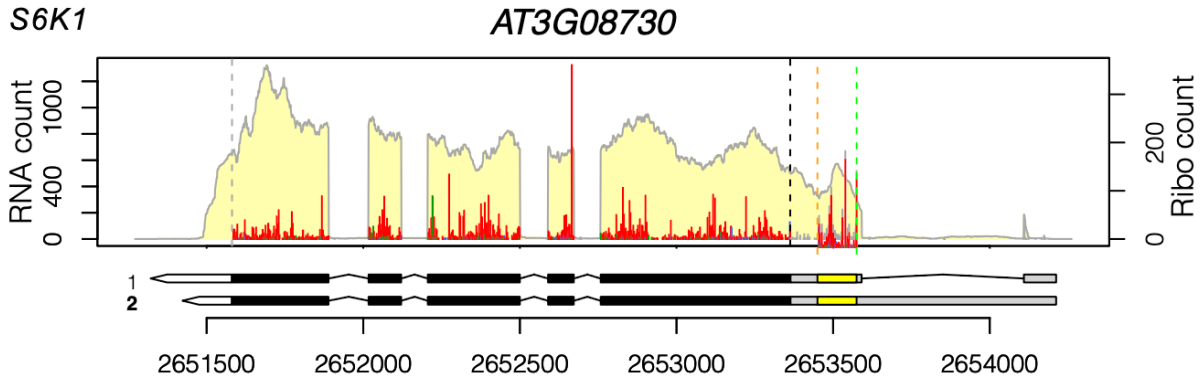
I. *PIF4*



J. *PIF7*



K. *S6K1*



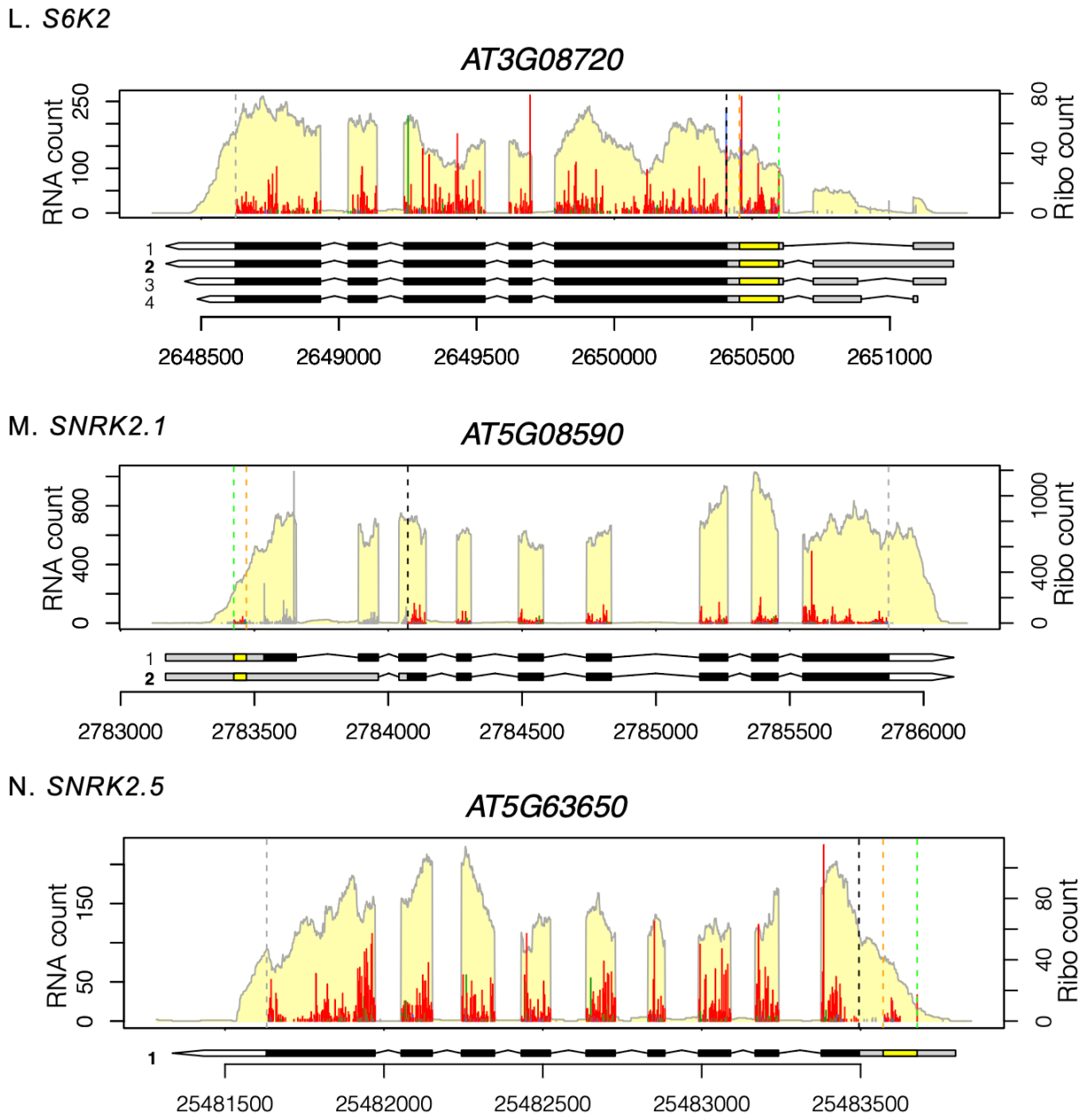


Fig. S2. Examples of ATuORFs in important regulatory genes. Data are represented as described in the Fig. 1E legend. Next to the gene models, the isoforms are numbered, and the isoform being considered is bolded. For genes that have multiple ATuORFs identified (B, E, and I), the 3-nt periodicity of each ATuORF is presented in separate panels.

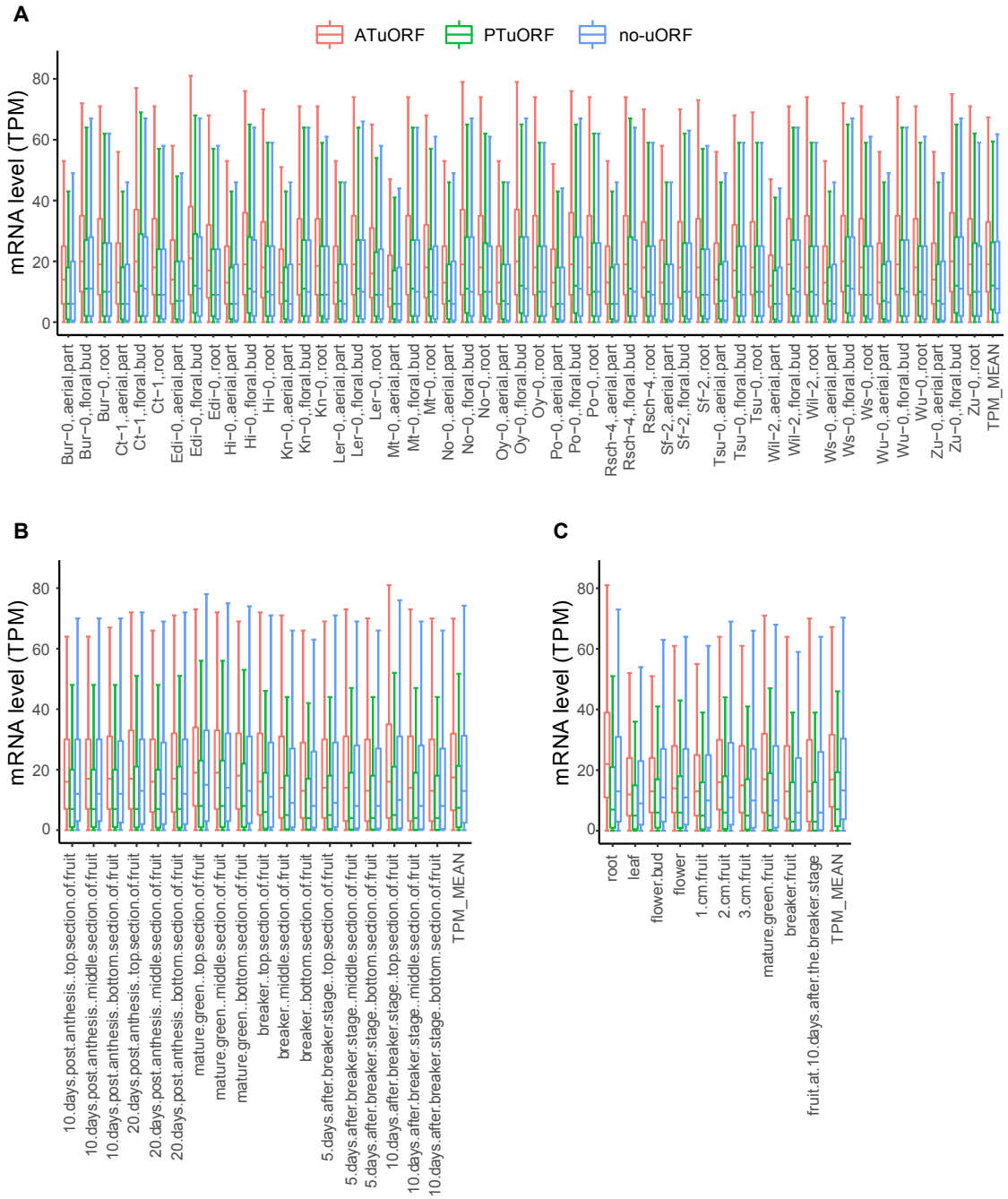


Fig. S3. mRNA levels of uORF-containing genes in different Arabidopsis ecotypes and different tomato growth stages or tissues

(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 of RNA-seq quantification were extracted from the EMBO-EBI expression atlas [92] (see Materials and Methods).

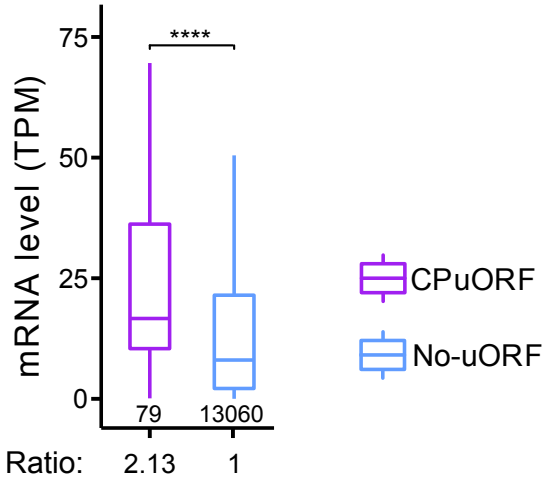


Fig. S4. mRNA levels of CPuORF-containing genes compared to no-uORF genes.

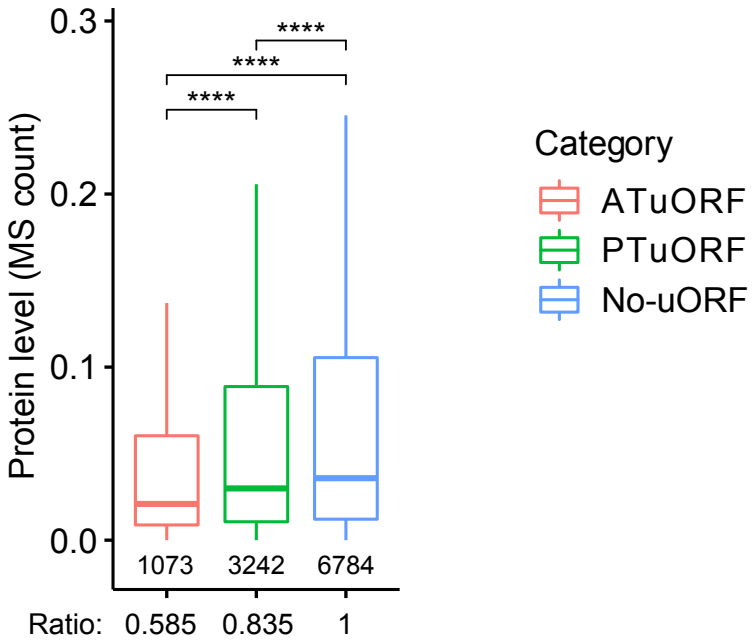


Fig. S5. Protein abundance of the uORF-containing genes in Arabidopsis root. The protein quantification data from Arabidopsis seedling roots were extracted from [66].