

LARGE-SCALE BIOLOGY

Improved Super-Resolution Ribosome Profiling Reveals Prevalent Translation of Upstream ORFs and Small ORFs in Arabidopsis

Hsin-Yen Larry Wu¹, Qiaoyun Ai¹, Rita Teresa Teixeira^{1,2}, Phong H. T. Nguyen¹, Gaoyuan Song³, Christian Montes³, J. Mitch Elmore³, Justin W. Walley³, Polly Yingshan Hsu^{1,#}

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

² Current address: Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisboa, 1749-016 Lisboa, Portugal

³ Department of Plant Pathology, Entomology, & Microbiology, Iowa State University, Ames, IA 50011 USA

Corresponding author: pollyhsu@msu.edu

Running title: The Arabidopsis ORFeome

The author 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).

ABSTRACT

A crucial step in functional genomics is identifying actively translated open reading frames (ORFs) and linking them to biological functions. The challenge lies in identifying short ORFs, as their identification is greatly influenced by data quality and depth. Here, we improved the coverage of super-resolution Ribo-seq in Arabidopsis (*Arabidopsis thaliana*), revealing uncharacterized translation events for nuclear, chloroplastic, and mitochondrial genes. Assisted by a transcriptome assembly, we identified 7,751 unconventional translation events, comprising 6,996 upstream ORFs (uORFs) and 209 downstream ORFs on annotated protein-coding genes, as well as 546 ORFs in presumed non-coding RNAs. Proteomics data confirmed the production of stable proteins from some of these unannotated translation events. We present evidence of active translation from primary transcripts of tasiRNAs (*TAS1–4*) and microRNAs (*pri-MIR163*, *pri-MIR169*), and periodic ribosome stalling supporting co-translational decay. Additionally, we developed a method for identifying extremely short uORFs, including 370 minimum uORFs (AUG-stop), and 2,921 tiny uORFs (2–10 amino acids), and 681 uORFs that overlap with each other. Remarkably, these short uORFs exhibit strong translational repression as do longer uORFs. We also systematically discovered 594 uORFs regulated by alternative splicing, suggesting widespread isoform-specific translational control. Finally, these prevalent uORFs are associated with numerous important pathways. In summary, our improved Arabidopsis translational landscape provides valuable resources to study gene expression regulation.

INTRODUCTION

Accurately defining gene models and determining translated open reading frames (ORFs) is fundamental for studying gene functions and monitoring cellular activity in all living organisms. Despite extensive efforts, our understanding of the translational landscape remains incomplete (Andrews and Rothnagel, 2014; Hellens et al., 2016; Orr et al., 2020; Wu et al., 2023). It has become increasingly clear that a substantial fraction of the transcriptomes in diverse plant species has prevalent unannotated translated ORFs. These unannotated translated ORFs include small ORFs (sORFs) encoded by presumed non-coding RNAs (ncRNAs), as well as short upstream ORFs (uORFs) present in the 5' untranslated regions (5' UTRs) of protein-coding RNAs (i.e., mRNAs) (Liu et al., 2013; Juntawong et al., 2014; Lei et al., 2015; Hsu et al., 2016; Bazin et al., 2017; Wu et al., 2019; Li and Liu, 2020; Kurihara et al., 2020; Sotta et al., 2022; Guo et al., 2023b; Qanmber et al., 2023; Zhu et al., 2023). Lately, downstream ORFs (dORFs) in the 3' UTRs of mRNAs were also reported in vertebrates (Wu et al., 2020). Serendipitous discoveries from early forward genetic screens and recent experimental evidence have shown that these relatively short ORFs can produce small proteins or peptides that play important roles in various aspects of signaling and physiology (Tavormina et al., 2015; Hsu and Benfey, 2018; Takahashi et al., 2019). In addition, the act of translation itself may have regulatory roles, even if the protein products are not functional (Orr et al., 2020). For example, uORFs have long been recognized as cis-regulatory elements suppressing protein synthesis of the downstream main ORFs (mORFs) and are associated with various plant phenotypes (Von Arnim et al., 2014; Xu et al., 2017; Zhang et al., 2018; Xing et al., 2020; Gage et al., 2022). More recently, the association of ribosomes with sORFs on the primary transcripts of trans-acting small interfering RNAs (tasiRNAs) has been suggested to bring the transcripts to the rough endoplasmic reticulum (ER) for microRNA (miRNA)-mediated tasiRNAs biogenesis (Li et al., 2016; Yoshikawa et al., 2016; Hou et al., 2016; Bazin et al., 2017; Iwakawa et al., 2021).

Despite their importance, short ORFs are commonly excluded by computational genome annotations. As a vast number of potential ORF sequences can occur randomly, computational annotations rely on several assumptions about ORFs to minimize false positive identifications. These assumptions typically include the translation of only one ORF, the longest one, from each transcript, and the requirement for the ORF to be greater than 100 codons (Basrai et al., 1997; Olsen et al., 2002; Lease and Walker, 2006). Although this approach is robust and efficient, genuine mRNAs encoding sORFs can be misannotated as ncRNAs. Moreover, the polycistronic potential of mRNAs is not considered, resulting in the exclusion of uORFs or dORFs in the

annotation. Thus, comprehensive identification of these hidden short ORFs is the first step toward understanding and characterizing their functions.

Bioinformatic approaches based on evolutionary conservation or sequence homology to known sORFs have been attempted (Olsen et al., 2002; Lease and Walker, 2006; Hanada et al., 2007, 2009; Zhou et al., 2013; de Bang et al., 2017; Feng et al., 2023; Li et al., 2023). However, accumulating evidence suggests that many sORFs have evolved recently during evolution (Ruiz-Orera et al., 2014; Sandmann et al., 2023), and some sORFs are only conserved in specific families or small groups of plants. For instance, the first plant peptide hormone involved in the wounding response in tomato (*Solanum lycopersicum*), systemin, is only present in the Solaneae, a subtribe of the Solanaceae (Pearce et al., 1991; Constabel et al., 1998). Additionally, a small protein called Qua-Quine Starch (QQS, 59 amino acids [aa]), involved in carbon and nitrogen allocation as well as pathogen susceptibility across species, is encoded by an orphan gene that exists only in Arabidopsis (*Arabidopsis thaliana*) (Li et al., 2009, 2015; Qi et al., 2019). In our previous analysis of the tomato translome, we identified many translated sORFs specific to the Solanaceae or exclusively present in tomato (Wu et al., 2019).

Similarly, although 30–70% of plant genes contain potential uORFs, only 119 uORFs encoding conserved peptides (CPuORFs) corresponding to 81 homology groups, have been identified in Arabidopsis despite extensive searches (Hayden and Jorgensen, 2007; Jorgensen and Dorantes-Acosta, 2012; Vaughn et al., 2012; Takahashi et al., 2012; Van Der Horst et al., 2019, 2020; Takahashi et al., 2020; Zhang et al., 2021). Some CPuORFs are also specific to certain plant families. For example, *MYB51* transcripts contains a Brassicaceae-specific CPuORF, likely because *MYB51* regulates the biosynthesis of glucosinolates, secondary metabolites mainly found in the order Brassicales (Hou et al., 2016). Therefore, these family- or species-specific sORFs and uORFs may have evolved to provide functions specific to certain groups of plants. Together, the small size, low conservation, and a limited number of experimentally characterized sORFs and uORFs available as training datasets restrict the power of bioinformatic prediction. For these reasons, a systematic experimental approach is necessary to uncover these hidden functional short ORFs.

Ribosome profiling (Ribo-seq) has emerged as a high-throughput approach for ORF discovery and quantification. Essentially, Ribo-seq combines ribosome footprinting with deep sequencing (Ingolia et al., 2009; Brar and Weissman, 2015). The procedure involves treating ribosome-bound mRNAs with ribonuclease to obtain ribosome-protected mRNA fragments (RPFs), or ribosome footprints. Sequencing RPFs enables the identification and quantification of ribosome occupancy on mRNAs throughout the transcriptome. Importantly, when using one

nucleotide (nt) to assign the position of RPFs on mRNAs, precisely digested RPFs exhibit enrichment in the expected reading frame along the coding sequences, which is called 3-nt periodicity. This periodic property reflects that ribosomes decode 3 nt at a time and is a benchmark for high-quality Ribo-seq data (Ingolia et al., 2009; Jiang et al., 2022). This 3-nt periodicity has been considered a reliable feature to distinguish real RPFs from contaminant RNA fragments protected by non-ribosomal protein complexes and to separate actively translating ribosomes from ribosomes stalled at certain regions of transcripts without engaging in translation (Guttman et al., 2013; Guydosh and Green, 2014; Jiang et al., 2022). For these reasons, the majority of ORF discovery software utilizes 3-nt periodicity as a key parameter to identify translated ORFs (Wang et al., 2019). Notably, imprecise digestion of RPFs can lead to out-of-frame mapping, decreasing the resolution of Ribo-seq data and the confidence of active translation. Therefore, strong 3-nt periodicity is critical for the success of ORF identification.

We previously optimized the footprinting buffer to improve the precise digestion of RPFs, resulting in over 90% of RPFs being enriched in the expected reading frame for samples derived from Arabidopsis seedling roots and shoots (Hsu et al., 2016). Although the high-quality data revealed numerous unannotated translation events, the number of relatively short ORFs, such as sORFs and uORFs, identified based on significant 3-nt periodicity (Calviello et al., 2016), was low (32 sORFs and 187 uORFs) (Hsu et al., 2016). This result is presumably due to insufficient coverage within these short ORFs. Additionally, multiple high-quality Ribo-seq datasets, including from zebrafish (*Danio rerio*), *Chlamydomonas reinhardtii*, Arabidopsis, and tomato (Bazzini et al., 2014; Chung et al., 2015; Hsu et al., 2016; Wu et al., 2019), have revealed specific out-of-frame mapping of RPFs at translation termination, likely resulted from structural rearrangement upon binding of eukaryotic Release Factors (eRFs) (Alkalaeva et al., 2006; Brown et al., 2015; Matheisl et al., 2015). This out-of-frame mapping at termination can significantly decrease the overall 3-nt periodicity of short ORFs. Furthermore, identifying uORFs is considered more challenging as they can be extremely short and often overlap with other uORFs, leading to low 3-nt periodicity. Therefore, it was proposed that more specialized software is needed for uORF identification (Wang et al., 2020).

Here, we present an improved Ribo-seq dataset with enhanced RPF coverage in Arabidopsis seedlings. Our new data substantially improves the identification of translated uORFs and sORFs, and provides evidence for supporting noncanonical translation associated with various regulation events of gene expression. Additionally, we developed a new computational method to address the issues of out-of-frame mapping at translation termination and overlapping between uORFs for uncovering relatively short uORFs. Combining our new

data and a new computational approach identified over 7000 unannotated translation events. Thus, our enhanced Arabidopsis translational landscape facilitates the discovery of translated ORFs and offers valuable resources for studying gene functions.

RESULTS

Enhanced super-resolution Ribo-seq improved the read coverage

Although our previous super-resolution Ribo-seq data revealed many unannotated translation events, we only detected 187 uORFs and 32 sORFs with significant 3-nt periodicity (Hsu et al., 2016), likely due to insufficient RPF coverage within these short ORFs. We propose here that increasing RPF coverage will improve the efficiency of identifying translated ORFs, especially for relatively short ORFs. We thus tested a new protocol using 7-d-old Arabidopsis seedlings. We generated both Ribo-seq and RNA-seq libraries. We made two major modifications in the Ribo-seq library preparation: 1) changing the order between RPF size selection and ribosomal RNA (rRNA) depletion to maximize the input materials; and 2) reducing the steps of RNA purification to minimize potential loss of RPFs (see MATERIALS AND METHODS for details). We observed that the number of PCR cycles needed for the final Ribo-seq library amplification dropped from 12 cycles (Hsu et al., 2016) to 9 cycles, suggesting that RPF yield increased.

After sequencing and data analysis, we observed excellent correlations among the respective biological replicates for Ribo-seq and RNA-seq samples (**Supplemental Fig. S1**, $r = 0.99$ – 1 for Ribo-seq; $r = 0.98$ – 1 for RNA-seq). Overall, the Ribo-seq samples showed good correlations with the RNA-seq samples (**Supplemental Fig. S1**, $r = 0.87$ – 0.93), as previously reported (Hsu et al., 2016).

Importantly, our new Ribo-seq data displayed characteristics expected for high-quality data, with strong 3-nt periodicity (92% in-frame reads for 28-nt RPFs), high enrichment for coding sequences (CDSs), and expected RPF lengths with 28 nt being the most abundant (**Fig. 1A–C**). To examine RPF diversity and coverage globally, we examined how many P-sites (the peptidyl tRNA-binding sites within translating ribosomes) these RPFs mapped to. From 20 million randomly selected reads, the number of unique P-sites detected in our new data was 4.01-fold higher than that in our previous dataset from Arabidopsis seedling shoots and roots grown under similar conditions (**Fig. 1D**), indicating an increased RPF diversity in the new data. The improved coverage was also evident in individual transcript profiles when comparing genes of similar mRNA levels. The overall RPF coverage within mORFs (and expected uORFs, if any)

was improved, and RPF levels were less noisy in our current data compared to the previous data (**Figs. 1E**, and **8A–B**).

ORF identification based on significant 3-nt periodicity

As the six samples were highly correlated (**Supplemental Fig. S1**), we pooled them to create one large Ribo-seq dataset and one large RNA-seq dataset for ORF identification, with the goal of identifying short ORFs and ORFs potentially translated at lower levels. In total, we obtained 298 million mapped Ribo-seq reads, corresponding to 11.2 million unique P-sites.

To capture translation events arising from unannotated transcripts, we also performed a reference-guided de novo transcriptome assembly using Araport11 annotation (**Supplemental Fig. S2**, workflow for ORF identification; **Supplemental File S1**, newly assembled transcriptome GTF). We then used RiboTaper (Calviello et al., 2016), a spectrum analysis software, to assess whether a significant 3-nt periodicity can be detected within each potential AUG-initiated ORF along both the transcripts annotated in Araport11 and the newly assembled transcripts. Theoretically, the smallest ORFs that RiboTaper could identify encode at least two amino acids.

Our new data dramatically increased the number of translated ORFs detected by RiboTaper (using the default cutoff, Multitaper F-test, $P < 0.05$). Among annotated protein-coding genes, we identified the translation of 2,113 uORFs, 35,191 annotated ORFs (referred to as ‘Conventional CDSs [CCDSs]’ by RiboTaper), and 209 dORFs (**Fig. 2A**, **2B** and **Supplemental Dataset S1A–C**). Additionally, we identified 546 ORFs (referred to as ‘ncORFs’ by RiboTaper) from transcripts that were not previously considered as protein coding (**Fig. 2A**, **2B** and **Supplemental Dataset S1D**). These ncORFs comprised 164 from newly assembled transcripts, 246 from annotated lncRNAs, miRNA precursors, other RNAs, and 136 from annotated pseudogenes (**Fig. 2C**). The number of ORFs identified in this study is in strong contrast to our previous datasets (187 uORFs, 10 dORFs, 64 ncORFs). Overall, these unannotated ORFs tend to be small in size. The median lengths for encoded peptides from ncORFs, dORFs, and uORFs were 46, 29, and 21 aa, respectively (**Fig. 2E**), which is consistent with the notion that ORFs smaller than 100 aa are often excluded from the annotation (Basrai et al., 1997).

The flanking sequence surrounding the start codon, termed the Kozak sequence, is known to be enriched for specific nucleotides (Kozak, 1981). We investigated whether these unannotated ORFs possess an optimal Kozak sequence. While the mORFs clearly showed an enrichment of A/G at the –3 position and G at the +4 position (**Supplemental Fig. S3A–B**),

consistent with a previous study (Liu et al., 2013), other types of ORFs (uORFs, sORFs, and dORFs), generally did not exhibit strong nucleotide preferences at the -3, +4, or other surrounding positions (**Supplemental Fig. S3C–J**). Lacking a predictable Kozak sequence in these non-canonical ORFs highlights the need for empirical approaches to identify these translation events.

Proteomic analysis using mass spectrometry confirmed the production of stable proteins from at least some of these ncORFs, dORFs, and uORFs, as evidenced by the detection of multiple peptides from each ORF (4th column in **Fig. 2C** and **Supplemental Dataset S2A–B**). Note that despite our proteomic analysis being relatively deep as indicated by the number of annotated ORFs detected, we only detected a small fraction of ncORFs, dORFs, and uORFs by this approach. In our previous study, three out of four sORFs tested in Arabidopsis were detected by immunoblot analysis (Hsu et al., 2016). The discrepancy between these two protein detection approaches suggests that a proteomic method specifically optimized for short ORFs is required.

Among the 467 ncORFs that were at least 20 aa, TargetP v2.0 (Armenteros et al., 2019) predicted that 4, 28 and 128 are targeted to chloroplasts, mitochondria, or are secreted, respectively (**Fig. 2D**, **Supplemental Dataset S3**), suggesting their potential function in these subcellular localizations.

We analyzed the evolutionary conservation of 255 single-exon ncORFs using tBLASTn. Notably, only 22 of these ncORFs had homologs outside of the Brassicaceae among the 13 genomes we searched (**Supplemental Fig. S4**). These results are consistent with our previous findings in tomato, where most of the identified tomato sORFs were family- or species-specific (Wu et al., 2019). Together, these observations align with the notion that many sORFs are likely de novo genes that have emerged recently in evolution in eukaryotes (Ruiz-Orera et al., 2014; Sandmann et al., 2023).

We investigated these unannotated ORFs individually using RiboPlotR (Wu and Hsu, 2021) to evaluate their translation patterns. RiboPlotR displays Ribo-seq and RNA-seq data in the context of gene and transcript structures, and the RPFs within ORFs are color-coded according to the reading frame. Examples of ORFs translated from the Araport11 annotated 'novel transcribed regions' (**Fig. 2F**), the primary transcripts of miR163 and miR169 (**Fig. 2G–H**), and a newly identified uORF (**Fig. 2I**) are presented. We also confirmed the translation of QQS (**Fig. 2J**), a functional sORF with de novo origin (Li et al., 2009). It has been shown that the application of synthetic peptides encoded by the sORFs within the primary transcripts of miRNAs increases the abundance of the corresponding miRNAs (Lauressergues et al., 2015).

However, no significant RPFs or 3-nt periodicity had previously been detected in the primary transcripts of miRNAs in previous Ribo-seq data in plants. The two ORFs detected in pri-*MIR163* encode peptides of 9 and 25 aa, respectively (**Fig. 2G**), while the ORF detected in pri-*MIR169* encodes a 50-aa peptide (**Fig. 2H**). Although these three peptides are not conserved, the second peptide from pri-*MIR163* is predicted to localize to mitochondria according to TargetP.

Taken together, our new data with improved coverage expands the number of identified translated ORFs, particularly those of smaller size that are often overlooked in the annotation.

Translation detected in plastids and mitochondria

In addition to RPFs mapped to nuclear genes (hereafter referred to as ‘nucleus RPFs’), the sample preparation of our current study also included the addition of chloramphenicol, enabling us to examine the RPFs mapped to plastid/chloroplast and mitochondrial genes (hereafter referred to as ‘plastid RPFs’ and ‘mitochondria RPFs’). Consistent with previous observations (Chotewutmontri and Barkan, 2016), plastid RPFs and mitochondria RPFs only composed a small fraction of total RPFs (**Fig. 1B** and **Supplemental Fig. S5**).

Unlike nucleus RPFs, which mainly mapped to CDSs (**Fig. 1B** and **Supplemental Fig. S5**), Ribo-seQC analysis (Calviello et al., 2019) revealed that a substantial fraction of plastid and mitochondria RPFs map to intergenic regions, particularly 33 nt for plastid RPFs and 28-nt for mitochondria RPFs (**Supplemental Fig. S5**). For RPFs mapped to CDSs, the predominant RPF length was 28 nt for nucleus and chloroplast RPFs, while it was 29 nt for mitochondria RPFs (**Supplemental Fig. S5**). Overall, the size distribution of these organelle RPFs was similar to previous observations reported in maize (*Zea mays*) and Arabidopsis for plastid and mitochondrial RPFs (Chotewutmontri and Barkan, 2016; Planchard et al., 2018).

We analyzed the frame enrichment in individual RPF lengths using Ribo-seQC and found that 28 nt provides the highest frame enrichment in all three genomes (**Supplemental Fig. S6**). However, at this length, nucleus and plastid RPFs used frame 1 (red), whereas mitochondria RPFs used frame 3 (teal) (**Supplemental Fig. S6**). In general, plastid RPFs exhibited strong 3-nt periodicity across the 22–30-nt range, with 71–83% RPFs being in frame. By contrast, mitochondria RPFs displayed relatively weak 3-nt periodicity, with the best in-frame percentage being around 52% at either 28 nt or 22 nt (**Supplemental Fig. S6**). These observations suggest that while our ribonuclease digestion condition works well for nucleus and plastid RPFs, there is still room for optimization for mitochondria RPFs.

We also examined the RPFs mapped to annotated start codons to infer the P-site position in each RPF length (**Supplemental Fig. S7–9**). Overall, plastid RPFs and nucleus RPFs behaved similarly (**Fig. 3A, left**). For 28-nt RPFs, their P-site was positioned 12 nt downstream from the 5' end of the RPF, starting from the 13th nt (**Fig. 3A, left**, and **Supplemental Fig. S7, S8**). As the RPF lengths decreased, the RPFs became shorter at the 5' end, while the distance from the P-site to the 3' end remained constant at 16 nt (including the first nt of the P-site) (**Fig. 3A, left**). By contrast, the mitochondria P-site was located 7 nt downstream from the 5' end for 28-nt RPFs (**Fig. 3A, right** and **Supplemental Fig. S9**), and the distance from the P-site to the 3' end remained constant at 21 nt (**Fig. 3A, right**).

In agreement with the above results, metaplots showed that plastid RPFs exhibit similar patterns as nucleus RPFs, with strong 3-nt periodicity and enrichment of frame 1 (red) (**Fig. 1A** and **3B**). However, plastid RPFs lacked the signature blue signal (from frame 2) observed in nucleus RPFs at the codon preceding the stop codon (**Fig. 1A** and **3B**). By contrast, mitochondria RPFs showed relatively weak 3-nt periodicity and mainly used frame 3 (teal) (**Fig. 3C**).

Inspecting individual plastid and mitochondria genes also confirmed that plastid genes preferentially use frame 1 (red) (**Fig. 3D–E**), while mitochondria genes predominantly used frame 3 (teal) (**Fig. 3F**). Interestingly, two mitochondrial genes (ATMG00280 and ATMG01320) used frame 1 (red) (**Fig. 3G–H**), and they were predicted to encode proteins that function in the chloroplasts: ATMG00280 is annotated as a large subunit of RUBISCO, while ATMG01320 is annotated as a plastoquinone/NADH-ubiquinone protein. The usage of frame 1 (red) in these two mitochondrial genes suggests that the detected RPFs are likely derived from their plastid- and nuclear homologs (ATCG00490 and AT2G07689), respectively, and that ATMG00280 and ATMG01320 are not translated.

We observed substantial intron retentions in many plastid and mitochondrial genes; however, the RPFs mainly mapped to the CDS/exon regions and are sparse in the intron regions (e.g., **Fig. 3E–F**). These observations suggest that translation occurs after the introns are spliced and that not all genes in plastids and mitochondria necessarily use a coupled transcription-translation mechanism like their prokaryotic ancestors (Xiong et al., 2022; Trösch, 2022).

Together these results demonstrate that our data are useful to validate the annotated gene models in plastids and mitochondria and to study translation in these organelles.

sORFs detected in tasiRNAs

Next, we investigated unconventional translation events in annotated ncRNAs, focusing on tasiRNAs. The biogenesis of tasiRNAs involves miRNA-guided cleavage of the primary *TAS* transcripts (reviewed in (Fei et al., 2013)). There are four *TAS* gene families in Arabidopsis. Previous studies have reported the detection of RPFs on the primary transcripts of *TAS1–4*, especially at a sORF upstream of the miRNA target site (Li et al., 2016; Hsu et al., 2016; Bazin et al., 2017; Iwakawa et al., 2021). The RPF stalling associated with the sORF was shown to be important for tasiRNA production (Iwakawa et al., 2021). However, it remains unknown whether ribosomes in fact translate these sORFs. Only the sORFs in *TAS3A* and *TAS4* have been detected with significant 3-nt periodicity (Hsu et al., 2016; Hsu and Benfey, 2018). With the improved coverage of our dataset, RiboTaper identified actively translated ORFs across all four *TAS* gene families (**Figs. 4 and 5**). It is worth noting that translation occurs not only in the ORFs immediately adjacent to the miRNA target sites but also in ORFs upstream or downstream of the miRNA target sites.

In *TAS1A*, RiboTaper identified two ORFs (**Fig. 4A**). Interestingly, *TAS1A* contains an unannotated intron, suggesting that *TAS1A* may also be regulated by alternative splicing. ORF2 is located within this retained intron and overlaps with the miR173 target site (**Fig. 4A**). Importantly, we observed RPF stalling 16 nt upstream of the miR173 target site (**Fig. 4A**). Similarly, in *TAS1B*, RiboTaper detected two ORFs, and we detected strong RPF stalling 14 nt upstream of the miR173 target site (**Fig. 4B**). *TAS1C* has three ORFs identified by RiboTaper (ORF1, ORF2, and ORF4). Additionally, another ORF (ORF3) overlapped with the miR173 target site, and we detected strong RPF pausing signals within this ORF at 14 and 17 nt upstream of the miR173 target site (**Fig. 4C**). It should be noted that ORF3 was likely excluded by RiboTaper due to its overlap with ORF2, which uses a blue reading frame (**Fig. 4C**).

In *TAS2*, we observed strong 3-nt periodicity in ORF1, accompanied by distinct RPF stalling 16 nt upstream of the miR173 target site (**Fig. 4D**). We observed another strong pausing 30 nt upstream of the pausing adjacent to the miR173 target site.

Unlike *TAS1* and *TAS2*, *TAS3A* possesses two miR390 target sites. In line with our previous findings (Hsu et al., 2016; Hsu and Benfey, 2018), RiboTaper identified the ORF located entirely upstream of the first miR390 target site, with strong 3-nt periodicity (**Fig. 5A**). Moreover, we detected strong stalling 16 nt upstream of the first miR390 target site (**Fig. 5A**).

TAS4 has relatively low expression in our data. Nevertheless, consistent with our previous findings (Hsu et al., 2016), RiboTaper detected an ORF in the only annotated transcript (isoform 1). However, when we visualized the RNA-seq and Ribo-seq profiles, we realized that

TAS4 has an unannotated isoform (isoform 2). The RPFs located within the ORF in isoform 2 overlap with the miR828 target site (**Fig. 5B**).

Together, our data provide evidence of extensive RPFs associated with *TAS1*, *TAS2*, *TAS3*, and *TAS4*. The significant 3-nt periodicity observed in these ORFs strongly supports their active translation. In a previous study on *TAS2*, the position of ORF termination relative to the miR173 target site was shown to be crucial for tasiRNA biogenesis (Yoshikawa et al., 2016). In *TAS1*, *TAS2*, and *TAS3*, we observed that RPFs accumulate until 14–17 nt upstream of the miRNA target site. Considering that our RPFs were plotted using the first nt of the P-site (the 13th nt for 28-nt RPFs) and the distance between the P-site to the 3' end remained constant at 16 nt (**Fig. 3A**), the observed distance between RPF pausing and the miRNA target site coincides with the approximate length of the RPF from the P-site to the 3' end (illustrations in **Fig. 5C**). In line with the previously proposed 'ribosome stalling complex' model (Iwakawa et al., 2021), our observations suggest that most ribosomes can translate until they encounter the miRNA–ARGONAUTE (AGO)–SUPPRESSOR OF GENE SILENCING 3 (SGS3) complex, leading to ribosome stalling in *TAS1*, *TAS2*, and *TAS3* (**Fig. 5C**). These results are also consistent with the previous finding that a mutation within 6 nt upstream of the miR173 target site does not have observable effects (Yoshikawa et al., 2016), as the last ribosomal P-site is 16 nt upstream of the miRNA target site (**Fig. 5C**). Thus, our improved coverage data enhance the resolution and identification of translated ORFs on *TAS* transcripts.

Translation of dORFs

Next, we examined unannotated ORFs in protein-coding genes. In our dataset, RiboTaper identified 209 dORFs (**Fig. 2A** and **2B**). This class of ORFs has not been described in detail in plants, and their functional significance remains unknown. Interestingly, we found that the mORFs of genes containing dORFs exhibit higher translation efficiency compared to genes without dORFs, suggesting that dORFs are associated with genes whose transcripts exhibit high translation efficiency (**Fig. 6A**). Indeed, a gene ontology (GO) term enrichment analysis on the dORF-containing genes revealed enrichment for genes expected to be highly expressed and translated, such as ribosome and translation machinery (**Supplemental Dataset S4**).

We further evaluated these dORFs by analyzing their translational profiles (**Fig. 6B–F**). While we confirmed that these dORFs display strong 3-nt periodicity, we observed that most of the dORFs have low translation levels (**Fig. 6B–E**). Notably, their patterns could be categorized into two groups and the patterns suggest that they are unlikely to be reinitiation events following

the termination of the mORFs. In Type 1 dORFs, the mORF has relatively high translation levels (Fig. 6B–D), and substantial RPFs are present between the mORF and the dORF (Supplemental Fig. S10A–C). This finding suggests that the RPFs in the dORF may result from a readthrough of the mORF using frame 1 (red) (Supplemental Fig. S10A) or continued translation from an alternative ORF using a different frame (e.g., blue, Supplemental Fig. S10B–C).

Type 2 dORFs appear to originate from unannotated small genes/transcripts that are located downstream or overlapping with the 3' UTRs of annotated genes. These dORFs are likely derived from independent genes/transcripts, as their mRNA levels differ from those of the annotated genes (Fig. 6E–F). For example, there is a small gene that was annotated as part of the 3' UTR of isoform 2 of AT3G46640 *LUX ARHYTHMO* (*LUX*, encoding a key transcription factor in the circadian clock), and RiboTaper identified two translated ORFs of 58 and 54 aa in this unannotated gene (Fig. 6E). Similarly, we identified a dORF in the 3' UTR of AT1G49980 (Fig. 6F). The start codon of the dORF overlaps with the stop codon of the mORF (zoomed-in section in Fig. 6F), indicating that these two ORFs cannot be translated sequentially on the same mRNA. Published Cap Analysis of Gene Expression (CAGE) data (Thieffry et al., 2020) revealed additional transcription start sites associated with these dORFs (Supplemental Fig. S11A,B), providing further support that these unannotated ORFs, hidden within the 3' UTRs or downstream regions of other genes, originate from independent genes/transcripts.

Interestingly, a tBLASTn search of the 153 dORFs greater than 20 aa revealed that 11 of these dORFs are conserved throughout evolution (Supplemental Fig. S12). Importantly, all 11 conserved dORFs belong to Type 2 dORFs (Supplemental Fig. S12). This finding suggests that the peptide sequences encoded by these unannotated genes (dORFs) likely correspond to functional proteins.

Taken together, the dORFs we detected are unlikely to be reinitiation events analogous to the uORF–mORF pairs. Type 1 dORFs potentially result from readthrough events, which have been recently reported in Arabidopsis (Sahoo et al., 2022), or they could be part of alternative ORFs (Fig. 6G, left); further investigation is needed to elucidate the underlying mechanisms. Type 2 dORFs revealed the presence of unannotated small genes hidden in the 3' region of other genes (Fig. 6G, right).

Ribosome queueing at CPuORFs links to co-translational decay

We next investigated uORFs, which represent the largest class of unannotated ORFs we identified (Fig. 2A). CPuORFs are of great interest as they are associated with substantial

ribosome pausing/repression, likely through an interaction between the nascent peptide and the exit tunnel within ribosomes. Notably, several CPuORFs have been reported to interact with certain metabolites (reviewed in (Van Der Horst et al., 2020)). CPuORFs are also a focus of co-translational decay studies (Yu et al., 2016; Hou et al., 2016; Guo et al., 2023a). Co-translational decay demonstrates that mRNAs can undergo degradation while bound by translating ribosomes. As a result, the degraded mRNA fragments display features associated with translation (Hu et al., 2009; Pelechano et al., 2015; Merret et al., 2015; Yu et al., 2016; Hou et al., 2016). Global analysis of 5' ends of mRNA degradation fragments in Arabidopsis detected mRNA fragments with ~30-nt periodicity (approximate length of RPFs) preceding the stop codons of several CPuORFs, supporting the idea that several ribosomes tightly stack behind a pausing ribosome in these regions (Yu et al., 2016; Hou et al., 2016; Guo et al., 2023a). However, the corresponding 30-nt periodicity of RPFs was not detected in previous Ribo-seq data (Hou et al., 2016).

In Arabidopsis, 119 CPuORFs have been identified, classified into 81 homology groups based on the conservation of the uORF peptide sequences (reviewed in (Van Der Horst et al., 2020)). We examined the RPFs in different homology groups of CPuORFs and found those in Homology Group 1 show clear periodic ribosome pausing patterns. Specifically, we observed 2–4 distinct RPF peak(s) in 30-nt intervals upstream of the stop codons of CPuORF1, CPuORF2, and CPuORF3, within *BASIC LEUCINE ZIPPER 2 (bZIP2)*, *bZIP11*, and *bZIP53* transcripts, respectively (**Fig. 7A–C**). These patterns are consistent with the expectation that multiple ribosomes line up behind a pausing ribosome prior to the stop codon of these CPuORFs (see illustrations of ribosome positions above the RPF peaks in **Fig. 7A–C**). We also observed a similar pattern in *MYB51* transcripts (**Fig. 7D**), which contain a Brassicaceae-specific CPuORF (Hou et al., 2016). Therefore, our enhanced Ribo-seq data provide evidence for periodic ribosome pausing upstream of the stop codons of CPuORFs, further supporting the occurrence of co-translational decay in plants.

Identification of uORFs using RiboTaper

Besides CPuORFs identified by evolutionary conservation, we also investigated uORFs identified through our experimental approach. uORFs represent the shortest unannotated ORFs (**Fig. 2E**), and their identification could be most affected by RPF coverage. Our new data substantially improved the detection of uORFs based on the 3-nt periodicity, increasing the number from the previous 187 uORFs to 2113 uORFs (**Fig. 2A,B, Supplemental Dataset S1A**). Strong 3-nt periodicity within the uORF regions in the translational profiles supports the active

translation of these uORFs (examples shown for *RIBOSOMAL S6 KINASE 1 (S6K1)* and *REPRESSOR OF GA 1 (RGA1)* in **Fig. 8A,B**). Comparing uORF genes with similar mRNA levels in our previous and current datasets, it is evident that both the uORF and mORF regions exhibited improved coverage in our current data (**Fig. 8A,B**).

Consistent with our expectations, these translated uORFs were associated with lower mORF translation efficiency (**Fig. 8C**). Furthermore, genes that have a higher number of translated uORFs exhibited lower mORF translation efficiency (**Fig. 8C**).

To validate that these uORFs function as translational repressors, we performed a dual-luciferase assay comparing wild-type uORFs with mutated uORFs, where the start codon ATG was mutated to AGG, in *S6K1* and *RGA1* (**Fig. 8D** and **8E**). Upon mutating the uORFs, the downstream firefly luciferase (FLUC) reporter showed increased protein levels/activity (**Fig. 8E**), while *FLUC* transcript levels showed no significant changes (**Fig. 8F**). These results suggest that these uORFs normally repress the translation of their downstream mORFs.

Identification of uORFs using CiPS

Although the uORFs identified by RiboTaper show significant 3-nt periodicity, we observed that globally, uORFs have the lowest 3-nt periodicity compared to CCDSs and ncORFs (**Fig. 9A**), and the 3-nt periodicity appears to be inversely correlated with ORF length among these ORF groups (**Fig. 9A** and **Fig. 2E**). We reasoned that the overall 3-nt periodicity is affected by the atypical RPFs observed at the codon proceeding the stop codon (the -1 codon). In this position, the P-sites of many RPFs mapped to frame 2 (blue) instead of the expected frame (red) (**Fig. 1A** and illustrations in **Fig. 9B–E**). This pattern likely arises from the binding of release factors, rather than a charged tRNA, to the A-site during termination, leading to a different ribosomal conformation and resulting in distinct RPF patterns (Alkalaeva et al., 2006; Brown et al., 2015; Matheisl et al., 2015). In addition, given that termination is a major rate-limiting step during translation, more RPFs are associated with this position (Wolin and Walter, 1988; Ingolia et al., 2009). Importantly, this out-of-frame signal (blue) can disproportionately and substantially lower the 3-nt periodicity of short ORFs and influence their identification based on 3-nt periodicity (illustrations in **Fig. 9B–E**). Indeed, comparing the size distribution of predicted uORFs (i.e., AUG-initiated ORFs in the 5' UTR) and those identified by RiboTaper revealed that RiboTaper clearly biases towards longer uORFs (**Fig. 9F**, median length of 9 and 21, respectively).

To improve the identification of short uORFs, we developed a new approach called CiPS (Count, in-frame Percentage and Site). In brief, we first classified the blue signal at the -1

codon as being in frame. We then applied three criteria to consider if a uORF is translated by evaluating 1) RPF counts, 2) in-frame RPF percentage, and 3) the percentage of occupied in-frame sites (see MATERIALS AND METHODS for detail).

In total, CiPS identified 6,539 translated uORFs (**Supplemental Dataset S5A**), including 370 minimum uORFs (1 aa) and 388 uORFs of 2 aa in length (**Supplemental Dataset S5B, C**). By contrast, RiboTaper identified no minimum uORFs and only one 2-aa uORF (**Supplemental Dataset S1A**). CiPS successfully identified a substantial number of short uORFs (**Fig. 9H–I**). Notably, the uORFs identified by CiPS exhibited a similar size distribution as predicted uORFs (**Fig. 9F**). RiboTaper and CiPS identified 1653 uORFs in common, while 460 uORFs only detected by RiboTaper and 4,886 uORFs only detected by CiPS (**Fig. 9G**). The uORFs identified exclusively by RiboTaper had a marginal effect on the translation efficiency of their mORFs (**Fig. 9J**). By contrast, the uORFs identified by CiPS only or by both methods showed significantly stronger repression (**Fig. 9J**).

CiPS also improved the number of uORFs identified within individual transcripts, with the highest number of 7 uORFs detected within a transcript (**Fig. 9K**). In the example of *AUXIN RESPONSE FACTOR 6* (*ARF6*) transcript (**Fig. 9L**), there are six potential AUG-initiated uORFs. CiPS identified uORF-1, uORF-4, uORF-5, and uORF-6. By contrast, RiboTaper only identified uORF-6. We also observed a noncanonical uORF initiated by an AUA start codon (**Fig. 9L**). uORF-1 and 4 were excluded by RiboTaper, likely due to the strong blue signal at the –1 codon, as well as the partial overlap between uORF-4 and neighboring uORFs (**Fig. 9L**).

Consistent with our expectations and the findings of RiboTaper (**Fig. 8C**), genes containing a higher number of CiPS-identified uORFs showed a correlation with lower mORF translation efficiency (**Fig. 9M**).

A previous survey of 23 plant uORFs in the literature suggested that longer uORFs are associated with stronger repression of their mORFs (Von Arnim et al., 2014). As the 23 uORFs characterized are enriched for CPuORFs, which are known to have special properties (Van Der Horst et al., 2020), we investigated whether there is a general relationship between uORF length and repression among diverse uORFs. We compared genes with one translated uORF encoding 1, 2–10, 11–20, 21–30, or > 30 aa, and found that they similarly repress the translation efficiency of their mORFs, compared to genes without any translated uORFs (**Fig. 9N**). This result suggests that these translated uORFs, regardless of their length, are equally powerful translational repressors on a global scale.

Characterization of minimum uORFs and tiny uORFs

By definition, 'AUG-stop' represents the shortest possible uORF, known as a minimum uORF (illustrations in **Fig. 9C**). In our dataset, 746 minimum uORFs contained at least one RPF. As minimum uORFs only have 1 codon, we removed the in-frame site criterion of CiPS accordingly and uncovered 370 minimum uORFs that passed the remaining criteria (**Supplemental Dataset S5B**). This list included the two minimum uORFs previously reported in *NOD26-LIKE INTRINSIC PROTEIN 5;1 (NIP5;1)* (**Fig. 10A**), and the minimum uORFs reported in *SKU5* and *BOR2* (**Fig. 10B,C**), which are all related to boron transport and response (Tanaka et al., 2016; Sotta et al., 2021). Our data did not detect the reported minimum uORF in *ABNORMAL SHOOT 2 (ABS2)*, as in our experimental conditions, *ABS2* uses a transcription start site downstream of the potential minimum uORF (**Fig. S13**). Notably, we detected new minimum uORFs in several important genes, including *PBS1-LIKE 36 (PBL36)*, encoding a receptor-like kinase involved in regulating shoot and root meristems; *SNF1-RELATED PROTEIN KINASE 2.2 (SnRK2.2)*, encoding a key kinase in the ABA signaling pathway; and *NONEXPRESSER OF PR GENES 1-LIKE PROTEIN 3 (NPR3)*, encoding a salicylic acid receptor and a negative regulator of plant pathogen response (**Fig. 10D-F**). As expected, these minimum uORFs showed the characteristic blue signal at the -1 codon, since minimum uORFs consist of only one codon, which is also immediately upstream of the stop codon (**Fig. 10A-F**, illustrations in **Fig. 9C**).

As shown in **Fig. 9N**, globally, minimum uORFs showed similar repression effects on downstream mORF translation, compared to longer uORFs. To validate the functions of these minimum uORFs, we performed a dual-luciferase assay to test the uORFs in *BOR2*, *SnRK2.2* and *NPR3* (**Fig. 10G** and **10H**). We observed that mutations in the uORFs lead to increased FLUC activity levels without significantly affecting *FLUC* transcript levels (**Fig. 10H** and **10I**). These results support the notion that these minimum uORFs normally repress the translation of their downstream mORFs.

Like minimum uORFs, other uORFs of extremely short lengths are difficult to identify based on 3-nt periodicity (illustrations in **Fig. 9D-E**). CiPS successfully identified 2,921 tiny uORFs (2–10 aa) (**Supplemental Dataset S5C**). Examples of 2-, 3-, 4-, 5-, and 8-aa uORFs are shown in **Fig. 11A-D** and **Supplemental Fig. S14A-B**. Similar to minimum uORFs, mutations in these tiny uORFs result in increased FLUC activity levels without significantly increasing *FLUC* transcript levels (**Fig. 11E**, **Supplemental Figs. S15** and **S14C,D**). These results support the idea that these tiny uORFs normally repress translation of their mORF and are consistent with their repressive effects observed in global analysis (**Fig. 9N**).

Overlap of translated uORFs

Case studies using reporter assays have suggested that the CPuORFs in Homology Groups 1 and 3 can be regulated by their overlapping uORFs (Hanfrey et al., 2002; Wiese et al., 2004; Hanfrey et al., 2005). In our dataset, we detected thousands of transcripts that possess multiple translated uORFs (**Fig. 9K**), with some uORFs overlapping with each other (e.g., **Fig. 9L**). However, the extent of uORF overlapping at a genome-wide scale has remained unclear, especially since a comprehensive list of translated uORFs was previously lacking. As CiPS permits some level of overlap between uORFs, our new dataset offers an opportunity to identify uORF stacking globally. We systematically identified the overlapping events of translated uORFs on the most abundant transcript isoforms in our data, resulting in the identification of 681 overlapping uORFs in 317 transcripts (**Supplemental Dataset S6A**).

It is noteworthy that out of the 92 CPuORF transcripts detected in our data, 18 possessed additional translated uORFs that stack with the CPuORF. Remarkably, we detected stacking uORFs in all five members of Homology Group 1 CPuORFs within *bZIP1*, *bZIP2*, *bZIP11*, *bZIP44*, and *bZIP53* mRNAs (**Supplemental Dataset S6B**). This result is in line with the notion that the five members in Homology Group 1 contain an overlapping uORF upstream of the CPuORFs, which is also conserved (Wiese et al., 2004). Examples of uORF/CPuORF overlapping within *bZIP11* (Homology Group 1), *S-ADENOSYLMETHIONINE DECARBOXYLASE 2* (*SAMDC2*, Homology Group 3), and *CBL-INTERACTING PROTEIN KINASE 6* (*CIPK6*, Homology Group 27) are presented in **Fig. 12A–C**. The translational regulation of *bZIP11* by uORFs has been extensively studied (Rook et al., 1998; Wiese et al., 2004; Rahmani et al., 2009; Roy et al., 2010). In *bZIP11*, three other uORFs stack with the CPuORF, but only the first uORF (18 aa) was detected as translated (**Fig. 12A**), which is consistent with the previous observation that the first uORF regulates the CPuORF in reporter assays (Roy et al., 2010). Similarly, in *SAMDC2*, one 3-aa uORF was translated and stacked with the CPuORF (**Fig. 12B**); in *CIPK6*, one 12-aa uORF was translated and stacked with the CPuORF (**Fig. 12C**). These uORF stacking events, particularly the conserved stacking pattern in Homology Group 1, suggest that these upstream stacking uORFs potentially regulate the translation of CPuORFs.

Translated uORFs affected by alternative splicing

As the presence of potential uORFs is determined by the 5' UTR sequences, the occurrence of translated uORFs may be regulated by alternative splicing in the 5' UTRs. Two such examples have been reported in *HCS1*, encoding a holocarboxylase synthetase, and

PHYTOCHROME-INTERACTING FACTOR 3 (PIF3), encoding a repressor of light responses. In the case of *HCS1*, the presence or absence of the isoform-specific uORF leads to different translational initiation sites and alters the subcellular targeting of *HCS1* (Puyaubert et al., 2008). In the case of *PIF3*, an intron retention within the 5' UTR is regulated by phytochrome B to repress the translation of *PIF3* transcript (Dong et al., 2020). Although the particular isoforms reported to affect the uORFs in *HCS1* and *PIF3* were not significantly expressed under our experimental conditions (**Supplemental Fig. S16A–B**), by examining the two most abundant isoforms of expressed genes in our data, we identified 594 translated uORFs that are affected by alternative splicing within 399 genes (**Supplemental Dataset S7A, 7B**).

These uORFs are linked to specific isoforms through various alternative splicing events, including intron retention, alternative 5' donor site, alternative 3' acceptor site, and exon skipping (**Fig. 13A–D**). Notably, *TOPELESS*, an important regulator in the auxin and jasmonic acid (JA) signaling pathways, contained a cassette exon, which includes a translated uORF in isoform 2 (**Fig. 13D**). These examples illustrate that the presence of and regulation by a uORF may be widely controlled by alternative splicing and suggest the prevalence of isoform-specific translational control.

Pathways regulated by uORFs

A prior bioinformatic analysis indicated that predicted uORFs are enriched in genes encoding transcription factors and protein kinases (Kim et al., 2007). To investigate what molecular functions and pathways are controlled by translated uORFs, we performed a GO term enrichment analysis on genes containing uORFs identified by RiboTaper or CiPS (**Supplemental Fig. S17**). Consistent with the previous prediction, we identified protein phosphorylation/modification, signal transduction, regulation of gene expression, in the 'Biological Process' terms, and protein kinase activity, transcription regulator activity, DNA binding in the 'Molecular Function' terms. Interestingly, we identified many membrane-associated terms, including peroxisomal membrane, endosomes, Golgi, vesicles, plasma membrane in the 'Cellular Component' terms (**Supplemental Fig. S17**).

We also surveyed whether genes in various critical pathways harbor translated uORFs. Our analysis revealed that many key components in the light signaling, circadian clock, eight phytohormone signaling pathways (auxin, abscisic acid [ABA], JA, ethylene, gibberellin [GA], salicylic acid [SA], cytokinin, and brassinosteroids), development, as well as processes related to translational control, RNA biology, or signaling, contain translated uORFs (**Table 1**). These findings suggest that these important plant pathways are under translational control through

uORFs. Therefore, our comprehensive uORF list serves as a valuable resource for further understanding gene expression regulation and plant growth.

DISCUSSION

In this study, we tackled the identification of short ORFs by improving the coverage of Ribo-seq and developing a new computational pipeline for uORF identification. Our results provide a clearer view of the translational landscape and useful resources for diverse gene regulation studies, especially for uORFs on protein-coding transcripts and sORFs within presumed non-coding RNAs.

While a large portion of sORFs may be simply misannotated due to constraints of computational annotations, the translation of sORFs within well-characterized non-coding RNAs, such as primary transcripts of tasiRNAs and miRNAs, highlights the existence of 'dual function' RNAs (reviewed in (Ulveling et al., 2011; Raina et al., 2018)). That is, some RNAs may function as regulatory RNAs while also possessing protein-coding capacity. In the case of tasiRNAs, the translation of sORFs within primary transcripts regulates the biosynthesis of tasiRNAs (Zhang et al., 2012; Yoshikawa et al., 2016; Li et al., 2016; Hou et al., 2016; Bazin et al., 2017; Iwakawa et al., 2021). In the case of miRNAs, application of synthetic micropeptides encoded by the sORFs in several pri-MIRNAs increases the abundance of the pri-MIRNAs or mature miRNAs (Lauressergues et al., 2015; Sharma et al., 2020). For both tasiRNAs and pri-miRNAs, the translation and the regulatory RNA act in the same pathways, but it remains possible that the two functions act in different pathways, as shown in bacteria (reviewed in (Raina et al., 2018)). Only a limited number of sORFs have been characterized in plants so far, but these sORFs clearly have wide ranges of functions (Hanada et al., 2013; Tavormina et al., 2015; Hsu and Benfey, 2018; Fesenko et al., 2019; Ong et al., 2022). Studies in bacteria and mammals revealed that dozens of sORFs encode membrane proteins (reviewed in (Orr et al., 2020)). These small proteins can be a component of a large protein complex, or involved in the recruitment of other proteins to membranes, the assembly of protein complexes within or near membranes, and controlling protein stability or activity of other proteins. The prediction that a substantial portion of Arabidopsis sORFs we detected are secreted proteins offers a promising direction for functional characterization of these sORFs.

We found that out-of-frame (frame 2) RPFs at the codon preceding the stop interfere with the identification of short uORFs. In eukaryotes, translational termination occurs when the ribosomal A-site encounters one of the three stop codons, followed by eRF1/eRF3 complex

binding to the A-site and releasing the nascent peptide (reviewed in (Schuller and Green, 2018)). Cryogenic electron microscopy (Cryo-EM) analysis of mammalian ribosomes at termination revealed that eRF1 binding ‘pulls’ the nucleotide downstream of the stop codon into the A-site, thus packing 4 nt of mRNA at the A-site (Brown et al., 2015; Matheisl et al., 2015). We reasoned that the frame-2 RPFs at the codon preceding the stop resulted from the structural rearrangement of ribosomes upon eRFs binding to the A-site during termination. Comparing the same length of RPFs from an elongating ribosome and a terminating ribosome, the packing caused by eRF1 binding would cause the RPF of a terminating ribosome to be 1 nt shorter at the 5’ end and 1 nt longer at the 3’ end. This 1-nt shifting toward the 3’ end causes the RPFs of terminating ribosomes to be mapped to the next reading frame (frame 2). Therefore, the frame-2 mapping of terminating RPFs we observed in Arabidopsis and tomato (Hsu et al., 2016; Wu et al., 2019) suggests that plant ribosomes behave like mammalian ribosomes at translation termination. Although plastid RPFs behave similarly to nucleus RPFs in multiple ways (**Figs. 1A**, and **3A–B**), plastid RPFs lacked frame-2 mapping at termination (**Fig. 3B**). Currently, little is known about the mechanism of translation termination in plastid and mitochondrial ribosomes, which have prokaryotic origins (Zoschke and Bock, 2018). In eukaryotes, eRF1 recognizes all three stop codons, whereas in prokaryotes, two different release factors recognize specific stop codons. Moreover, the release factors in eukaryotes and prokaryotes are evolutionarily unrelated (reviewed in (Buskirk and Green, 2017)). Our observations that plastid RPFs lack frame-2 mapping at termination suggest that the mRNA packing caused by release factor binding may not occur in plastids. This phenomenon is likely the case for plant mitochondria as well, although specialized Ribo-seq optimizing for mitochondrial ribosomes will be necessary to draw a clear conclusion.

Our CiPS pipeline, which accepts the frame-2 mapping of terminating RPFs as in-frame, facilitates the identification of uORFs with relatively short lengths. Minimum uORFs in particular have only one codon and are therefore impossible to be identified based on 3-nt periodicity. Given that the codon usage of uORFs are similar to random triplet sequences in 5’ UTRs and most uORF peptides are not evolutionarily conserved, it is believed that most uORFs are selected for their regulatory role, rather than their encoded peptides (Von Arnim et al., 2014; Fields et al., 2015; Johnstone et al., 2016; Van Der Horst et al., 2020). Modulating uORFs has emerged as a promising approach to controlling gene expression and selection for agricultural traits (Xu et al., 2017; Zhang et al., 2018; Xing et al., 2020; Gage et al., 2022). As our results show that short uORFs are equally powerful as longer uORFs and most short uORFs are largely unexplored in previous studies, our uORF list offers new opportunities for gene

regulation in diverse plant pathways, especially for genes encoding transcription factors and protein kinases, critical regulators in cellular signaling.

MATERIALS AND METHODS

Plant growth conditions and lysate preparation

Arabidopsis (Arabidopsis thaliana) Col-0 seeds were surface sterilized with 70% (v/v) ethanol for 5 min, followed by 33% (v/v) bleach with 0.03% (v/v) Tween 20 for 10 min, then rinsed with sterile water 5 times. The seeds were stratified at 4°C in the dark for 2 d, then grown hydroponically in sterile liquid medium (2.15 g/L Murashige and Skoog salts [Caisson Laboratories], 1% [w/v] sucrose, 0.5 g/L MES, pH 5.7) while shaking at 85 rpm under a 16-h light (75–80 $\mu\text{mol m}^{-2}\text{s}^{-1}$ from cool white fluorescent bulbs [Philips F17T8]) and 8-h dark cycle at 22°C for 7 d. At Zeitgeber time 4 (4 h after lights on), DMSO corresponding to 0.1% of the medium volume was added to the medium (these were control samples of our large-scale chemical treatment experiment). After 20 or 60 min, three biological replicates (~300 seedlings per sample) were harvested at each time point and immediately flash-frozen with liquid nitrogen.

Plant lysates were prepared as previously described (Hsu et al., 2016). Briefly, each 0.1 g of ground tissue powder was resuspended in 400 μL of lysis buffer (100 mM Tris-HCl [pH 8.0], 40 mM KCl, 20 mM MgCl_2 , 2% [v/v] polyoxyethylene (10) tridecyl ether [Sigma, P2393], 1% [w/v] sodium deoxycholate [Sigma, D6750], 1 mM dithiothreitol, 100 $\mu\text{g/mL}$ cycloheximide [Sigma, C4859], 100 $\mu\text{g/mL}$ chloramphenicol [Sigma R4408], and 10 units/mL DNase I [Epicenter, D9905K]). The lysates were spun at 3,000 g at 4°C for 3 min, and the supernatant was transferred to a new tube and subsequently centrifuged at 20,000 g at 4°C for 10 min. The supernatant was transferred to a new tube and the RNA concentration was determined with 10x dilutions using a 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 processing.

Ribo-seq library construction

Briefly, ribosome footprints were processed using 200 μL of the lysates described above, and sequencing libraries were constructed according to our previous method (Hsu et al., 2016) with a few modifications described below. Please note that the TruSeq Mammalian Ribo Profile Kit (illumina, RPHMR12126) and RiboZero Plant Leaf kit (illumina, MRZPL1224) described here

have been discontinued; researchers who are interested in our method should reference our custom library construction protocol (Wu and Hsu, 2022) instead.

The 200 μ L of the lysates described above were treated with RNase I (50 units nuclease per 40 μ g of RNA; the nuclease was included in the TruSeq Mammalian Ribo Profile Kit, Illumina, RPHMR12126) for 1 h at room temperature with gentle mixing. Then, 15 μ L of SUPERase-IN (Invitrogen, AM2696) was added, and the lysate was passed through a size exclusion column (Illustra MicroSpin S-400 HR Columns; GE Healthcare, 27-5140-01). RNA > 17 nt was isolated with a RNA Clean & Concentrator-25 kit (Zymo Research, R1017) and separated on 15% (w/v) urea-Tris borate EDTA (TBE) gels (Invitrogen, EC68852BOX). Gel slices roughly between 27 nt and 31 nt were isolated, and the RNAs were purified as previously described. Next, rRNA depletion was performed using a RiboZero Plant Leaf kit (Illumina, MRZPL1224) in one quarter of the recommended reaction volume. Ribo-seq libraries were constructed using a 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 a Hi-Seq 4000 sequencer using single-end 50-bp sequencing.

We reasoned that the lower coverage in our previous method may have been caused by 1) some lost footprints during the purification, 2) too little starting materials. In our previous protocol, we purified the RNA twice (first >17 nt and then <200 nt) after the size exclusion column step to concentrate ribosome footprints before rRNA depletion by RiboZero. We believed that some footprints might be lost during the two rounds of RNA purification. In addition, the rRNA depletion by RiboZero prior to size selection by gel limited the amount of input RNA that could be used according to the manufacturer's recommendations. Thus, we made the following changes to our protocol: reduced the RNA purification to one round (>17 nt), isolated RPFs between 27 nt and 31 nt, and then performed rRNA depletion by RiboZero. These changes are intended to maximize the input and minimize the footprint loss during the procedure.

RNA-seq library construction

Total RNA greater than 200 nt was purified from 100 μ L of the lysates described above using a RNA Clean & Concentrator-25 kit (Zymo Research, R1017) as previously described (Hsu et al., 2016). 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 μ g of RNA per sample was subjected to rRNA depletion using a RiboZero Plant Leaf kit (Illumina, MRZPL1224) following the manufacturer's recommendations. Then, 100 ng of

rRNA-depleted RNA was fragmented to around 200 nt in length based on the RIN reported by the Bioanalyzer, and strand-specific sequencing libraries were made using a 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 a Hi-Seq 4000 using paired-end 100-bp sequencing.

Sequencing data pre-processing and analysis

Data pre-processing and analysis were performed as previously described (Hsu et al., 2016), except that the Araport11 annotation (Cheng et al., 2017) 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) (Langmead and Salzberg, 2012) to remove rRNA, tRNA, snRNA, and snoRNA contaminant sequences.

Transcriptome assembly, ORF identification using RiboTaper, statistical analysis and data visualization

For transcriptome assembly, the RNA-seq data from all six samples were first combined and mapped with STAR aligner (Dobin et al., 2013) with the following parameters --alignIntronMax 5000, --alignIntronMin 15, --outFilterMismatchNmax 2, --outFilterMultimapNmax 20, --outFilterType BySJout, --alignSJoverhangMin 8 and --alignSJDBoverhangMin 2. The resulting bam file was used for reference-guided transcriptome assembly with Stringtie (Pertea et al., 2015) following our previous pipeline (Wu et al., 2019). We used gffcompare (Pertea and Pertea, 2020) to compare Stringtie output gtf file with the Araport11 annotation and selected the newly assembled transcripts (i.e., transcript types i, x, y, o, u and s).

Both the RNA-seq and Ribo-seq reads were mapped to the newly assembled gtf file (**Supplemental File S1**, which contains both Araport11 and newly assembled transcripts) with STAR aligner (RNA-seq parameters: identical to above mentioned prior to the transcriptome assembly; Ribo-seq parameters: --alignSJoverhangMin 4, --alignSJDBoverhangMin 1, --outSAMmultNmax 1, and the remaining parameters were identical to RNA-seq). Combining all six samples, the total reads mapped to the genome for Ribo-seq and RNA-seq were 298.01 million and 180.67 million pairs, respectively.

To identify translated ORFs, we used the bam files mapped above from both RNA-seq and Ribo-seq as input for RiboTaper (v1.3.1a) (Calviello et al., 2016). The Ribo-seq metaplot and the distribution of Ribo-seq reads in different genome features were generated using Ribo-

seQC (Calviello et al., 2019) with 10% randomly selected reads. The Ribo-seq read lengths and offsets used in RiboTaper analysis were 24, 25, 26, 27, 28, and 8, 9, 10, 11, 12, respectively, as determined from the metaplots for nuclear and chloroplast genes. The resulting unannotated ORFs and annotated mORFs were extracted from the RiboTaper output ORF_max_filt file. Since the P-site offsets for mitochondrial genes are different from the nuclear and chloroplast genes, the translational profiles of mitochondria-encoded genes were manually visualized using RiboPlotR (Wu and Hsu, 2021).

All data visualization and statistical analysis were performed in R (v4.0.3) (Da Rosa et al., 2004). For Ribo-seq and RNA-seq data visualization, the P_sites_all files from RiboTaper were first processed using the following code: `cut -f 1,3,6 P_sites_all | sort | uniq -c | sed -r 's/^( *[^ ]+)+\1\t' > output.txt` to aggregate the read counts at each P-site. The Ribo-seq and RNA-seq profiles were plotted using RiboPlotR (Wu and Hsu, 2021). RiboPlotR presents the Ribo-seq data in the context of gene and transcript structure with exon-intron junctions, and the RPFs within ORFs are color-coded to indicate the reading frames.

Proteomics sample preparation and analysis

Proteomic experiments were carried out using shoot and root tissues from 4- and 21-d-old plants as well as the root tissue from 12-d-old seedlings (Arabidopsis accession Col-0). The protein extraction and digestion using trypsin and Lys-C were carried out based on established methods (Song et al., 2018b, 2018a). Samples for datasets that include tandem mass tag (TMT) labeling and/or phosphopeptide enrichment were prepared as previously described (Clark et al., 2021; Montes et al., 2022; Song et al., 2020). Two-dimensional high-performance liquid chromatography (HPLC) fractionation was performed using either strong cation exchange or basic reversed phase fractionation. Fractionated peptides were delivered to a Q Exactive Plus mass spectrometer using either an Agilent 1260 quaternary or a Thermo U3000 HPLC. Data-dependent acquisition was performed using Xcalibur 4.0 software in positive ion mode with a capillary temperature of 275°C and an RF of 60. MS1 spectra were measured at a resolution of 70,000 while MS2 spectra were measured at a resolution of 17,500 (label free) or 35,000 (tandem mass tag [TMT] labeled). All raw data were analyzed together using MaxQuant version 1.6.7.0 (Tyanova et al., 2016). Spectra were searched against a custom protein database generated from the RiboTaper output file ORFs_max_filt (Supplemental Dataset S1), which was complemented with reverse decoy sequences and common contaminants by MaxQuant. Carbamidomethyl cysteine was set as a fixed modification while methionine oxidation and protein N-terminal acetylation were set as variable modifications.

Phosphorylation (S, T, Y) was also set as a variable modification for samples that were phosphopeptide-enriched. Digestion parameters were set to “specific” and “Trypsin/P;LysC”. For the TMT experiments, the sample type was set to “Reporter Ion MS2”. Up to two missed cleavages were allowed. A false discovery rate of less than 0.01 at both the peptide spectral match and protein identification level was required. The “second peptide” option was used to identify co-fragmented peptides.

Calculating mORF translation efficiency

We first used STAR to map the RNA-seq and Ribo-seq reads to the CDS of annotated protein-coding genes. The resulting bam files were used to quantify the transcripts per million (TPM) of each gene using RSEM (v1.3.1) (Li and Dewey, 2011). Then, the mORF translation efficiency was calculated by dividing the Ribo-seq TPM to the RNA-seq TPM.

uORF identification using CiPS (Count, in-frame Percentage and Site)

The processed RiboTaper P-site file (i.e., output.txt, see above) was analyzed for identification of small uORFs. To accommodate the expected ribosome conformational change at termination, we accepted the mapping of the second reading frame (blue) at the -1 codon as in-frame. We then applied the following criteria to consider if a uORF is translated: 1) ≥ 10 RPF counts, 2) in-frame RPF percentage $\geq 50\%$ (identical to the RiboTaper cut-off), 3) occupied RPF site $\geq 30\%$ in-frame (i.e., only the Ribo-seq occupied P-sites were evaluated and the in-frame % need to be $\geq 30\%$); this last criterion allows some tolerance if a uORF overlaps with other uORFs. These cut-offs were empirically determined based on our data. For simplicity, for each gene, the two most abundant transcript isoforms (determined by Kallisto (Bray et al., 2016)) with identical mORF starts were used for CiPS analysis. If the two isoforms had different mORF starts, the one with the more upstream mORF start was used for the analysis to avoid false positives of uORF identification. The duplicated uORFs shared by the two isoforms were removed. For minimum uORFs, since there is only one codon, the ‘occupied RPF site $\geq 30\%$ in-frame’ criterion was eliminated.

To identify translated uORFs that overlap with each other, we used CiPS-identified uORFs within the most abundant isoforms for the analysis.

For uORFs regulated by alternative splicing, we considered the two most abundant isoforms and identified translated uORFs (detected by CiPS) that only exist in one of the two isoforms.

R codes for Ribo-seq analysis, RiboTaper pipeline, identifying CiPS uORFs, the overlapping between uORFs, and uORF regulated by alternative splicing are available at: <https://github.com/hsinyenwu/ORFeome>

Constructs for dual luciferase assays

For the dual luciferase plasmid, the cauliflower mosaic virus (CaMV) 35S promoter sequence upstream of *Renilla Luciferase (REN)* within the pGreen II 0800 Luc plasmid (Hellens et al., 2005) was synthesized by BioBasic and cloned into the upstream of *Firefly LUCIFERASE (FLUC)* reporter gene via KpnI and XhoI restriction sites. The resulting construct was renamed pHsu-133. All 5' UTR sequences tested were synthesized by BioBasic and cloned into pHsu-133 between the 35S promoter and *FLUC* via BamHI and NotI restriction sites. The exact 5' UTR sequences tested are listed in **Supplemental File S2**. For *S6K1*, *RGA1* constructs, annotated 5' UTR sequences containing the wild-type or mutated uORF (ATG was changed to AGG) were used. For minimum uORF constructs, annotated 5' UTR sequences containing the wild-type or mutated uORF (ATG was changed to ATC) were used. For tiny uORF constructs, 5' UTR sequences containing the wild-type or mutated uORF (ATG was changed to AGG), including 40 nt upstream of the uORF to the end of 5' UTR, were used.

Dual luciferase assays and RT-qPCR

Arabidopsis protoplast preparation and transformation were modified from (Reis et al., 2020). Protoplasts were isolated from fully expanded rosette leaves of 20- to 21-d-old Col-0 plants grown on soil under a 16-h light ($\sim 100 \mu\text{mol m}^{-2}\cdot\text{s}^{-1}$ from cool white fluorescent bulbs) and 8-h dark cycle at 22°C. Finely cut leaf slices were immersed in enzyme solution (1% [w/v] cellulase, 0.25% [w/v] macerozyme, 0.4 M mannitol, 20 mM KCl, 20 mM MES, and 10 mM CaCl_2) followed by vacuum infiltration for 30 min and gentle shaking at 40 rpm in the dark for 2–2.5 h at room temperature to release the protoplasts. The protoplasts were passed through a 70- μm cell strainer, centrifuged at 100 g at 4°C for 3 min, and washed with cold W5 solution (154 mM NaCl, 125 mM CaCl_2 , 5mM KCl, and 2 mM MES) twice. The protoplasts were counted and resuspended in MMG solution (4.5 mM MES [pH 5.7], 0.4 M mannitol, and 15 mM MgCl_2) at 1×10^5 protoplasts/150 μL . For protoplast transformation, 1×10^5 protoplasts were combined with 5 μg plasmid DNA and 170 μL polyethylene glycol (PEG) solution (40% [w/v] PEG4000, 0.2 M mannitol, and 100 mM CaCl_2), and incubated for 5 min. After four consecutive washes with W5 solution, the transformed protoplasts were incubated in the dark for 16 to 18 h. Typically, 8 replicates of transformation were performed for each plasmid DNA in one experiment, and the

experiments were performed two or three times with similar results. The plasmid DNA was prepared using ZymoPURE II Plasmid Midiprep Kit (Zymo #D4201). After overnight incubation, protoplasts were harvested by centrifugation at 2250 g for 3 min at 4°C. The lysates were generated by adding 100 µL 1X Passive Lysis Buffer included in the Dual-luciferase assay kit (Promega E1960) to the protoplasts and vigorously shaking at room temperature for 15 min. The lysates were cleared by centrifugation at 2250 g for 3 min, and 20 µL of the supernatant was used for the dual-luciferase assay measured in a GloMax Navigator Plate Reader (Promega, GM2010) as specified by the manufacturer. FLUC luminescence was normalized to their corresponding REN luminescence.

For quantifying *FLUC* and *REN* transcript levels, 8 µL 10% (w/v) SDS was added to the remaining 80 µL lysates above, and total RNA was isolated using a Zymo RNA Clean and Concentrator Kit-5 (Zymo, R1014). RNA was converted to first-strand cDNA using LunaScript RT SuperMix (New England Biolabs, 3010) in a final 10-µL reaction. Ten-fold diluted cDNA was used in RT-qPCR using Luna Universal qPCR Master Mix (New England Biolabs #M3003E) in a final 10-µL reaction on a QuantStudio 3 Real-Time PCR machine (Thermo Fisher Scientific). Primers for *FLUC* and *REN* were previously described in (Zhang et al., 2018). The quantification was determined using the standard curve method, and relative *FLUC* transcript levels were normalized to their corresponding *REN* transcripts levels.

Accession numbers

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo>) under accession number GSE183264. R codes for Ribo-seq analysis, RiboTaper pipeline, identifying CiPS uORFs, the overlapping between uORFs, and uORF regulated by alternative splicing are available at GitHub (<https://github.com/hsinyenwu/ORFeome>). Analysis files, proteomic database fasta files, and the RiboPlotR output of the 2113 translated uORFs are available on Mendeley Data (<https://data.mendeley.com/datasets/89j7snbm2r/2>). The original MS proteomics raw data may be downloaded from MassIVE (<http://massive.ucsd.edu>) using the identifier “MSV000085044”; the three key files (i.e., evidence.txt, peptides.txt and proteinGroups.txt) are located in the /search/combined/txt subfolder. Genes mentioned in this article with their AGI codes are listed in Dataset S8.

COMPETING INTEREST STATEMENT

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS AND ACKNOWLEDGEMENTS

HLW and PYH designed the research and interpreted the data. PYH performed the sequencing experiments. HLW analyzed the sequencing data and prepared figures and tables. QA, RTT and PHTN performed dual luciferase assays. QA and PHTN performed RT-qPCR. GS and JME generated the proteomics datasets. CM configured the mass-spectrometry database and performed the search. PYH and JWW supervised the research. HLW and PYH wrote the manuscript with input from all other authors.

We thank Drs. Rodrigo Reis and Yves Poirier at University of Lausanne for sharing their protocol for dual luciferase assays, and we thank Isaiah Kaufman at Michigan State University for his critical review of this manuscript. pGreen II 0800 Luc plasmid is a gift from Dr. Roger Hellens at Queensland University of Technology. 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 (MCB-2051885) to PYH and Iowa State University Plant Science Institute, USDA NIFA Hatch project IOW3808, and National Science Foundation grant (IOS-1759023) to JWW.

Table 1. Important genes possessing translated uORFs.

Pathway	Gene Name
Light Signaling	<i>phyA, phyB, phyC, phyD, phyE, PHOT1, PHOT2, CRY2, PHH1, PIF3, PIF4, NPH3</i>
Circadian clock	<i>CCA1, LHY, GI, TIC, BOA, XCT, RVE1, RVE4, ZTL, PRR1, PRR2, PRR3, PRR7, PRR8, LIP1, AFR2, LNK1, LNK4</i>
Auxin	<i>TOPLESS, TPR1, TPR2, ARF3, ARF4, ARF5, ARF6, ARF7, ARF8, ARF9, ARF10, ARF11, ARF18, IAA8, IAA9, IAA10, IAA18, IAA27, TIR1, ABCB1, ABCB4, ABCB19, ARF1-BP, AFB2, PIN4, COI1</i>

Absciscic acid (ABA)	<i>ABI1, ABI2, SNRK2.1, SNRK2.2, SNRK2.5, SNRK2.6, SNRK2.8, SNRK2.10, GPA1, ABA1, ABA2, ABA4, ABF1, ABF3, ABF4, PYL5, PYR1</i>
Jasmonic acid (JA)	<i>TOPELESS, NINJA, JASSY, COI1, JAZ4, JAZ6, JAZ9, JAM1, JAM2, bHLH13, ASK2</i>
Ethylene	<i>ACO1, ACO2, ACO4, EIN2, EIN3, EIN4, EIN5, EIN6, ERF8, RAP2.4, CTR1, CRF2, CRF3, CRF8, CRF10, CRF11, CRF12, ETR2, ETO1, ETP1, ERF040, ERF104, ERS</i>
Gibberellic acid (GA)	<i>AGL20, RGA1, ATKAO1, ATKAO2, SPY</i>
Salicylic acid (SA)	<i>NPR1, NPR3, NPR4, EDS1, SNC1, FLS2, MPK4, LSD1, MKK5, TGA1, TGA2, TGA3, TGA5, TGA7, WRKY1, NSL2, DIR1, NIMIN-3</i>
Cytokinin	<i>CRE1, WOODEN LEG, AHK3, AHK4, SPY, ARR2, ARR7, AHP5, CRF2, CRF3, CRF8, CRF10, CRF11, CRF12</i>
Brassinosteroids	<i>ASKTHETA, BIN2, BRX, BRL3, BRS1, BUL1, BRI1, BIN1, BON</i>
Development	<i>SCARECROW, BP, ERECTA, ELK2, ELK4 (ERECTA-LIKE), YODA</i>
Others	<i>TOR, LST8-1, YAK1, GCN2, S6K1, S6K2, ATSNK1, ATSNK2, RDR6, APX1, CENH3, ATPT1, PSY, IRE1-1, UPF3, SMG7</i>

FIGURE LEGENDS

Fig. 1. Enhanced Ribo-seq data with improved coverage

- (A) Metagene analysis of 28-nt RPFs mapped to regions at the beginning, middle, or end of annotated ORFs in *Arabidopsis thaliana*. The RPFs are presented with their first nt at the P-site, which is the 13th nt for 28-nt RPFs. The RPFs are colored in red, blue, and teal to indicate they are in the first (expected), second, and third reading frames, respectively. Most footprints mapped to the coding sequence (CDS) in the expected reading frame (92% in frame).
- (B) Genomic features mapped by RPFs. Reads that mapped to nuclear (Nuc), mitochondrial (Mt), and plastid (Pt) genes are shown.
- (C) Length distribution of RPFs. Reads that mapped to nuclear genes are shown.
- (D) Distinct P-sites detected in 1 to 20 million randomly selected RPFs from our current and previous datasets (Hsu et al., 2016).

(E) RNA-seq and Ribo-seq profiles of AT1G04020 from the current study and our previous shoot data in (Hsu et al., 2016) are presented. 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 teal to indicate they are in the first (expected), second, and third reading frames, respectively. Reads outside of the open reading frame (ORF) range are colored in gray. In-frame site coverage indicates percentage of in-frame sites contain mapped Ribo-seq reads. Within the gene models, black boxes represent the annotated main ORFs (mORFs), and gray and white regions indicate 5' untranslated regions (5' UTRs) and 3' UTRs, respectively. The specific isoform being plotted is indicated to the left of the gene model and bolded. Black and gray vertical dashed lines represent the translation start and stop, respectively, for the annotated mORF.

Fig. 2. Translated ORFs identified in this study.

- (A) Number and position of ORFs detected within annotated protein-coding mRNAs and RNAs presumed to be non-coding. Most ORFs were identified by RiboTaper. *, additional upstream ORFs (uORFs) were identified by a separate method, CiPS (see below).
- (B) Categories of translated ORFs identified by RiboTaper. RiboTaper defines non-coding ORFs (ncORFs) as ORFs detected within presumed non-coding RNAs.
- (C) Number of translated ORFs and proteins identified from either newly assembled RNAs or various annotated transcript types. The 4th column indicates the number of proteins, peptides, and spectra detected by mass spectrometry for each class of ORFs. Tx, transcripts; MS, mass spectrometry.
- (D) Subcellular localization of proteins encoded by ncORFs of length between 20 and 100 amino acids predicted by TargetP. C, chloroplast; M, mitochondria; S, secreted.
- (E) Size distribution of annotated ORFs (CCDSs) and other ORFs identified by RiboTaper.
- (F–J) Examples of translated ORFs in various transcript types. Ribo-seq and RNA-seq profiles are presented as described in Fig. 1E. The additional ORF in (G) or uORF in (I) are shown by a yellow box in the gene model, and their translation start and stop are indicated by light blue and orange vertical dashed lines, respectively, within the profiles.
- (F) A translated ORF in a novel transcribed region defined by Araport 11.
- (G–H) Translated ORFs in primary transcripts of miR163 and miR169. The first ORF within *pre-MIR163* was visually identified.
- (I) A translated uORF in *ATKAO1*, involved in GA biosynthesis.

(J) A translated ORF in QQS, a sORF identified from an orphan gene in a previous study (Li et al., 2009).

Fig. 3. Translation in plastids and mitochondria.

(A) P-site inferred in RPFs mapped to nuclear, plastid, and mitochondrial genes.

(B–C) Metagene analysis of 28-nt RPFs at the beginning, middle, or end of annotated plastid and mitochondrial ORFs. The RPFs are colored in red, blue, and teal to indicate they are in the first, second, or third reading frames, respectively.

(D–E) Examples of translational profiles of plastid genes, which use frame 1 (red). Note that ATCG00130 (E) contains an intron.

(F) An example of translational profiles of a mitochondrial gene, which uses frame 3 (teal). Note that it also contains an intron.

(G–H) Two annotated mitochondrial genes predicted to function in chloroplasts and have RPFs mapped to frame 1 (red), suggesting that these RPFs are from their plastid and nuclear homologs, respectively.

In (E, F, H), the number above the blue curved line indicates the RNA-seq read count across exon-exon junctions.

Fig. 4. Translation and ribosome stalling of primary transcripts for *TAS1* and *TAS2*.

(A–D) Expression profiles of *TAS1A–C* and *TAS2*. In the gene models, ORFs identified by RiboTaper are marked with ‘R.’ Above the gene models, miRNA target sites are indicated by magenta rectangles. Note that strong RPF peaks were observed upstream of the miRNA target sites (A–D). The coordinates of the strong RPF peaks and the miRNA target sites are shown. The bold numbers in parentheses indicate the distance (nt) between the strong RPF peak and the downstream miRNA target site. The ORF3 in (C), and the ORF1 in (D) were manually curated based on their strong in-frame RPF peaks. These two ORFs were likely excluded by RiboTaper due to another ORF overlapping with them using a different reading frame (blue). Note that ORF1 in (D) was experimentally validated by a previous study (Yoshikawa et al., 2016). *TAS1A* contains an unannotated intron, which is indicated by a blue curved line, and the number above the blue curved line indicates the RNA-seq read count across the exon-exon junction.

Fig. 5. Translation of primary transcripts of *TAS3* and *TAS4* and proposed models.

(A–B) Expression profiles of *TAS3A* and *TAS4*. ORFs identified by RiboTaper are labeled with ‘(R)’ next to the gene models. Above the gene models, miRNA target sites are indicated by magenta rectangles. In (A), a strong RPF peak was observed upstream of the first miRNA390

target site in *TAS3A*. The coordinates of the strong RPF peaks and the miRNA target sites are indicated. In (B), although RiboTaper identified an ORF within the only annotated transcript, our data revealed that *TAS4* has an additional isoform (isoform 2), and the ORF within this unannotated isoform is more likely translated.

(C) Illustration of translated ORFs and RPF stalling relative to the miRNA target sites in *TAS1/2*, *TAS3*, and *TAS4*. Light blue boxes indicate translated ORFs. In *TAS1/2*, multiple ORFs are translated. One particular ORF overlaps with the miR173 target site; translation continues until the ribosome encounters the miR173 target site, where the first nt within the ribosomal P-site (red vertical line) corresponds to 16 nt upstream of the miR173 target site. In *TAS3*, only one ORF is translated, and the ORF is entirely upstream of the first miR390 target site. Similarly, translation continues until the ribosome encounters the first miR390 target site, where the ribosomal P-site corresponds to 16 nt upstream of the miR390 target site. In *TAS4*, only one ORF is translated, and the ORF overlaps with the miR828 target site.

Fig. 6. Examples of downstream ORFs (dORFs) and their translation efficiency.

(A) Comparison of mORF translation efficiency between genes with and without dORFs. The statistical significance of the difference between the two distributions was determined by a Kolmogorov-Smirnov (KS) test.

(B–F) Examples of dORFs. In the gene models, the dORFs are indicated by orange box(es), and their translation start and stop are indicated by light blue and orange vertical dashed lines, respectively, within the profiles.

(B–D) Type 1 dORFs: the mORF have high translation levels, while the dORFs have low translation levels. The RPFs in the mORFs are shown in gray. The RPFs in the dORFs are magnified and presented in three colors to indicate their reading frame.

(E–F) Type 2 dORFs: an additional gene/transcript is present in the annotated 3' UTR or downstream of the mORF gene. These additional genes/transcripts have distinct mRNA levels compared to the mORF genes. The zoom-in in (F) shows that the dORF start overlaps with the mORF stop, supporting the idea that these two ORFs are unlikely to be translated sequentially on the same mRNA.

(G) Illustrations of Type 1 and Type 2 dORFs. Type 1 dORFs may be potential readthrough from the mORF, or continuous translation from an alternative ORF (blue dashed box) overlapping with the mORF. Type 2 dORFs result from mis-annotation, in which a hidden gene/transcript is located in the 3' UTR or downstream of the mORF gene. These hidden genes

have independent transcription start sites supported by published CAGE data (see **Fig. S11**), and they have distinct RNA-seq (and Ribo-seq) levels, compared to the upstream mORF.

Fig. 7. Periodic ribosome stalling in CPuORFs. (A-D) Examples of uORFs encoding conserved peptides (CPuORFs) with periodic ribosome stalling. In the gene model, the CPuORF is shown by a yellow box, and its translation start and stop indicated by light blue and orange vertical dashed lines, respectively, within the profiles. The 30-nt internals between RPF peaks are indicated, and the inferred ribosome positions are illustrated.

Fig. 8. Newly discovered uORFs by RiboTaper.

(A–B) Two examples of translated uORFs identified by RiboTaper in the current data compared to our previous data (Hsu et al., 2016). Note that the in-frame site coverage by Ribo-seq reads is much improved in the current data.

(C) Cumulative plot comparing the mORF translation efficiency (TE) of genes containing 0, 1, 2, or ≥ 3 translated uORFs identified by RiboTaper. The different superscript lowercase letters indicate significant differences between groups (KS test, $P < 0.05$).

(D) Diagrams of dual-luciferase constructs for testing uORF functions. The start codon of uORF (ATG) was mutated to AGG in the mutated version. 35Sp, 35S promoter; CaMV term, CaMV terminator.

(E) Relative FLUC luminescence comparing 5' UTRs carrying the wild-type uORF or mutated uORF. FLUC luminescence levels were normalized to REN luminescence levels.

(F) Relative *FLUC* mRNA levels comparing 5' UTRs carrying the wild-type uORF or mutated uORF. *FLUC* mRNA levels were normalized to *REN* mRNA levels.

The statistical significance for boxplots in (E–F) was determined by Wilcoxon rank sum test (*: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$). In the boxplots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

Fig. 9. Identification of uORFs using CiPS.

(A) Cumulative plot comparing the in-frame percentage of CCDs, ncORFs, dORFs, and uORFs identified by RiboTaper. Different superscript lowercase letters indicate significant differences between groups (KS test, $P < 0.05$).

(B–E) Illustrations of expected RPF distribution of a long uORF (B), a minimum uORF (C), a 2-aa uORF (D), and a 3-aa uORF (E). Note that the codon preceding the stop (–1) is expected to have a significant amount of RPFs mapped to frame 2 (blue). For minimum uORFs (C), the start codon is also the –1 codon.

(F) Distribution of uORF peptide length of predicted uORFs (AUG-start), RiboTaper-identified uORFs, and CiPS-identified uORFs.

(G) Venn diagram showing the extent of overlap between uORFs identified by CiPS and/or RiboTaper.

(H) Distribution of uORF peptide length, in terms of number, comparing uORFs identified by RiboTaper-only, CiPS-only, or both methods. The gray dashed line marks 20 aa.

(I) Distribution of uORF peptide length, in terms of percentage, comparing uORFs identified by RiboTaper-only, CiPS-only, or both methods. The gray dashed line marks 20 aa.

(J) Cumulative plot comparing mORF translation efficiency of genes without any translated uORFs and with uORFs identified by RiboTaper-only, CiPS-only, or both methods. Different superscript lowercase letters indicate significant differences between groups (KS test, $P < 0.05$).

(K) Comparison of the number of translated uORFs identified per gene between RiboTaper and CiPS.

(L) Expression profile of *ARF6*, which contains multiple uORFs. Zoom-ins of each AUG-start uORF (yellow box) and an AUA-start uORF (orange box) are shown below the main diagram to evaluate the performance of RiboTaper and CiPS.

(M) Cumulative plot comparing the mORF translation efficiency of genes containing 0, 1, 2, ≥ 3 translated uORFs identified by CiPS. Different superscript lowercase letters indicate significant differences between groups (KS test, $P < 0.05$).

(N) Boxplot comparing the mORF translation efficiency of genes containing no, 1 aa (minimum), 2–10 aa, 11–20 aa, 21–30 aa and > 30 aa translated uORFs identified by CiPS. The gray horizontal line in the background marks the median of genes without translated uORFs. Different superscript lowercase letters indicate significant differences between groups (Wilcoxon rank sum test, $P < 0.05$). In the boxplots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range.

Fig. 10. Translation of minimum uORFs.

(A–F) Examples of minimum uORFs. The minimum uORF positions are indicated by orange triangles. RPFs mapped to minimum uORFs show the characteristic of –1 codon, where a high fraction of RPFs mapped to frame 2 (blue).

(G) Diagrams of dual-luciferase constructs for testing minimum uORF functions. The start codon of minimum uORF (ATG) was mutated to ATC in the mutated version.

(H) Relative FLUC luminescence comparing 5' UTRs carrying the wild-type uORF or mutated uORF. FLUC luminescence levels are normalized to REN luminescence levels.

(I) Relative *FLUC* mRNA levels comparing 5' UTRs carrying the wild-type uORF or mutated uORF. *FLUC* mRNA levels are normalized to *REN* mRNA levels.

The statistical significance for boxplots in (H–I) was determined by Wilcoxon rank sum test (*: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$). In the boxplots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

Fig. 11. Translation of tiny uORFs.

(A–D) Examples of tiny uORFs of varying uORF length. Ribo-seq read and frame information for the tiny uORFs are shown in the zoom-in boxes.

(E) Relative FLUC luminescence comparing 5' UTRs carrying the wild-type uORF (ATG) or mutated uORF (AGG). FLUC luminescence levels were normalized to REN luminescence levels. Statistical significance was determined by Wilcoxon rank sum test (*: $0.01 < P < 0.05$, **: $0.001 < P < 0.01$, ***: $0.0001 < P < 0.001$). In the boxplots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

Fig. 12. Overlapping of translated uORFs.

(A–C) Translational profiles of three CPuORF genes showing the overlapping of translated uORFs. The positions of CPuORFs (red boxes) and other uORFs (blue or teal boxes, depending on their reading frame) are indicated. In (B–C), zoom-ins of the overlapping region between the CPuORF and the stacking uORF (suORF). Note that *bZIP11* (A) is also shown in Fig. 7B as an example of periodic ribosome stalling upstream of the CPuORF stop codon.

Fig. 13. Translated uORFs regulated by alternative splicing.

(A–D) Examples of translated uORFs affected by various types of alternative splicing. The number above the blue curved line indicates the RNA-seq read count across exon-exon junctions. The specific isoform number being plotted is indicated to the left of the gene model and bolded. The yellow boxes within the gene models represent the uORFs.

References

Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L., and Pestova, T.V. (2006). In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors eRF1 and eRF3. *Cell* 125: 1125–1136.

- 1212 Andrews, S.J. and Rothnagel, J.A. (2014). Emerging evidence for functional peptides encoded
1213 by short open reading frames. *Nature Reviews Genetics* 15: 193–204.
- 1214 Armenteros, J.J.A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A.,
1215 and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep
1216 learning. *Life Science Alliance* 2: e201900429.
- 1217 de Bang, T.C. et al. (2017). Genome-Wide Identification of Medicago Peptides Involved in
1218 Macronutrient Responses and Nodulation. *Plant Physiol* 175: 1669–1689.
- 1219 Basrai, M.A., Hieter, P., and Boeke, J.D. (1997). Small open reading frames: Beautiful needles
1220 in the haystack. *Genome Research* 7: 768–771.
- 1221 Bazin, J., Baerenfaller, K., Gosai, S.J., Gregory, B.D., Crespi, M., and Bailey-Serres, J. (2017).
1222 Global analysis of ribosome-associated noncoding RNAs unveils new modes of
1223 translational regulation. *Proceedings of the National Academy of Sciences of the United*
1224 *States of America* 114: E10018–E10027.
- 1225 Bazzini, A.A., Johnstone, T.G., Christiano, R., MacKowiak, S.D., Obermayer, B., Fleming, E.S.,
1226 Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., and Giraldez, A.J. (2014).
1227 Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary
1228 conservation. *EMBO Journal* 33: 981–993.
- 1229 Brar, G.A. and Weissman, J.S. (2015). Ribosome profiling reveals the what, when, where and
1230 how of protein synthesis. *Nature Reviews Molecular Cell Biology* 16: 651–664.
- 1231 Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-
1232 seq quantification. *Nature Biotechnology* 34: 525–527.
- 1233 Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for
1234 stop codon recognition in eukaryotes. *Nature* 524: 493–496.
- 1235 Buskirk, A.R. and Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and
1236 eukaryotes. *Philos Trans R Soc Lond B Biol Sci* 372: 20160183.
- 1237 Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M.,
1238 Obermayer, B., and Ohler, U. (2016). Detecting actively translated open reading frames
1239 in ribosome profiling data. *Nature Methods* 13: 165–170.
- 1240 Calviello, L., Sydow, D., Harnett, D., and Ohler, U. (2019). Ribo-seQC: comprehensive analysis
1241 of cytoplasmic and organellar ribosome profiling data. *bioRxiv*: 601468.
- 1242 Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D.
1243 (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference
1244 genome. *Plant Journal* 89: 789–804.
- 1245 Chotewutmontri, P. and Barkan, A. (2016). Dynamics of Chloroplast Translation during
1246 Chloroplast Differentiation in Maize. *PLoS Genetics* 12: e1006106.

- 1247 Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C., and
1248 Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-
1249 friendly software package for Ribo-seq data analysis. *Rna* 21: 1731–1745.
- 1250 Clark, N.M., Nolan, T.M., Wang, P., Song, G., Montes, C., Valentine, C.T., Guo, H., Sozzani, R.,
1251 Yin, Y., and Walley, J.W. (2021). Integrated omics networks reveal the temporal
1252 signaling events of brassinosteroid response in *Arabidopsis*. *Nat Commun* 12: 5858.
- 1253 Constabel, C.P., Yip, L., and Ryan, C.A. (1998). Prosystemin from potato, black nightshade,
1254 and bell pepper: Primary structure and biological activity of predicted systemin
1255 polypeptides. *Plant Molecular Biology* 36: 55–62.
- 1256 Da Rosa, E.L.S., Oleskovicz, C.F., and Aragão, B.N. (2004). Rapid prototyping in maxillofacial
1257 surgery and traumatology: Case report. *Brazilian Dental Journal* 15: 243–247.
- 1258 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,
1259 and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:
1260 15–21.
- 1261 Dong, J., Chen, H., Deng, X.W., Irish, V.F., and Wei, N. (2020). Phytochrome B induces intron
1262 retention and translational inhibition of phytochrome-interacting factor3. *Plant Physiology*
1263 182: 159–166.
- 1264 Fei, Q., Xia, R., and Meyers, B.C. (2013). Phased, secondary, small interfering RNAs in
1265 posttranscriptional regulatory networks. *Plant Cell* 25: 2400–2415.
- 1266 Feng, Y., Jiang, M., Yu, W., and Zhou, J. (2023). Identification of short open reading frames in
1267 plant genomes. *Frontiers in Plant Science* 14.
- 1268 Fesenko, I. et al. (2019). Distinct types of short open reading frames are translated in plant cells.
1269 *Genome Res* 29: 1464–1477.
- 1270 Fields, A.P., Rodriguez, E.H., Jovanovic, M., Stern-Ginossar, N., Haas, B.J., Mertins, P.,
1271 Raychowdhury, R., Hacohen, N., Carr, S.A., Ingolia, N.T., Regev, A., and Weissman,
1272 J.S. (2015). A Regression-Based Analysis of Ribosome-Profiling Data Reveals a
1273 Conserved Complexity to Mammalian Translation. *Molecular Cell* 60: 816–827.
- 1274 Gage, J.L., Mali, S., McLoughlin, F., Khaipho-Burch, M., Monier, B., Bailey-Serres, J., Vierstra,
1275 R.D., and Buckler, E.S. (2022). Variation in upstream open reading frames contributes to
1276 allelic diversity in maize protein abundance. *Proceedings of the National Academy of*
1277 *Sciences of the United States of America* 119: e2112516119.
- 1278 Guo, R., Yu, X., and Gregory, B.D. (2023a). The identification of conserved sequence features
1279 of co-translationally decayed mRNAs and upstream open reading frames in angiosperm
1280 transcriptomes. *Plant Direct* 7: e479.
- 1281 Guo, Y. et al. (2023b). The translational landscape of bread wheat during grain development.
1282 *The Plant Cell: Online ahead of print.*

- 1283 Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome
1284 profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154:
1285 240–251.
- 1286 Guydosh, N.R. and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell*
1287 156: 950–962.
- 1288 Hanada, K. et al. (2013). Small open reading frames associated with morphogenesis are hidden
1289 in plant genomes. *Proceedings of the National Academy of Sciences of the United*
1290 *States of America* 110: 2395–2400.
- 1291 Hanada, K., Akiyama, K., Sakurai, T., Toyoda, T., Shinozaki, K., and Shiu, S.H. (2009). sORF
1292 finder: A program package to identify small open reading frames with high coding
1293 potential. *Bioinformatics* 26: 399–400.
- 1294 Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H., and Shiu, S.H. (2007). A large number of novel
1295 coding small open reading frames in the intergenic regions of the *Arabidopsis thaliana*
1296 genome are transcribed and/or under purifying selection. *Genome Research* 17: 632–
1297 640.
- 1298 Hanfrey, C., Elliott, K.A., Franceschetti, M., Mayer, M.J., Illingworth, C., and Michael, A.J. (2005).
1299 A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-
1300 responsive translation. *Journal of Biological Chemistry* 280: 39229–39237.
- 1301 Hanfrey, C., Franceschetti, M., Mayer, M.J., Illingworth, C., and Michael, A.J. (2002). Abrogation
1302 of upstream open reading frame-mediated translational control of a plant S-
1303 adenosylmethionine decarboxylase results in polyamine disruption and growth
1304 perturbations. *Journal of Biological Chemistry* 277: 44131–44139.
- 1305 Hayden, C.A. and Jorgensen, R.A. (2007). Identification of novel conserved peptide uORF
1306 homology groups in *Arabidopsis* and rice reveals ancient eukaryotic origin of select
1307 groups and preferential association with transcription factor-encoding genes. *BMC*
1308 *Biology* 5: 32.
- 1309 Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam,
1310 S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional
1311 genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods*
1312 1: 13.
- 1313 Hellens, R.P., Brown, C.M., Chisnall, M.A.W., Waterhouse, P.M., and Macknight, R.C. (2016).
1314 The Emerging World of Small ORFs. *Trends in Plant Science* 21: 317–328.
- 1315 Hou, C.Y., Lee, W.C., Chou, H.C., Chen, A.P., Chou, S.J., and Chen, H.M. (2016). Global
1316 analysis of truncated RNA ends reveals new insights into Ribosome Stalling in plants.
1317 *Plant Cell* 28: 2398–2416.
- 1318 Hsu, P.Y. and Benfey, P.N. (2018). Small but Mighty: Functional Peptides Encoded by Small
1319 ORFs in Plants. *Proteomics* 18: 1700038.
- 1320 Hsu, P.Y., Calviello, L., Wu, H.Y.L., Li, F.W., Rothfels, C.J., Ohler, U., and Benfey, P.N. (2016).
1321 Super-resolution ribosome profiling reveals unannotated translation events in

- 1322 Arabidopsis. Proceedings of the National Academy of Sciences of the United States of
1323 America 113: E7126–E7135.
- 1324 Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Collier, J. (2009). Co-translational
1325 mRNA decay in *Saccharomyces cerevisiae*. *Nature* 461: 225–229.
- 1326 Ingolia, N.T., Ghaemmamghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide
1327 analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*
1328 324: 218–223.
- 1329 Iwakawa, H. oki, Lam, A.Y.W., Mine, A., Fujita, T., Kiyokawa, K., Yoshikawa, M., Takeda, A.,
1330 Iwasaki, S., and Tomari, Y. (2021). Ribosome stalling caused by the Argonaute-
1331 microRNA-SGS3 complex regulates the production of secondary siRNAs in plants. *Cell*
1332 Reports 35: 109300.
- 1333 Jiang, M., Ning, W., Wu, S., Wang, X., Zhu, K., Li, A., Li, Y., Cheng, S., and Song, B. (2022).
1334 Three-nucleotide periodicity of nucleotide diversity in a population enables the
1335 identification of open reading frames. *Briefings in Bioinformatics* 23: 1–13.
- 1336 Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J. (2016). Upstream ORF s are prevalent
1337 translational repressors in vertebrates. *The EMBO Journal* 35: 706–723.
- 1338 Jorgensen, R.A. and Dorantes-Acosta, A.E. (2012). Conserved peptide upstream open reading
1339 frames are associated with regulatory genes in angiosperms. *Frontiers in Plant Science*
1340 3: 191.
- 1341 Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J. (2014). Translational dynamics
1342 revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. *Proceedings of*
1343 *the National Academy of Sciences of the United States of America* 111: E203–E212.
- 1344 Kim, B.H., Cai, X., Vaughn, J.N., and Von Arnim, A.G. (2007). On the functions of the h subunit
1345 of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome Biology* 8:
1346 R60.
- 1347 Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the aug initiator codon
1348 by eukaryotic ribosomes. *Nucleic Acids Research* 9: 5233–5252.
- 1349 Kurihara, Y., Makita, Y., Shimohira, H., Fujita, T., Iwasaki, S., and Matsui, M. (2020).
1350 Translational landscape of protein-coding and non-protein-coding RNAs upon light
1351 exposure in *Arabidopsis*. *Plant and Cell Physiology* 61: 536–545.
- 1352 Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
1353 Methods 9: 357–359.
- 1354 Lauressergues, D., Couzigou, J.M., San Clemente, H., Martinez, Y., Dunand, C., Bécard, G.,
1355 and Combier, J.P. (2015). Primary transcripts of microRNAs encode regulatory peptides.
1356 *Nature* 520: 90–93.
- 1357 Lease, K.A. and Walker, J.C. (2006). The *Arabidopsis* unannotated secreted peptide database,
1358 a resource for plant peptidomics. *Plant Physiology* 142: 831–838.

- 1359 Lei, L. et al. (2015). Ribosome profiling reveals dynamic translational landscape in maize
1360 seedlings under drought stress. *Plant Journal* 84: 1206–1208.
- 1361 Li, B. and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data
1362 with or without a reference genome. *BMC Bioinformatics* 12: 323.
- 1363 Li, L. et al. (2015). QQS orphan gene regulates carbon and nitrogen partitioning across species
1364 via NF-YC interactions. *Proceedings of the National Academy of Sciences of the United*
1365 *States of America* 112: 14734–14739.
- 1366 Li, L., Foster, C.M., Gan, Q., Nettleton, D., James, M.G., Myers, A.M., and Wurtele, E.S. (2009).
1367 Identification of the novel protein QQS as a component of the starch metabolic network
1368 in *Arabidopsis* leaves. *Plant J* 58: 485–498.
- 1369 Li, S. et al. (2016). Biogenesis of phased siRNAs on membrane-bound polysomes in
1370 *Arabidopsis*. *eLife* 5: e22750.
- 1371 Li, Y.R. and Liu, M.J. (2020). Prevalence of alternative AUG and non-AUG translation initiators
1372 and their regulatory effects across plants. *Genome Research* 30: 1418–1433.
- 1373 Li, Z., Jin, J., Wang, Y., Long, W., Ding, Y., Hu, H., and Wei, L. (2023). ExamPLe: explainable
1374 deep learning framework for the prediction of plant small secreted peptides.
1375 *Bioinformatics* 39: btad108.
- 1376 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).
1377 Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell* 25: 3699–3710.
- 1378 Matheisl, S., Berninghausen, O., Becker, T., and Beckmann, R. (2015). Structure of a human
1379 translation termination complex. *Nucleic Acids Res* 43: 8615–8626.
- 1380 Merret, R., Nagarajan, V.K., Carpentier, M.C., Park, S., Favory, J.J., Descombin, J., Picart, C.,
1381 Charnig, Y.Y., Green, P.J., Deragon, J.M., and Bousquet-Antonelli, C. (2015). Heat-
1382 induced ribosome pausing triggers mRNA co-translational decay in *Arabidopsis thaliana*.
1383 *Nucleic Acids Research* 43: 4121–4132.
- 1384 Montes, C., Wang, P., Liao, C.Y., Nolan, T.M., Song, G., Clark, N.M., Elmore, J.M., Guo, H.,
1385 Bassham, D.C., Yin, Y., and Walley, J.W. (2022). Integration of multi-omics data reveals
1386 interplay between brassinosteroid and Target of Rapamycin Complex signaling in
1387 *Arabidopsis*. *New Phytologist* 236: 893–910.
- 1388 Olsen, A.N., Mundy, J., and Skriver, K. (2002). Peptomics, identification of novel cationic
1389 *Arabidopsis* peptides with conserved sequence motifs. *In Silico Biology* 2: 441–451.
- 1390 Ong, S.N., Tan, B.C., Al-Idrus, A., and Teo, C.H. (2022). Small open reading frames in plant
1391 research: from prediction to functional characterization. *3 Biotech* 12: 76.
- 1392 Orr, M.W., Mao, Y., Storz, G., and Qian, S.-B. (2020). Alternative ORFs and small ORFs:
1393 shedding light on the dark proteome. *Nucleic Acids Res* 48: 1029–1042.
- 1394 Pearce, G., Strydom, D., Johnson, S., and Ryan, C.A. (1991). A polypeptide from tomato leaves
1395 induces wound-inducible proteinase inhibitor proteins. *Science* 253: 895–898.

- 1396 Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread co-translational RNA decay
1397 reveals ribosome dynamics. *Cell* 161: 1400–1412.
- 1398 Perte, G. and Perte, M. (2020). GFF Utilities: GffRead and GffCompare.
- 1399 Perte, M., Perte, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L.
1400 (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq
1401 reads. *Nature Biotechnology* 33: 290–295.
- 1402 Planchard, N., Bertin, P., Quadrado, M., Dargel-Graffin, C., Hatin, I., Namy, O., and Mireau, H.
1403 (2018). The translational landscape of Arabidopsis mitochondria. *Nucleic Acids*
1404 *Research* 46: 6218–6228.
- 1405 Puyaubert, J., Denis, L., and Alban, C. (2008). Dual targeting of Arabidopsis holocarboxylase
1406 synthetase1: A small upstream open reading frame regulates translation initiation and
1407 protein targeting. *Plant Physiology* 146: 478–491.
- 1408 Qanmber, G., You, Q., Yang, Z., Fan, L., Zhang, Z., Chai, M., Gao, B., Li, F., and Yang, Z.
1409 (2023). Transcriptional and translational landscape fine-tune genome annotation and
1410 explores translation control in cotton. *J Adv Res: S2090-1232(23)00142-X*.
- 1411 Qi, M. et al. (2019). QQS orphan gene and its interactor NF-YC4 reduce susceptibility to
1412 pathogens and pests. *Plant Biotechnol J* 17: 252–263.
- 1413 Rahmani, F., Hummel, M., Schuurmans, J., Wiese-Klinkenberg, A., Smeekens, S., and Hanson,
1414 J. (2009). Sucrose control of translation mediated by an upstream open reading frame-
1415 encoded peptide. *Plant Physiology* 150: 1356–1367.
- 1416 Raina, M., King, A., Bianco, C., and Vanderpool, C.K. (2018). Dual-Function RNAs. *Microbiol*
1417 *Spectr* 6.
- 1418 Reis, R.S., Deforges, J., Sokoloff, T., and Poirier, Y. (2020). Modulation of shoot phosphate
1419 level and growth by PHOSPHATE1 upstream open reading frame. *Plant Physiology* 183:
1420 1145–1156.
- 1421 Rook, F., Gerrits, N., Kortstee, A., Van Kampen, M., Borrias, M., Weisbeek, P., and Smeekens,
1422 S. (1998). Sucrose-specific signalling represses translation of the Arabidopsis ATB2
1423 bZIP transcription factor gene. *Plant Journal* 15: 253–263.
- 1424 Roy, B., Vaughn, J.N., Kim, B.H., Zhou, F., Gilchrist, M.A., and Von Arnim, A.G. (2010). The h
1425 subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring
1426 upstream open reading frames. *Rna* 16: 748–761.
- 1427 Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M. (2014). Long non-coding RNAs
1428 as a source of new peptides. *eLife* 3: e03523.
- 1429 Sahoo, S., Singh, D., Singh, A., Pandit, M., Vasu, K., Som, S., Pullagurla, N.J., Laha, D., and
1430 Eswarappa, S.M. (2022). Identification and functional characterization of mRNAs that
1431 exhibit stop codon readthrough in Arabidopsis thaliana. *Journal of Biological Chemistry*
1432 298: 102173.

- 1433 Sandmann, C.-L. et al. (2023). Evolutionary origins and interactomes of human, young
1434 microproteins and small peptides translated from short open reading frames. *Mol Cell* 83:
1435 994-1011.e18.
- 1436 Schuller, A.P. and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation.
1437 *Nature Reviews Molecular Cell Biology* 19: 526–541.
- 1438 Sharma, A., Badola, P.K., Bhatia, C., Sharma, D., and Trivedi, P.K. (2020). Primary transcript of
1439 miR858 encodes regulatory peptide and controls flavonoid biosynthesis and
1440 development in *Arabidopsis*. *Nature Plants* 6: 1262–1274.
- 1441 Song, G., Brachova, L., Nikolau, B.J., Jones, A.M., and Walley, J.W. (2018a). Heterotrimeric G-
1442 Protein-Dependent Proteome and Phosphoproteome in Unstimulated *Arabidopsis* Roots.
1443 *Proteomics* 18: 1800323.
- 1444 Song, G., Hsu, P.Y., and Walley, J.W. (2018b). Assessment and Refinement of Sample
1445 Preparation Methods for Deep and Quantitative Plant Proteome Profiling. *Proteomics* 18:
1446 e1800220.
- 1447 Song, G., Montes, C., and Walley, J.W. (2020). Quantitative Profiling of Protein Abundance and
1448 Phosphorylation State in Plant Tissues Using Tandem Mass Tags. In *Plant Proteomics:
1449 Methods and Protocols*, J.V. Jorin-Novo, L. Valledor, M.A. Castillejo, and M.-D. Rey,
1450 eds, *Methods in Molecular Biology*. (Springer US: New York, NY), pp. 147–156.
- 1451 Sotta, N., Chiba, Y., Aoyama, H., Takamatsu, S., Suzuki, T., Miwa, K., Yamashita, Y., Naito, S.,
1452 and Fujiwara, T. (2022). Translational Landscape of a C4 Plant, *Sorghum bicolor*, Under
1453 Normal and Sulfur-Deficient Conditions. *Plant and Cell Physiology* 63: 592–604.
- 1454 Sotta, N., Chiba, Y., Miwa, K., Takamatsu, S., Tanaka, M., Yamashita, Y., Naito, S., and
1455 Fujiwara, T. (2021). Global analysis of boron-induced ribosome stalling reveals its
1456 effects on translation termination and unique regulation by AUG-stops in *Arabidopsis*
1457 shoots. *Plant Journal* 106: 1455–1467.
- 1458 Takahashi, F., Hanada, K., Kondo, T., and Shinozaki, K. (2019). Hormone-like peptides and
1459 small coding genes in plant stress signaling and development. *Current Opinion in Plant
1460 Biology* 51: 88–95.
- 1461 Takahashi, H. et al. (2020). Comprehensive genome-wide identification of angiosperm upstream
1462 ORFs with peptide sequences conserved in various taxonomic ranges using a novel
1463 pipeline, ESUCA. *BMC Genomics* 21: 260.
- 1464 Takahashi, H., Takahashi, A., Naito, S., and Onouchi, H. (2012). BAIUCAS: A novel BLAST-
1465 based algorithm for the identification of upstream open reading frames with conserved
1466 amino acid sequences and its application to the *Arabidopsis thaliana* genome.
1467 *Bioinformatics* 28: 2231–2241.
- 1468 Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., Chiba, Y., Hirai, M.Y.,
1469 Akiyama, T., Onouchi, H., Naito, S., and Fujiwara, T. (2016). The Minimum open reading
1470 frame, AUG-stop, induces boron-dependent ribosome stalling and mRNA degradation.
1471 *Plant Cell* 28: 2830–2849.

- 1472 Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammuea, B.P.A. (2015). The
1473 plant peptidome: An expanding repertoire of structural features and biological functions.
1474 *Plant Cell* 27: 2095–2118.
- 1475 Thieffry, A., Vigh, M.L., Bornholdt, J., Ivanov, M., Brodersen, P., and Sandelin, A. (2020).
1476 Characterization of arabidopsis thaliana promoter bidirectionality and antisense RNAs by
1477 inactivation of nuclear RNA decay pathways. *Plant Cell* 32: 1845–1867.
- 1478 Trösch, R. (2022). Chloroplast transcription and translation are linked. *Nat. Plants* 8: 867–867.
- 1479 Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass
1480 spectrometry-based shotgun proteomics. *Nature Protocols* 11: 2301–2319.
- 1481 Ulveling, D., Francastel, C., and Hubé, F. (2011). When one is better than two: RNA with dual
1482 functions. *Biochimie* 93: 633–644.
- 1483 Van Der Horst, S., Filipovska, T., Hanson, J., and Smeekens, S. (2020). Metabolite control of
1484 translation by conserved peptide uORFs: The ribosome as a metabolite multisensor.
1485 *Plant Physiology* 182: 110–122.
- 1486 Van Der Horst, S., Snel, B., Hanson, J., and Smeekens, S. (2019). Novel pipeline identifies new
1487 upstream ORFs and non-AUG initiating main ORFs with conserved amino acid
1488 sequences in the 5' leader of mRNAs in *Arabidopsis thaliana*. *RNA* 25: 292–304.
- 1489 Vaughn, J.N., Ellingson, S.R., Mignone, F., and Von Arnim, A. (2012). Known and novel post-
1490 transcriptional regulatory sequences are conserved across plant families. *Rna* 18: 368–
1491 384.
- 1492 Von Arnim, A.G., Jia, Q., and Vaughn, J.N. (2014). Regulation of plant translation by upstream
1493 open reading frames. *Plant Science* 214: 1–12.
- 1494 Wang, H., Wang, Y., and Xie, Z. (2019). Computational resources for ribosome profiling: from
1495 database to Web server and software. *Brief Bioinform* 20: 144–155.
- 1496 Wang, Y., Zhang, H., and Lu, J. (2020). Recent advances in ribosome profiling for deciphering
1497 translational regulation. *Methods* 176: 46–54.
- 1498 Wiese, A., Elzinga, N., Wobbes, B., and Smeekens, S. (2004). A conserved upstream open
1499 reading frame mediates sucrose-induced repression of translation. *Plant Cell* 16: 1717–
1500 1729.
- 1501 Wolin, S.L. and Walter, P. (1988). Ribosome pausing and stacking during translation of a
1502 eukaryotic mRNA. *The EMBO journal* 7: 3559–3569.
- 1503 Wu, H.Y.L. and Hsu, P.Y. (2022). A custom library construction method for super-resolution
1504 ribosome profiling in *Arabidopsis*. *Plant Methods* 18: 1–8.
- 1505 Wu, H.Y.L. and Hsu, P.Y. (2021). RiboPlotR: a visualization tool for periodic Ribo-seq reads.
1506 *Plant Methods* 17: 1–7.

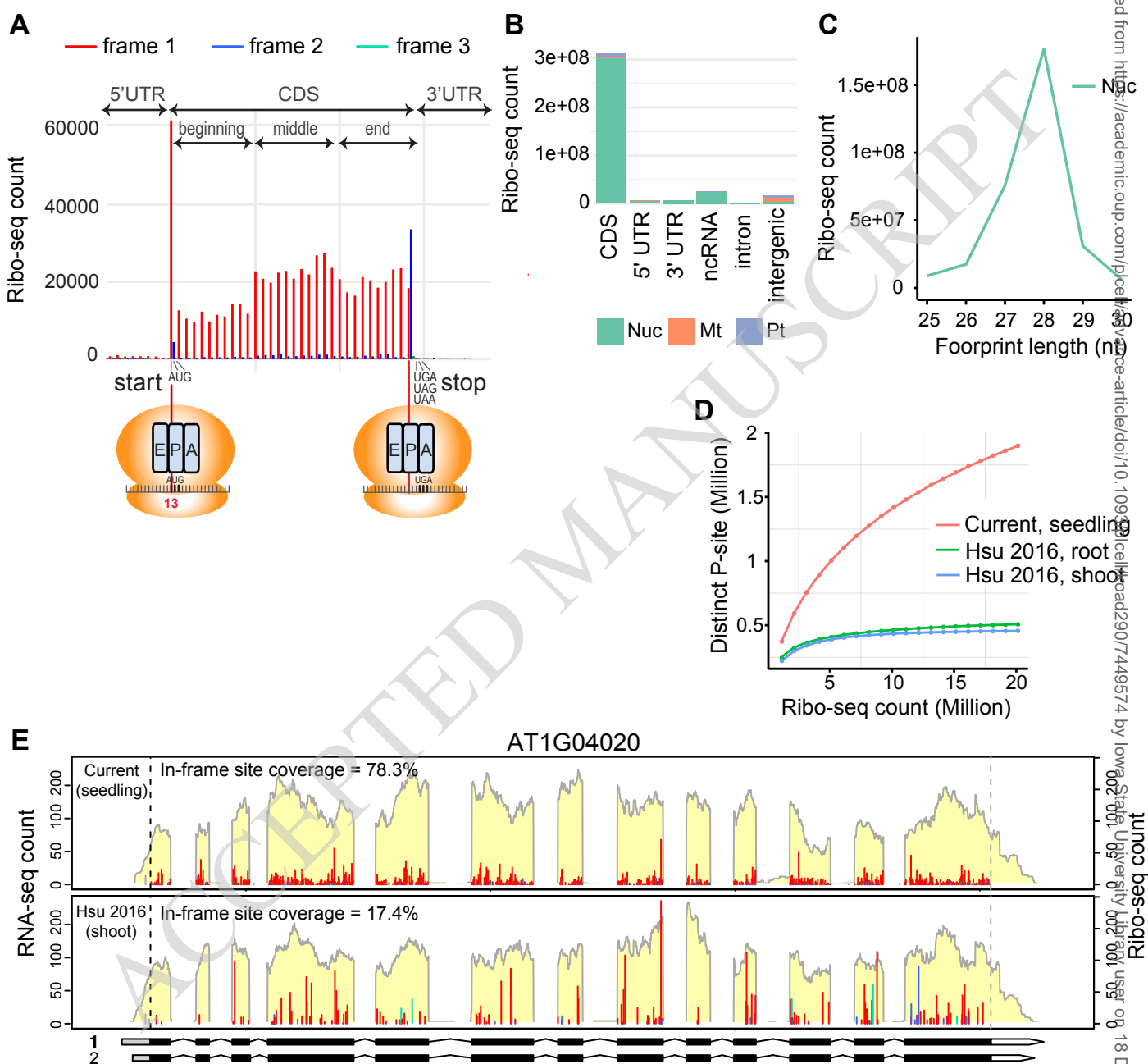
- 1507 Wu, H.-Y.L., Jen, J., and Hsu, P.Y. (2023). What, where, and how: Regulation of translation and
1508 the translational landscape in plants. *The Plant Cell*: koad197.
- 1509 Wu, H.Y.L., Song, G., Walley, J.W., and Hsu, P.Y. (2019). The tomato translational landscape
1510 revealed by transcriptome assembly and ribosome profiling. *Plant Physiology* 181: 367–
1511 380.
- 1512 Wu, Q., Wright, M., Gogol, M.M., Bradford, W.D., Zhang, N., and Bazzini, A.A. (2020).
1513 Translation of small downstream ORFs enhances translation of canonical main open
1514 reading frames. *The EMBO Journal* 39.
- 1515 Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., and Gao, C. (2020). Fine-tuning sugar
1516 content in strawberry. *Genome Biology* 21: 230.
- 1517 Xiong, H.-B., Pan, H.-M., Long, Q.-Y., Wang, Z.-Y., Qu, W.-T., Mei, T., Zhang, N., Xu, X.-F.,
1518 Yang, Z.-N., and Yu, Q.-B. (2022). AtNusG, a chloroplast nucleoid protein of bacterial
1519 origin linking chloroplast transcriptional and translational machineries, is required for
1520 proper chloroplast gene expression in *Arabidopsis thaliana*. *Nucleic Acids Res* 50:
1521 6715–6734.
- 1522 Xu, G., Yuan, M., Ai, C., Liu, L., Zhuang, E., Karapetyan, S., Wang, S., and Dong, X. (2017).
1523 UORF-mediated translation allows engineered plant disease resistance without fitness
1524 costs. *Nature* 545: 491–494.
- 1525 Yoshikawa, M., Iki, T., Numa, H., Miyashita, K., Meshi, T., and Ishikawa, M. (2016). A short
1526 open reading frame encompassing the microRNA173 target site plays a role in trans-
1527 acting small interfering RNA biogenesis. *Plant Physiology* 171: 359–368.
- 1528 Yu, X., Willmann, M.R., Anderson, S.J., and Gregory, B.D. (2016). Genome-wide mapping of
1529 uncapped and cleaved transcripts reveals a role for the nuclear mrna cap-binding
1530 complex in cotranslational RNA decay in *Arabidopsis*. *Plant Cell* 28: 2385–2397.
- 1531 Zhang, C., Ng, D.W.-K., Lu, J., and Chen, Z.J. (2012). Roles of target site location and
1532 sequence complementarity in trans-acting siRNA formation in *Arabidopsis*. *The Plant*
1533 *Journal* 69: 217–226.
- 1534 Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018). Genome
1535 editing of upstream open reading frames enables translational control in plants. *Nature*
1536 *Biotechnology* 36: 894–900.
- 1537 Zhang, H., Wang, Y., Wu, X., Tang, X., Wu, C., and Lu, J. (2021). Determinants of genome-
1538 wide distribution and evolution of uORFs in eukaryotes. *Nature Communications* 12: 1–
1539 17.
- 1540 Zhou, P., Silverstein, K.A.T., Gao, L., Walton, J.D., Nallu, S., Guhlin, J., and Young, N.D. (2013).
1541 Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery
1542 Application). *BMC Bioinformatics* 14: 335.
- 1543 Zhu, X.-T., Zhou, R., Che, J., Zheng, Y.-Y., Tahir Ul Qamar, M., Feng, J.-W., Zhang, J., Gao, J.,
1544 and Chen, L.-L. (2023). Ribosome profiling reveals the translational landscape and
1545 allele-specific translational efficiency in rice. *Plant Commun* 4: 100457.

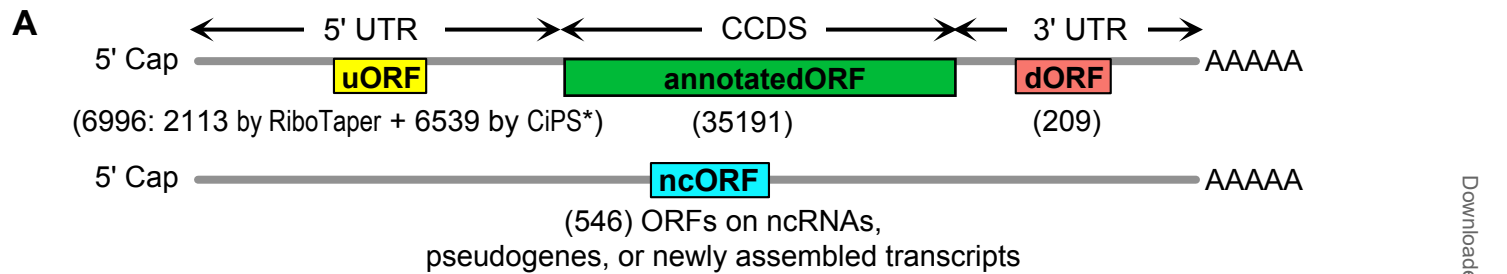
1546 Zoschke, R. and Bock, R. (2018). Chloroplast translation: Structural and functional organization,
1547 operational control, and regulation. *Plant Cell* 30: 745–770.

1548

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



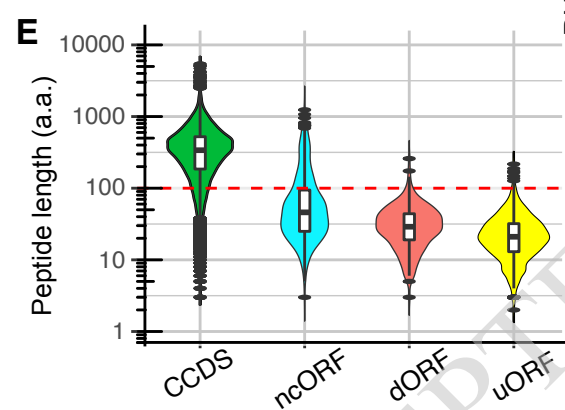


B ORFs by RiboTaper categories

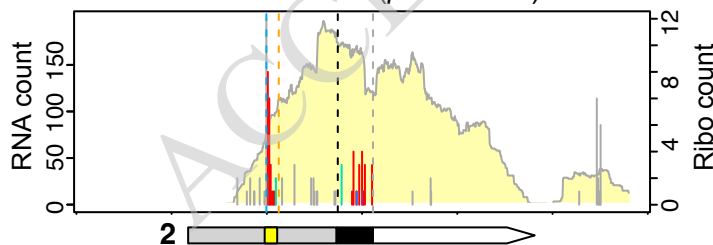
ORF type	ORFs by RiboTaper	# txs/genes for identified ORFs
CCDS	35191	31252/21038
uORF	2113	1884/1785
dORF	209	208/206
ncORF	546	449/397

D

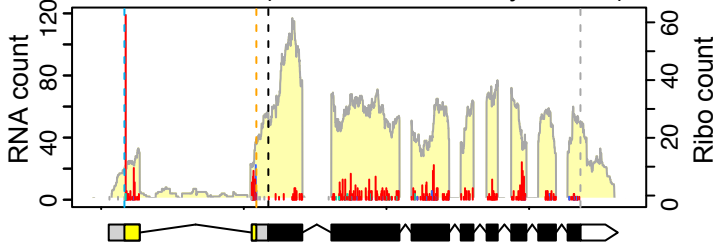
Localization	-	C	M	S
Number	307	4	28	128



G AT1G66725 (*pri-miR163*)



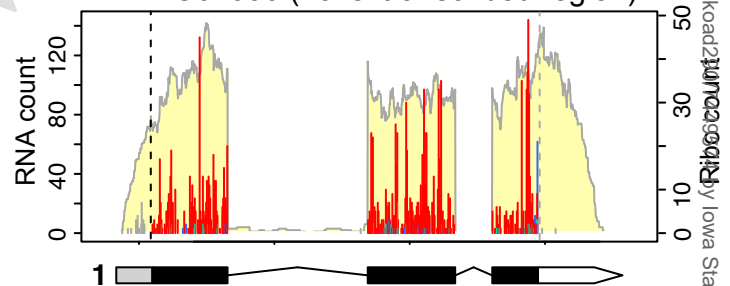
I AT1G05160 (*ATKAO1*, GA biosynthesis)



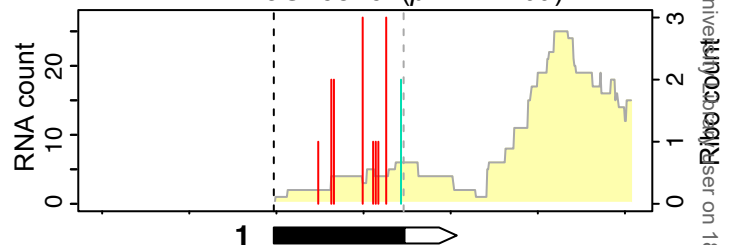
C ORFs by newly assembled RNAs and Araport11 transcript types

Transcript type	ORFs by RiboTaper	# txs/genes for identified ORFs	MS: #protein/ #peptide/ #spectra
Newly assembled RNA	164	135/130	13/70/382
Antisense lnc-RNA	54	42/40	3/10/39
Lnc-RNA	27	24/20	0/0/0
Novel transcribed region	66	57/49	4/12/27
Other RNA	97	81/69	2/2/7
miRNA precursor	2	2/2	0/0/0
Pseudogene	136	108/87	19/61/235
CCDS	35191	31252/21038	24874/215128/5726211
uORF	2113	1884/1785	2/3/32
dORF	209	208/206	4/6/52

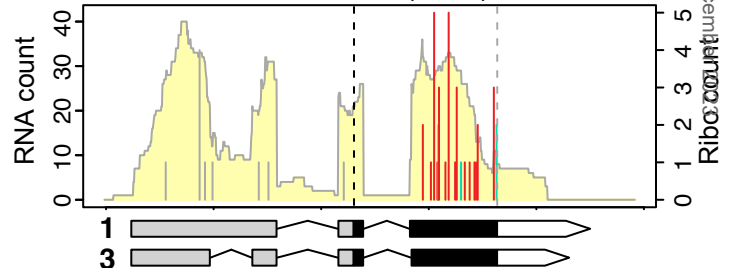
F AT4G02995 (novel transcribed region)

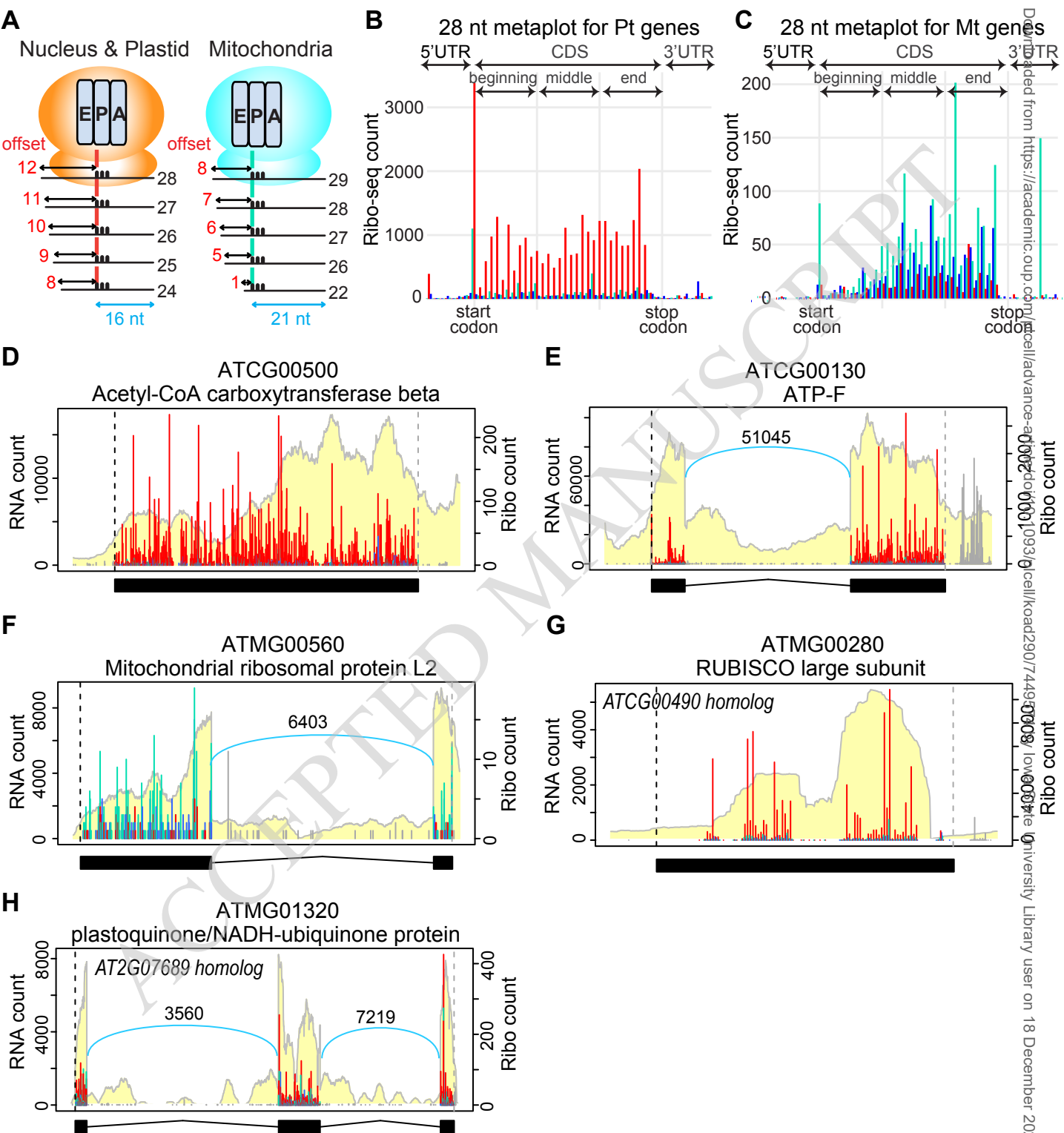


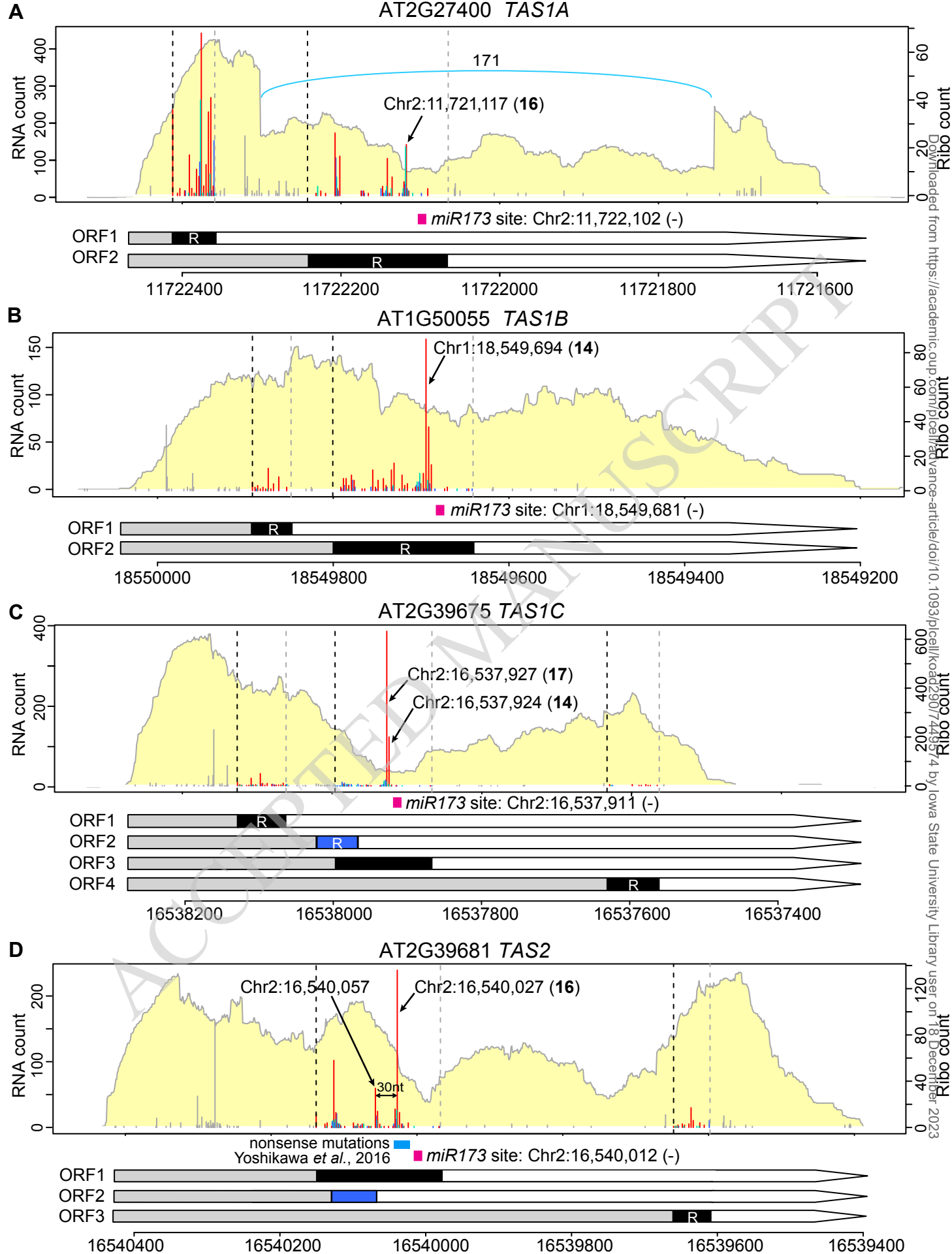
H AT3G26816 (*pri-miR169*)

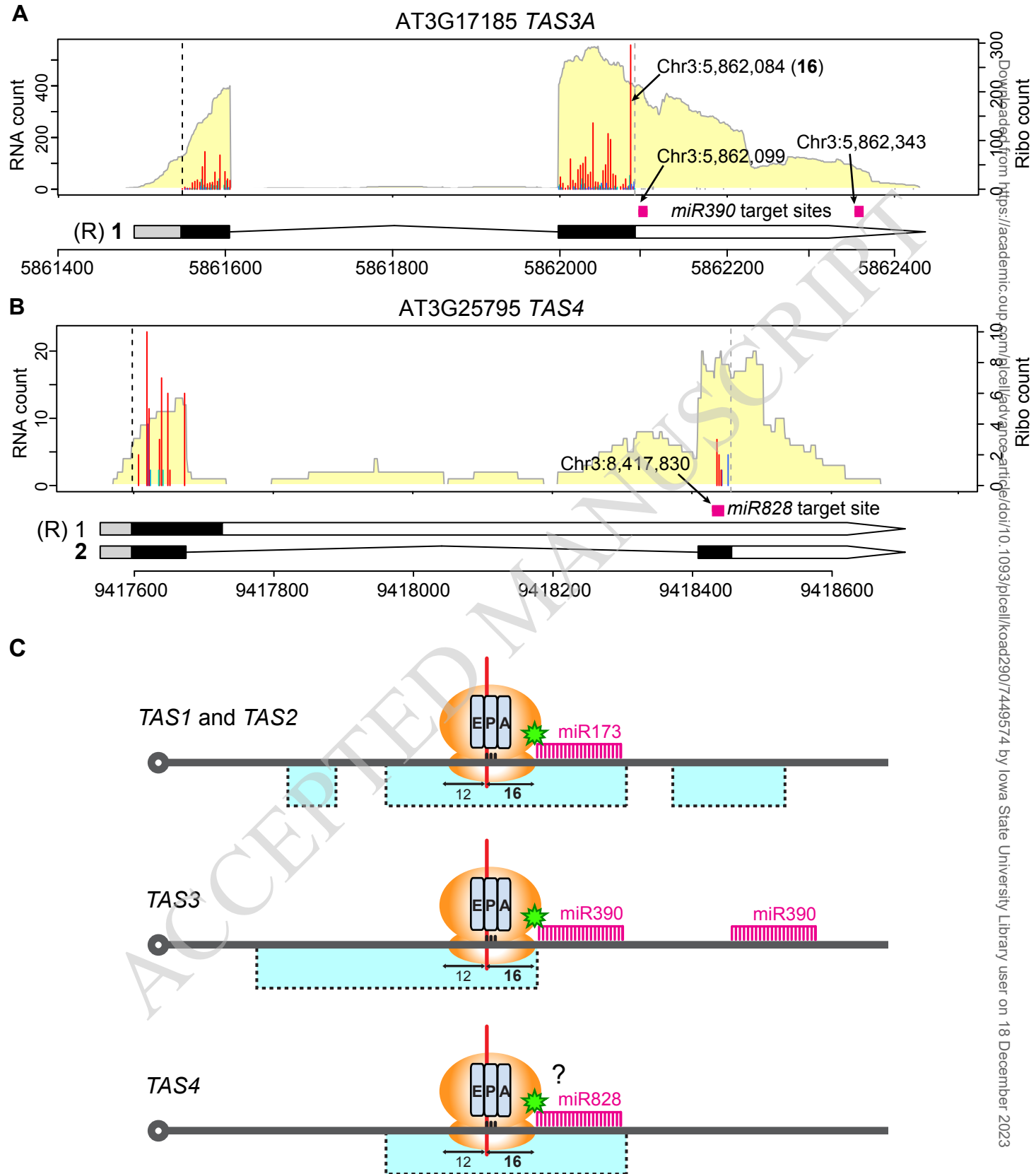


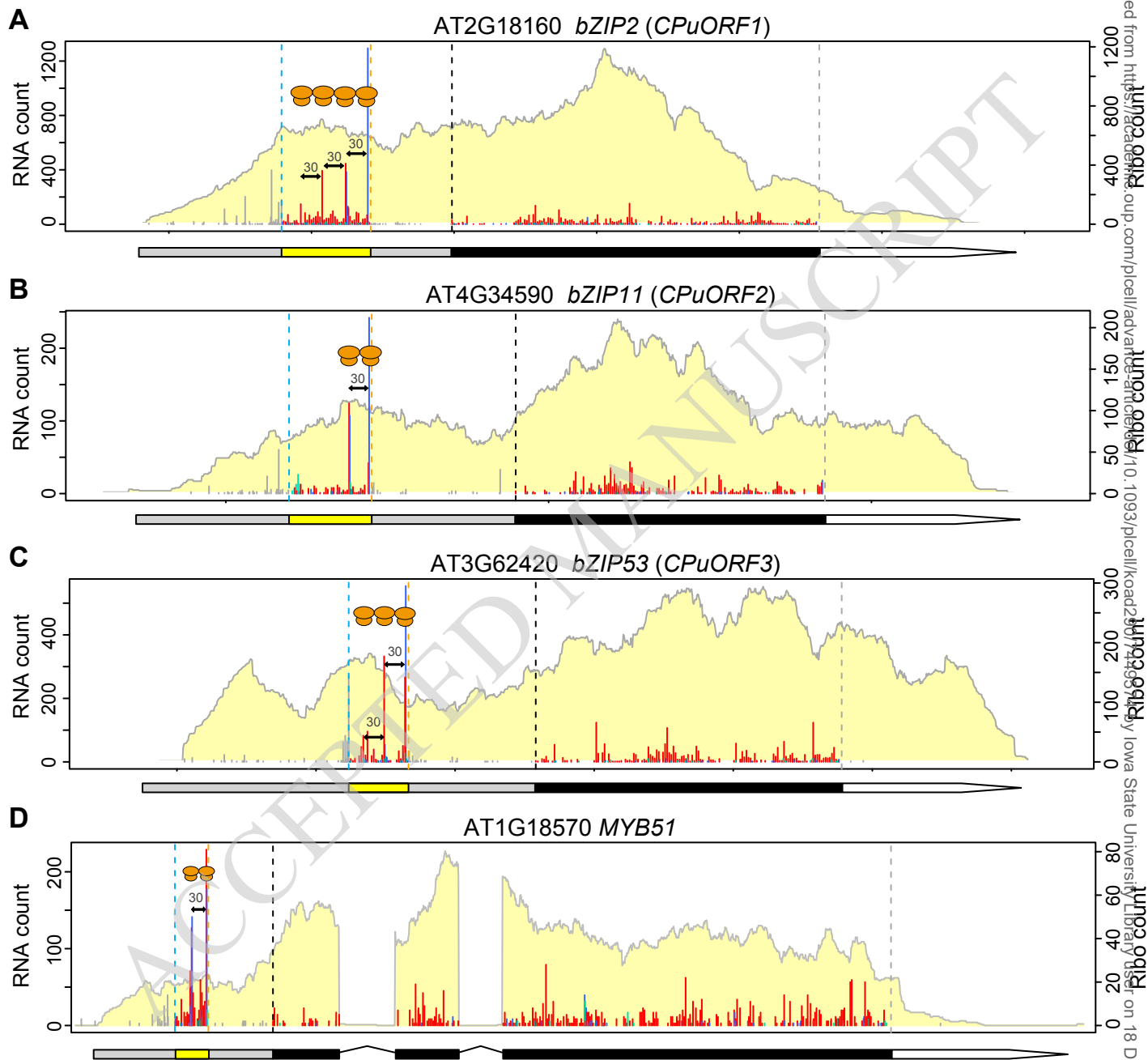
J AT3G30720 (*QQS*)

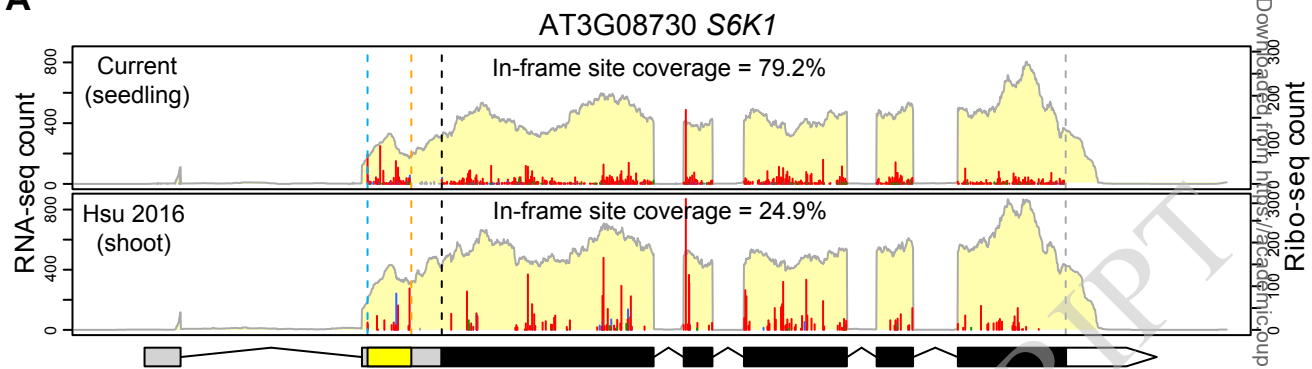
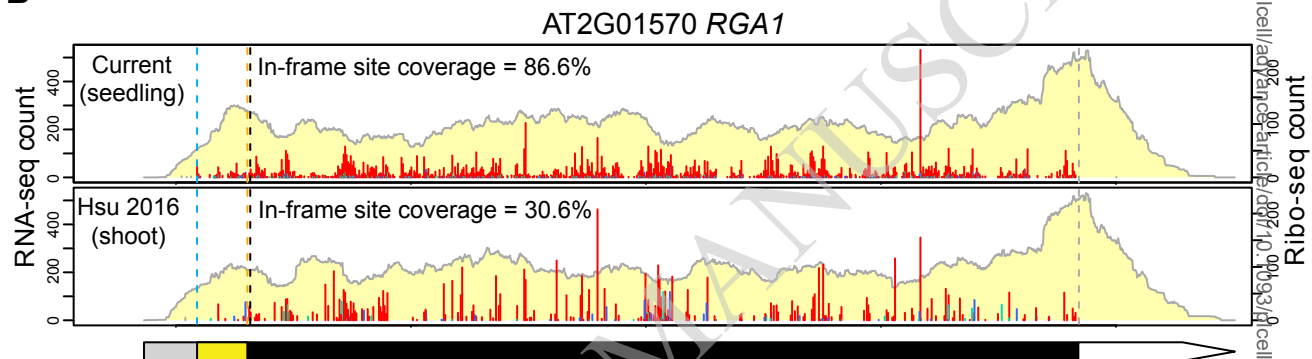
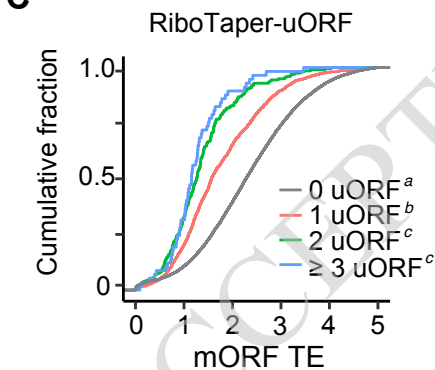
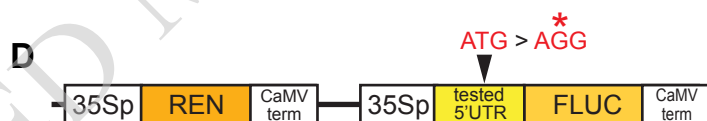
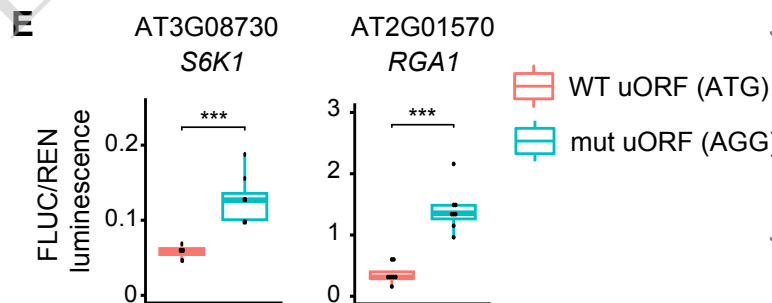
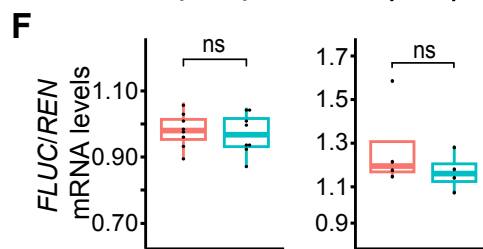


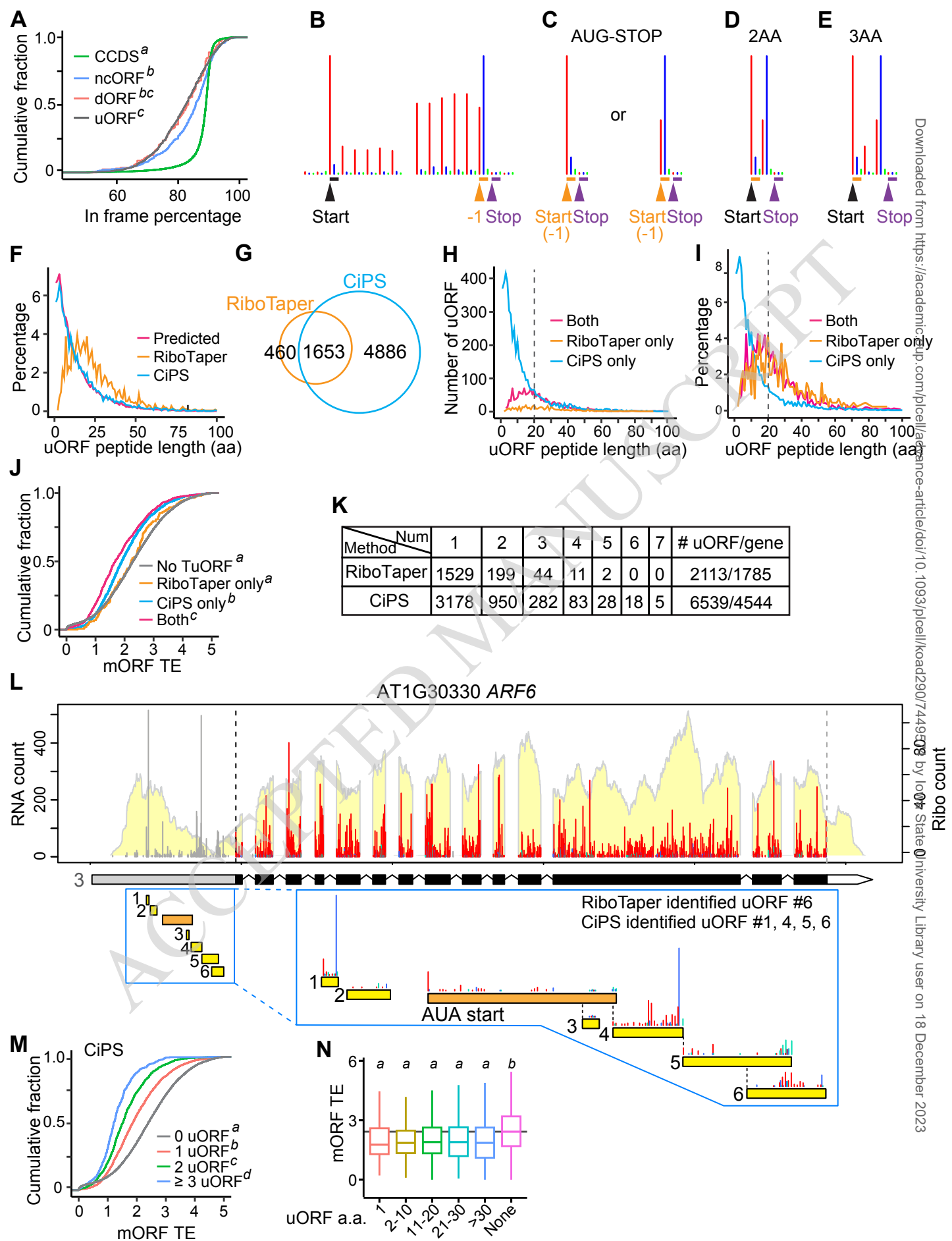


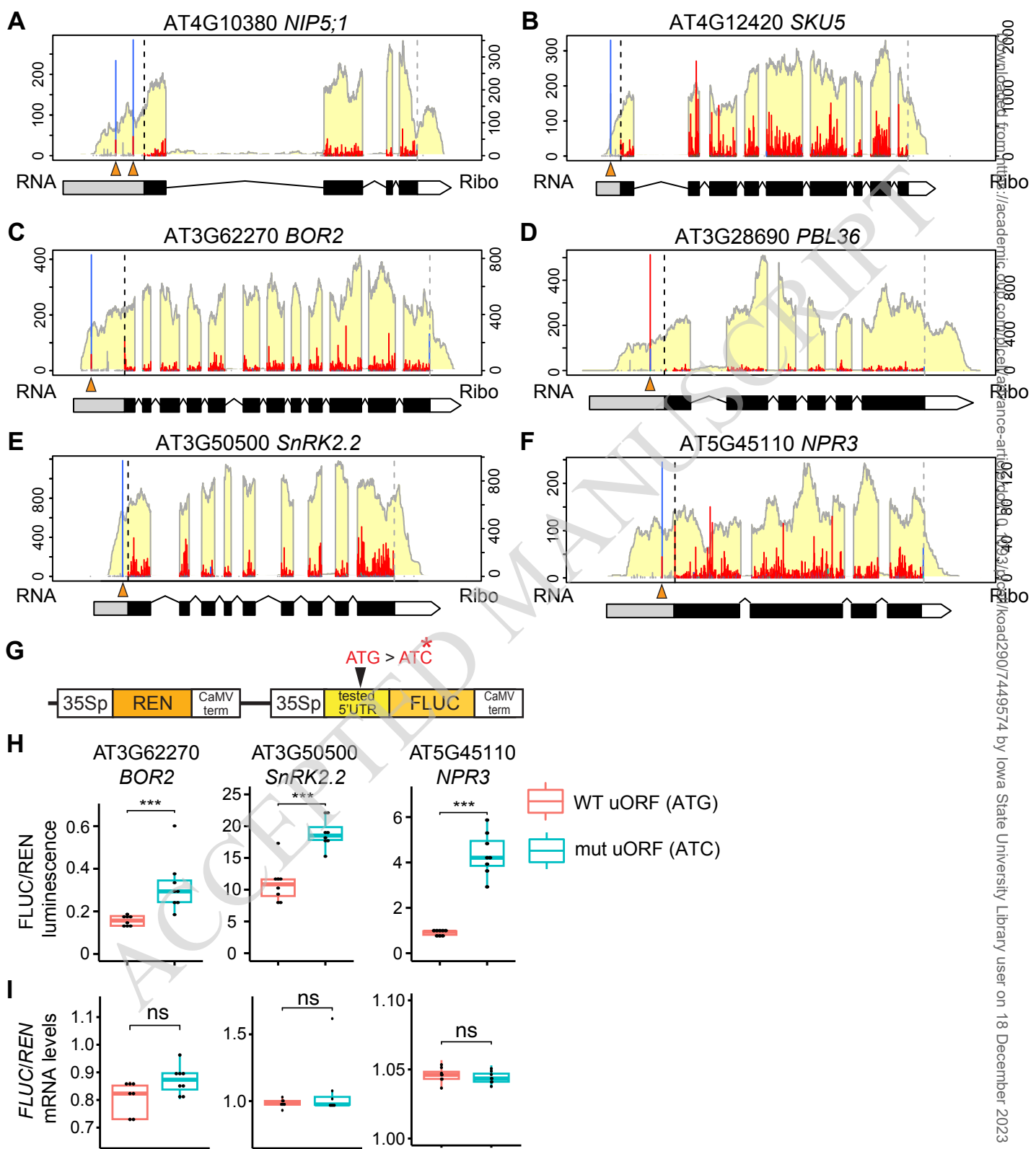


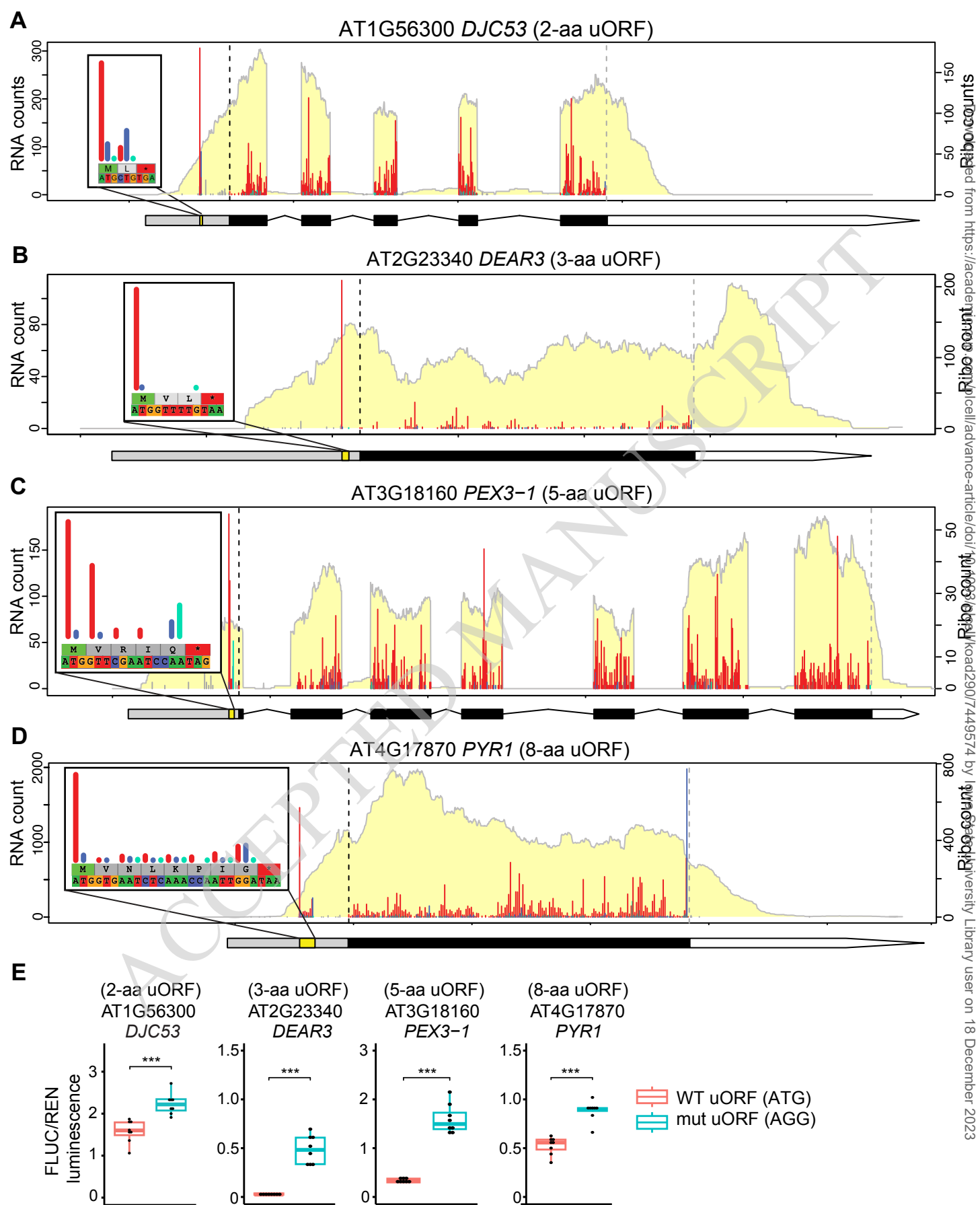


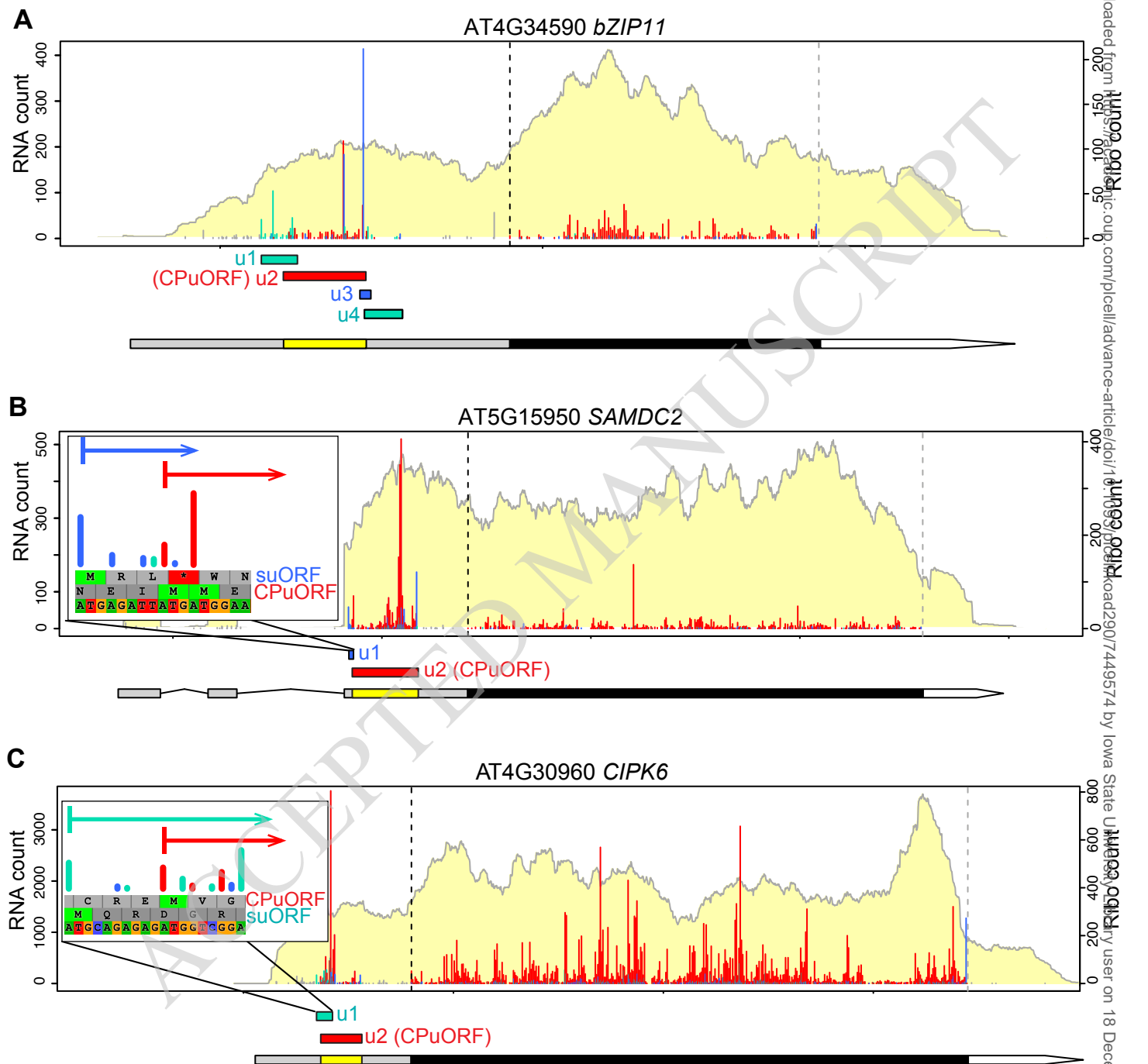


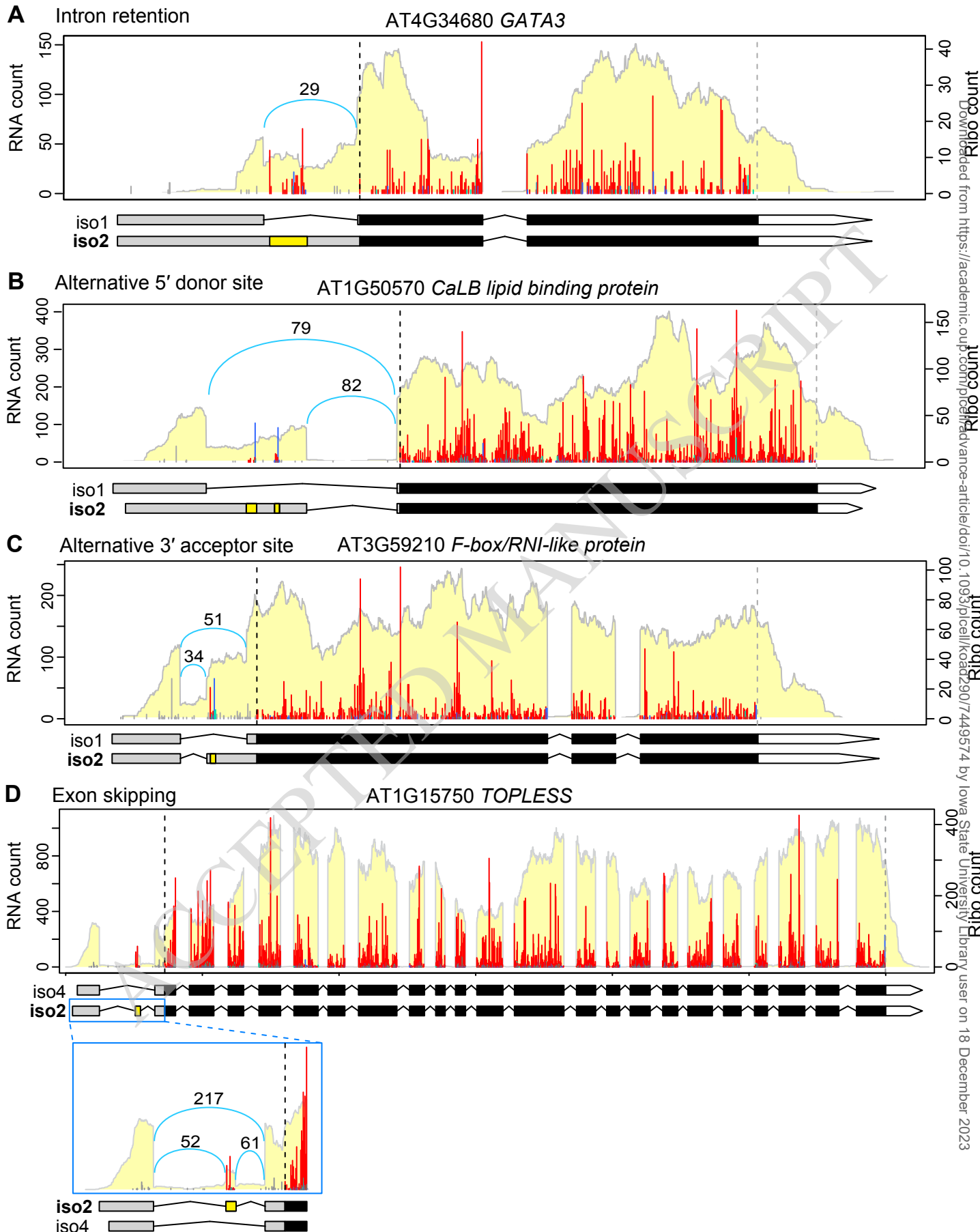
A**B****C****D****E****F**











Parsed Citations

Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L., and Pestova, T.V. (2006). In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors eRF1 and eRF3. *Cell* 125: 1125–1136.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Andrews, S.J. and Rothnagel, J.A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics* 15: 193–204.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Armenteros, J.J.A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* 2: e201900429.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

de Bang, T.C. et al. (2017). Genome-Wide Identification of Medicago Peptides Involved in Macronutrient Responses and Nodulation. *Plant Physiol* 175: 1669–1689.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Basrai, M.A., Hieter, P., and Boeke, J.D. (1997). Small open reading frames: Beautiful needles in the haystack. *Genome Research* 7: 768–771.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Bazin, J., Baerenfaller, K., Gosai, S.J., Gregory, B.D., Crespi, M., and Bailey-Serres, J. (2017). 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* 114: E10018–E10027.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Bazzini, A.A., Johnstone, T.G., Christiano, R., MacKowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., and Giraldez, A.J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO Journal* 33: 981–993.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Brar, G.A. and Weissman, J.S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology* 16: 651–664.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34: 525–527.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. *Nature* 524: 493–496.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Buskirk, A.R. and Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos Trans R Soc Lond B Biol Sci* 372: 20160183.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Calviello, L., Sydow, D., Harnett, D., and Ohler, U. (2019). Ribo-seQC: comprehensive analysis of cytoplasmic and organellar ribosome profiling data. *bioRxiv*: 601468.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant Journal* 89: 789–804.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Chotewutmontri, P. and Barkan, A. (2016). Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLoS Genetics* 12: e1006106.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C., and Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *Rna* 21: 1731–1745.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Clark, N.M., Nolan, T.M., Wang, P., Song, G., Montes, C., Valentine, C.T., Guo, H., Sozzani, R., Yin, Y., and Walley, J.W. (2021). Integrated omics networks reveal the temporal signaling events of brassinosteroid response in *Arabidopsis*. *Nat Commun* 12:

Google Scholar: [Author Only Title Only Author and Title](#)

Constabel, C.P., Yip, L., and Ryan, C.A. (1998). Prosystemin from potato, black nightshade, and bell pepper: Primary structure and biological activity of predicted systemin polypeptides. *Plant Molecular Biology* 36: 55–62.

Google Scholar: [Author Only Title Only Author and Title](#)

Da Rosa, E.L.S., Oleskovicz, C.F., and Aragão, B.N. (2004). Rapid prototyping in maxillofacial surgery and traumatology: Case report. *Brazilian Dental Journal* 15: 243–247.

Google Scholar: [Author Only Title Only Author and Title](#)

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.

Google Scholar: [Author Only Title Only Author and Title](#)

Dong, J., Chen, H., Deng, X.W., Irish, V.F., and Wei, N. (2020). Phytochrome B induces intron retention and translational inhibition of phytochrome-interacting factor3. *Plant Physiology* 182: 159–166.

Google Scholar: [Author Only Title Only Author and Title](#)

Fei, Q., Xia, R., and Meyers, B.C. (2013). Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* 25: 2400–2415.

Google Scholar: [Author Only Title Only Author and Title](#)

Feng, Y., Jiang, M., Yu, W., and Zhou, J. (2023). Identification of short open reading frames in plant genomes. *Frontiers in Plant Science* 14.

Google Scholar: [Author Only Title Only Author and Title](#)

Fesenko, I. et al. (2019). Distinct types of short open reading frames are translated in plant cells. *Genome Res* 29: 1464–1477.

Google Scholar: [Author Only Title Only Author and Title](#)

Fields, A.P., Rodriguez, E.H., Jovanovic, M., Stern-Ginossar, N., Haas, B.J., Mertins, P., Raychowdhury, R., Hacohen, N., Carr, S.A., Ingolia, N.T., Regev, A., and Weissman, J.S. (2015). A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. *Molecular Cell* 60: 816–827.

Google Scholar: [Author Only Title Only Author and Title](#)

Gage, J.L., Mali, S., McLoughlin, F., Khaipho-Burch, M., Monier, B., Bailey-Serres, J., Vierstra, R.D., and Buckler, E.S. (2022). Variation in upstream open reading frames contributes to allelic diversity in maize protein abundance. *Proceedings of the National Academy of Sciences of the United States of America* 119: e2112516119.

Google Scholar: [Author Only Title Only Author and Title](#)

Guo, R., Yu, X., and Gregory, B.D. (2023a). The identification of conserved sequence features of co-translationally decayed mRNAs and upstream open reading frames in angiosperm transcriptomes. *Plant Direct* 7: e479.

Google Scholar: [Author Only Title Only Author and Title](#)

Guo, Y. et al. (2023b). The translational landscape of bread wheat during grain development. *The Plant Cell: Online ahead of print*.

Google Scholar: [Author Only Title Only Author and Title](#)

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154: 240–251.

Google Scholar: [Author Only Title Only Author and Title](#)

Guydosh, N.R. and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell* 156: 950–962.

Google Scholar: [Author Only Title Only Author and Title](#)

Hanada, K. et al. (2013). Small open reading frames associated with morphogenesis are hidden in plant genomes. *Proceedings of the National Academy of Sciences of the United States of America* 110: 2395–2400.

Google Scholar: [Author Only Title Only Author and Title](#)

Hanada, K., Akiyama, K., Sakurai, T., Toyoda, T., Shinozaki, K., and Shiu, S.H. (2009). sORF finder: A program package to identify small open reading frames with high coding potential. *Bioinformatics* 26: 399–400.

Google Scholar: [Author Only Title Only Author and Title](#)

Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H., and Shiu, S.H. (2007). A large number of novel coding small open reading frames in the intergenic regions of the *Arabidopsis thaliana* genome are transcribed and/or under purifying selection. *Genome Research* 17: 632–640.

Google Scholar: [Author Only Title Only Author and Title](#)

Hanfrey, C., Elliott, K.A., Franceschetti, M., Mayer, M.J., Illingworth, C., and Michael, A.J. (2005). A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-responsive translation. *Journal of Biological Chemistry* 280: 39229–

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hanfey, C., Franceschetti, M., Mayer, M.J., Illingworth, C., and Michael, A.J. (2002). Abrogation of upstream open reading frame-mediated translational control of a plant S-adenosylmethionine decarboxylase results in polyamine disruption and growth perturbations. *Journal of Biological Chemistry* 277: 44131–44139.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hayden, C.A and Jorgensen, R.A (2007). Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC Biology* 5: 32.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* 1: 13.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hellens, R.P., Brown, C.M., Chisnall, M.A.W., Waterhouse, P.M., and Macknight, R.C. (2016). The Emerging World of Small ORFs. *Trends in Plant Science* 21: 317–328.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hou, C.Y., Lee, W.C., Chou, H.C., Chen, A.P., Chou, S.J., and Chen, H.M. (2016). Global analysis of truncated RNA ends reveals new insights into Ribosome Stalling in plants. *Plant Cell* 28: 2398–2416.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hsu, P.Y. and Benfey, P.N. (2018). Small but Mighty: Functional Peptides Encoded by Small ORFs in Plants. *Proteomics* 18: 1700038.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Collier, J. (2009). Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* 461: 225–229.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ingolia, N.T., Ghaemmighani, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324: 218–223.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Iwakawa, H. oki, Lam, A.Y.W., Mine, A., Fujita, T., Kiyokawa, K., Yoshikawa, M., Takeda, A., Iwasaki, S., and Tomari, Y. (2021). Ribosome stalling caused by the Argonaute-microRNA-SGS3 complex regulates the production of secondary siRNAs in plants. *Cell Reports* 35: 109300.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Jiang, M., Ning, W., Wu, S., Wang, X., Zhu, K., Li, A., Li, Y., Cheng, S., and Song, B. (2022). Three-nucleotide periodicity of nucleotide diversity in a population enables the identification of open reading frames. *Briefings in Bioinformatics* 23: 1–13.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J. (2016). Upstream ORF s are prevalent translational repressors in vertebrates. *The EMBO Journal* 35: 706–723.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Jorgensen, R.A and Dorantes-Acosta, A.E. (2012). Conserved peptide upstream open reading frames are associated with regulatory genes in angiosperms. *Frontiers in Plant Science* 3: 191.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J. (2014). 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* 111: E203–E212.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kim, B.H., Cai, X., Vaughn, J.N., and Von Arnim, A.G. (2007). On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome Biology* 8: R60.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the aug initiator codon by eukaryotic ribosomes. *Nucleic*

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

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

Lauressergues, D., Couzigou, J.M., San Clemente, H., Martinez, Y., Dunand, C., Bécard, G., and Combier, J.P. (2015). Primary transcripts of microRNAs encode regulatory peptides. *Nature* 520: 90–93.

Lease, K.A. and Walker, J.C. (2006). The *Arabidopsis* unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiology* 142: 831–838.

Lei, L. et al. (2015). Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant Journal* 84: 1206–1208.

Li, B. and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323.

Li, L. et al. (2015). QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. *Proceedings of the National Academy of Sciences of the United States of America* 112: 14734–14739.

Li, L., Foster, C.M., Gan, Q., Nettleton, D., James, M.G., Myers, A.M., and Wurtele, E.S. (2009). Identification of the novel protein QQS as a component of the starch metabolic network in *Arabidopsis* leaves. *Plant J* 58: 485–498.

Li, S. et al. (2016). Biogenesis of phased siRNAs on membrane-bound polysomes in *Arabidopsis*. *eLife* 5: e22750.

Li, Y.R. and Liu, M.J. (2020). Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants. *Genome Research* 30: 1418–1433.

Li, Z., Jin, J., Wang, Y., Long, W., Ding, Y., Hu, H., and Wei, L. (2023). ExamPLe: explainable deep learning framework for the prediction of plant small secreted peptides. *Bioinformatics* 39: btad108.

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). Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell* 25: 3699–3710.

Matheisl, S., Berninghausen, O., Becker, T., and Beckmann, R. (2015). Structure of a human translation termination complex. *Nucleic Acids Res* 43: 8615–8626.

Merret, R., Nagarajan, V.K., Carpentier, M.C., Park, S., Favory, J.J., Descombin, J., Picart, C., Charng, Y.Y., Green, P.J., Deragon, J.M., and Bousquet-Antonelli, C. (2015). Heat-induced ribosome pausing triggers mRNA co-translational decay in *Arabidopsis thaliana*. *Nucleic Acids Research* 43: 4121–4132.

Montes, C., Wang, P., Liao, C.Y., Nolan, T.M., Song, G., Clark, N.M., Elmore, J.M., Guo, H., Bassham, D.C., Yin, Y., and Walley, J.W. (2022). Integration of multi-omics data reveals interplay between brassinosteroid and Target of Rapamycin Complex signaling in *Arabidopsis*. *New Phytologist* 236: 893–910.

Olsen, A.N., Mundy, J., and Skriver, K. (2002). Peptomics, identification of novel cationic *Arabidopsis* peptides with conserved sequence motifs. *In Silico Biology* 2: 441–451.

Ong, S.N., Tan, B.C., Al-Idrus, A., and Teo, C.H. (2022). Small open reading frames in plant research: from prediction to functional characterization. *3 Biotech* 12: 76.

Orr, M.W., Mao, Y., Storz, G., and Qian, S.-B. (2020). Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res* 48: 1029–1042.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pearce, G., Strydom, D., Johnson, S., and Ryan, C.A. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science* 253: 895–898.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* 161: 1400–1412.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pertea, G. and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33: 290–295.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Planchard, N., Bertin, P., Quadrado, M., Dargel-Graffin, C., Hatin, I., Namy, O., and Mireau, H. (2018). The translational landscape of Arabidopsis mitochondria. *Nucleic Acids Research* 46: 6218–6228.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Puyaubert, J., Denis, L., and Alban, C. (2008). Dual targeting of Arabidopsis holocarboxylase synthetase1: A small upstream open reading frame regulates translation initiation and protein targeting. *Plant Physiology* 146: 478–491.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Qanmber, G., You, Q., Yang, Z., Fan, L., Zhang, Z., Chai, M., Gao, B., Li, F., and Yang, Z. (2023). Transcriptional and translational landscape fine-tune genome annotation and explores translation control in cotton. *J Adv Res: S2090-1232(23)00142-X*.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Qi, M. et al. (2019). QQS orphan gene and its interactor NF-YC4 reduce susceptibility to pathogens and pests. *Plant Biotechnol J* 17: 252–263.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Rahmani, F., Hummel, M., Schuurmans, J., Wiese-Klinkenberg, A., Smeekens, S., and Hanson, J. (2009). Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiology* 150: 1356–1367.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Raina, M., King, A., Bianco, C., and Vanderpool, C.K. (2018). Dual-Function RNAs. *Microbiol Spectr* 6.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Reis, R.S., Deforges, J., Sokoloff, T., and Poirier, Y. (2020). Modulation of shoot phosphate level and growth by PHOSPHATE1 upstream open reading frame. *Plant Physiology* 183: 1145–1156.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Rook, F., Gerrits, N., Kortstee, A., Van Kampen, M., Borrias, M., Weisbeek, P., and Smeekens, S. (1998). Sucrose-specific signalling represses translation of the Arabidopsis ATB2 bZIP transcription factor gene. *Plant Journal* 15: 253–263.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Roy, B., Vaughn, J.N., Kim, B.H., Zhou, F., Gilchrist, M.A., and Von Arnim, A.G. (2010). The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. *Rna* 16: 748–761.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ruiz-Orera, J., Messegue, X., Subirana, J.A., and Alba, M.M. (2014). Long non-coding RNAs as a source of new peptides. *eLife* 3: e03523.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sahoo, S., Singh, D., Singh, A., Pandit, M., Vasu, K., Som, S., Pullagurla, N.J., Laha, D., and Eswarappa, S.M. (2022). Identification and functional characterization of mRNAs that exhibit stop codon readthrough in Arabidopsis thaliana. *Journal of Biological Chemistry* 298: 102173.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sandmann, C.-L. et al. (2023). Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. *Mol Cell* 83: 994-1011.e18.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Schuller, A.P. and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology* 19: 526–541.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sharma, A., Badola, P.K., Bhatia, C., Sharma, D., and Trivedi, P.K. (2020). Primary transcript of miR858 encodes regulatory peptide

and controls flavonoid biosynthesis and development in *Arabidopsis*. *Nature Plants* 6: 1262–1274.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Song, G., Brachova, L., Nikolau, B.J., Jones, A.M., and Walley, J.W. (2018a). Heterotrimeric G-Protein-Dependent Proteome and Phosphoproteome in Unstimulated *Arabidopsis* Roots. *Proteomics* 18: 1800323.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Song, G., Montes, C., and Walley, J.W. (2020). Quantitative Profiling of Protein Abundance and Phosphorylation State in Plant Tissues Using Tandem Mass Tags. In *Plant Proteomics: Methods and Protocols*, J.V. Jorri n-Novo, L. Villedor, M.A. Castillejo, and M.-D. Rey, eds, *Methods in Molecular Biology*. (Springer US: New York, NY), pp. 147–156.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sotta, N., Chiba, Y., Aoyama, H., Takamatsu, S., Suzuki, T., Miwa, K., Yamashita, Y., Naito, S., and Fujiwara, T. (2022). Translational Landscape of a C4 Plant, *Sorghum bicolor*, Under Normal and Sulfur-Deficient Conditions. *Plant and Cell Physiology* 63: 592–604.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sotta, N., Chiba, Y., Miwa, K., Takamatsu, S., Tanaka, M., Yamashita, Y., Naito, S., and Fujiwara, T. (2021). Global analysis of boron-induced ribosome stalling reveals its effects on translation termination and unique regulation by AUG-stops in *Arabidopsis* shoots. *Plant Journal* 106: 1455–1467.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Takahashi, F., Hanada, K., Kondo, T., and Shinozaki, K. (2019). Hormone-like peptides and small coding genes in plant stress signaling and development. *Current Opinion in Plant Biology* 51: 88–95.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Takahashi, H. et al. (2020). Comprehensive genome-wide identification of angiosperm upstream ORFs with peptide sequences conserved in various taxonomic ranges using a novel pipeline, ESUCA. *BMC Genomics* 21: 260.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Takahashi, H., Takahashi, A., Naito, S., and Onouchi, H. (2012). BAUCAS: A novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the *Arabidopsis thaliana* genome. *Bioinformatics* 28: 2231–2241.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B.P.A. (2015). The plant peptidome: An expanding repertoire of structural features and biological functions. *Plant Cell* 27: 2095–2118.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Thieffry, A., Vigh, M.L., Bornholdt, J., Ivanov, M., Brodersen, P., and Sandelin, A. (2020). Characterization of *Arabidopsis thaliana* promoter bidirectionality and antisense RNAs by inactivation of nuclear RNA decay pathways. *Plant Cell* 32: 1845–1867.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tr sch, R. (2022). Chloroplast transcription and translation are linked. *Nat. Plants* 8: 867–867.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols* 11: 2301–2319.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ulveling, D., Francastel, C., and Hub , F. (2011). When one is better than two: RNA with dual functions. *Biochimie* 93: 633–644.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Van Der Horst, S., Filipovska, T., Hanson, J., and Smeekens, S. (2020). Metabolite control of translation by conserved peptide uORFs: The ribosome as a metabolite multisensor. *Plant Physiology* 182: 110–122.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Van Der Horst, S., Snel, B., Hanson, J., and Smeekens, S. (2019). Novel pipeline identifies new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5' leader of mRNAs in *Arabidopsis thaliana*. *RNA* 25: 292–304.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Vaughn, J.N., Ellington, S.R., Mignone, F., and Von Arnim, A. (2012). Known and novel post-transcriptional regulatory sequences

are conserved across plant families. *Rna* 18: 368–384.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wang, H., Wang, Y., and Xie, Z. (2019). Computational resources for ribosome profiling: from database to Web server and software. *Brief Bioinform* 20: 144–155.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wang, Y., Zhang, H., and Lu, J. (2020). Recent advances in ribosome profiling for deciphering translational regulation. *Methods* 176: 46–54.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wiese, A., Elzinga, N., Wobbes, B., and Srneekens, S. (2004). A conserved upstream open reading frame mediates sucrose-induced repression of translation. *Plant Cell* 16: 1717–1729.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wolin, S.L. and Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO journal* 7: 3559–3569.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, H.Y.L. and Hsu, P.Y. (2022). A custom library construction method for super-resolution ribosome profiling in Arabidopsis. *Plant Methods* 18: 1–8.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, H.Y.L. and Hsu, P.Y. (2021). RiboPlotR: a visualization tool for periodic Ribo-seq reads. *Plant Methods* 17: 1–7.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, H.-Y.L., Jen, J., and Hsu, P.Y. (2023). What, where, and how: Regulation of translation and the translational landscape in plants. *The Plant Cell*: koad197.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, H.Y.L., Song, G., Walley, J.W., and Hsu, P.Y. (2019). The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiology* 181: 367–380.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, Q., Wright, M., Gogol, M.M., Bradford, W.D., Zhang, N., and Bazzini, A.A. (2020). Translation of small downstream ORFs enhances translation of canonical main open reading frames. *The EMBO Journal* 39.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., and Gao, C. (2020). Fine-tuning sugar content in strawberry. *Genome Biology* 21: 230.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Xiong, H.-B., Pan, H.-M., Long, Q.-Y., Wang, Z.-Y., Qu, W.-T., Mei, T., Zhang, N., Xu, X.-F., Yang, Z.-N., and Yu, Q.-B. (2022). AtNusG, a chloroplast nucleoid protein of bacterial origin linking chloroplast transcriptional and translational machineries, is required for proper chloroplast gene expression in *Arabidopsis thaliana*. *Nucleic Acids Res* 50: 6715–6734.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Xu, G., Yuan, M., Ai, C., Liu, L., Zhuang, E., Karapetyan, S., Wang, S., and Dong, X. (2017). UORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* 545: 491–494.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Yoshikawa, M., Iki, T., Numa, H., Miyashita, K., Meshi, T., and Ishikawa, M. (2016). A short open reading frame encompassing the microRNA173 target site plays a role in trans-acting small interfering RNA biogenesis. *Plant Physiology* 171: 359–368.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Yu, X., Willmann, M.R., Anderson, S.J., and Gregory, B.D. (2016). Genome-wide mapping of uncapped and cleaved transcripts reveals a role for the nuclear mRNA cap-binding complex in cotranslational RNA decay in *Arabidopsis*. *Plant Cell* 28: 2385–2397.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhang, C., Ng, D.W.-K., Lu, J., and Chen, Z.J. (2012). Roles of target site location and sequence complementarity in trans-acting siRNA formation in *Arabidopsis*. *The Plant Journal* 69: 217–226.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018). Genome editing of upstream open reading frames enables translational control in plants. *Nature Biotechnology* 36: 894–900.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhou, P., Silverstein, K.A.T., Gao, L., Walton, J.D., Nallu, S., Guhlin, J., and Young, N.D. (2013). Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinformatics* 14: 335.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhu, X.-T., Zhou, R., Che, J., Zheng, Y.-Y., Tahir Ul Qamar, M., Feng, J.-W., Zhang, J., Gao, J., and Chen, L.-L. (2023). Ribosome profiling reveals the translational landscape and allele-specific translational efficiency in rice. *Plant Commun* 4: 100457.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zoschke, R. and Bock, R. (2018). Chloroplast translation: Structural and functional organization, operational control, and regulation. *Plant Cell* 30: 745–770.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

ACCEPTED MANUSCRIPT