Plant Physiology Preview. Published on June 27, 2019, as DOI:10.1104/pp.19.00541

Short title: Tomato translatome revealed by ribosome profiling

* To whom correspondence should be addressed: pollyhsu@msu.edu

The tomato translational landscape revealed by transcriptome assembly and ribosome profiling

Hsin-Yen Larry Wu[1], Gaoyuan Song[2], Justin W Walley[2], and Polly Yingshan Hsu[1,*]

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

[2] Department of Plant Pathology & Microbiology, Iowa State University, Ames, IA 50011 USA

One-sentence summary:

Ribosome profiling revealed previously unannotated ORFs, elucidated evolutionarily conserved and unique translational features, and identifies regulatory mechanisms hidden in the tomato genome.

FOOTNOTES

Author contributions

HLW and PYH designed the research; PYH performed the sequencing experiments; HLW and PYH analyzed the sequencing data; GS and JWW performed the proteomic experiments and analyzed the proteomic data; HLW and PYH wrote the paper with input from all authors.

The author responsible for distributing materials integral to the findings presented in this article in accordance with the policy described in the guidelines from the Committee on Publication

Ethics (COPE) (http://publicationethics.org/resources/guidelines) is: Polly Hsu (pollyhsu@msu.edu)

Funding information:

This work was supported by a US Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA) postdoctoral fellowship [2016-67012-24720] and Michigan State University startup grant to P.Y.H., and National Science Foundation [1759023], USDA NIFA Hatch project 3808, and the Iowa State University Plant Sciences Institute awards to J.W.W.

ABSTRACT

Recent applications of translational control in *Arabidopsis thaliana* (Arabidopsis) highlight the potential power of manipulating mRNA translation for crop improvement. However, to what extent translational regulation is conserved between Arabidopsis and other species is largely unknown and the translatome of most crops remains poorly studied. Here, we combined *de novo* transcriptome assembly and ribosome profiling to study global mRNA translation in tomato (*Solanum lycopersicum*) roots. Exploiting features corresponding to active translation, we discovered widespread unannotated translation events, including 1329 upstream ORFs (uORFs) within the 5′ UTRs of annotated coding genes and 354 small ORFs (sORFs) among unannotated transcripts. uORFs may repress translation of their downstream main ORFs, whereas sORFs may encode signaling peptides. Besides evolutionarily conserved sORFs, we uncovered 96 Solanaceae-specific sORFs, revealing the importance of studying translatomes directly in crops. Proteomic analysis confirmed that some of the unannotated ORFs generate stable proteins *in planta*. In addition to defining the translatome, our results reveal the global regulation by uORFs and microRNAs. Despite diverging over 100 million years ago, many translational features are well conserved between Arabidopsis and tomato. Thus, our approach provides a high-throughput method to discover unannotated ORFs, elucidates evolutionarily

conserved and unique translational features, and identifies regulatory mechanisms hidden in a crop genome.

INTRODUCTION

Besides being an essential step in gene expression, mRNA translation directly shapes the proteome, which contributes to cellular structure, function, and activity in all organisms. The characterization of translational regulation in Arabidopsis has enabled crop improvement, including increasing tomato (*Solanum lycopersicum*) sweetness, rice (*Oryza sativa*) immunity and lettuce (*Lactuca sativa*) resistance to oxidative stress (Sagor et al., 2016; Xu et al., 2017b; Zhang et al., 2018). However, not everything in Arabidopsis is applicable to other plants and how the Arabidopsis translatome compares to other species is largely unknown. Moreover, due to limited genomic resources and methods, translational landscapes and their underlying regulations in crops remain understudied.

Ribosome profiling, or Ribo-seq, has emerged as a high-throughput technique to study global translation (Ingolia et al., 2009; Brar and Weissman, 2015; Andreev et al., 2017). In a Ribo-seq experiment, ribosomes in the sample of interest are immobilized, and the lysate is treated with nucleases to obtain ribosome-protected mRNA fragments (i.e. ribosome footprints). Finally, sequencing of the ribosome footprints reveals the quantity and positions of ribosomes on a given transcript. Because ribosomes decipher mRNA every 3 nucleotide (nt), the periodic feature of ribosome footprints can be used to uncover previously unannotated translation events (Bazzini et al., 2014; Fields et al., 2015; Ji et al., 2015; Calviello et al., 2016; Hsu et al., 2016). For example, upstream open reading frames (uORFs) in the 5′ leader sequence or 5′ untranslated region (UTR) have been shown to be widespread in many protein-coding genes in humans, mouse, zebrafish, yeast, and plants (Brar et al., 2012; Liu et al., 2013; Ji et al., 2015; Lei et al., 2015; Chew et al., 2016; Hsu et al., 2016; Johnstone et al., 2016). Several well-characterized examples and global analyses indicate that uORFs can modulate the translation

of their downstream main ORFs (von Arnim et al., 2013; Liu et al., 2013; Lei et al., 2015; Chew et al., 2016; Johnstone et al., 2016; Hsu and Benfey, 2018). Moreover, numerous presumed non-coding RNAs have been found to possess translated small ORFs (sORFs), usually below 100 codons (Bazzini et al., 2014; Hsu et al., 2016; Bazin et al., 2017; Ruiz-Orera and Albà, 2019). The small size of the protein products of sORFs suggest that they may serve as signaling peptides (Hsu and Benfey, 2018; Ruiz-Orera and Albà, 2019). Despite their importance, uORFs and sORFs are often missing in annotations because computational predictions often assume that 1) protein-coding sequences encode proteins greater than 100 amino acids, and 2) only the longest ORF in a transcript is translated (Basrai et al., 1997; Claverie, 1997). Thus, ribosome profiling provides an unparalleled opportunity to experimentally identify translated ORFs genome-wide in an unbiased manner.

In plant research, ribosome profiling has been used to study translational regulation in diverse aspects of plant development and response to stress including photomorphogenesis, chloroplast differentiation, cotyledon development, hypoxia, hormone responses, nutrient deprivation, drought, pathogen responses, and biogenesis of small interfering RNAs (Liu et al., 2013; Zoschke et al., 2013; Juntawong et al., 2014; Lei et al., 2015; Merchante et al., 2015; Chotewutmontri and Barkan, 2016; Li et al., 2016; Bazin et al., 2017; Xu et al., 2017a; Shamimuzzaman and Vodkin, 2018). We previously optimized the resolution of this technique to resolve 3-nt periodicity, which enabled us to precisely define translated regions within individual transcripts, in Arabidopsis. As a result, we were able to identify previously unannotated translation events, including usage of non-AUG start site, uORFs in 5′ UTRs, and sORFs in annotated non-coding RNAs (Hsu et al., 2016). To date, systematically identifying translated ORFs in plants has only been attempted in Arabidopsis (Hsu et al., 2016; Bazin et al., 2017).

Tomato is the most widely cultivated vegetable worldwide (Schwarz et al., 2014). It belongs to the Solanaceae, whose members produce important foods, spices, and medicines. Like other crops, tomato has limited genomic resources or optimized methods. For instance, the

latest annotation, ITAG3.2 for the 'Heinz 1706' cultivar, only contains predicted protein-coding genes whereas non-coding RNAs and uORFs are not included (Fernandez-Pozo et al., 2015). We chose seedling roots to establish the protocol for translatomic analysis for several reasons: 1) the root plays an essential role in water/nutrient uptake as well as interaction between plants and other organisms or the environment; 2) the root is composed of diverse cell types, which is beneficial for surveying translation events, as we observed in our previous work in Arabidopsis seedlings (Hsu et al., 2016). Here, we performed ribosome profiling in combination with *de novo* transcriptome assembly to discover non-coding RNAs, uORFs and sORFs, and chart the translational landscape in tomato roots. The mapping and quantification of ribosome footprints in tomato not only uncovered numerous unannotated translation events but also revealed global features involved in translational regulation.

RESULTS

Establishment of an experimental and data analysis pipeline to map the tomato translatome

To map actively translated ORFs, we isolated the roots of tomato seedlings (*S. lycopersicum*, 'Heinz 1706' cultivar) and performed strand-specific RNA-seq and Ribo-seq in parallel (Figure 1A and 1B). RNA-seq reveals transcript identity and abundance, whereas Ribo-seq maps and quantifies ribosome occupancy on a given transcript (Brar and Weissman, 2015). We adapted our protocol and pipeline for Arabidopsis (Hsu et al., 2016) with two major modifications: 1) we increased the amount of RNase I used in tomato ribosome footprinting to achieve comparable resolution (see Methods for details); 2) we performed paired-end 100-bp RNA-seq followed by reference-guided *de novo* transcriptome assembly to capture transcripts missing from the ITAG3.2 reference annotation (Figure 1C, see Methods for details). This strategy allowed us to map the translated regions in both annotated and previously unannotated transcripts in an unbiased manner using the ORF-finding tool, RiboTaper (Calviello et al., 2016).

As the quality of ribosome footprints is critical for finding ORFs (Hsu et al., 2016), we first systematically evaluated the Ribo-seq results by mapping the reads to the ITAG3.2 annotation. Consistent with observations in other non-plant organisms and Arabidopsis (Ingolia et al., 2009; Bazzini et al., 2014; Hsu et al., 2016), the dominant ribosome footprints in tomato were 28 nt long (Figure 2A). Moreover, in contrast to RNA-seq, the Ribo-seq reads predominantly mapped to the annotated coding sequences (CDSs) and were sparse in the 5′ UTRs and 3′ UTRs (Figure 2B and 2C). The three biological replicates were highly correlated, as indicated by the Pearson correlation, in both Ribo-seq ($r = 0.998\sim1$) and RNA-seq ($r = 0.998\sim0.999$) (Supplemental Figure S1A and B). Overall, the RNA-seq and Ribo-seq datasets also showed a strong positive correlation (Pearson correlation after removing two extreme outliers, $r = 0.878–0.880$; Spearman correlation with all data points, $\rho = 0.912–0.915$) (Supplemental Figure S1C-F). Most importantly, the distribution of ribosome footprints within the CDS displayed clear 3-nt periodicity, a signature of translating ribosomes, which decipher 3 nt at a time (Figure 2C, and Supplemental Figure S2). Analyzing the distribution of footprints relative to the annotated translation start/stop sites allowed us to infer that the codon at the P-site within the ribosome is located between the 13^{th} and 15^{th} nts for 28-nt footprints, and so on for specific footprint lengths (Supplemental Figure S2 and Supplemental Figure S3). To visualize the position of the codon being translated, hereafter, we use the first nucleotides of the P-sites (denoted as P-site signals) to indicate the positions of the footprints on the transcripts (Figure 2C). The robustness of the 3-nt periodicity can be quantified based on the percentage of reads in the expected reading frame (shown in red in Figure 2C and hereafter). At a global level, our 28-nt footprints resulted in 85.5% in-frame reads. Together, these results demonstrate that our tomato Ribo-seq dataset is of high quality compared to datasets from plants and other organisms (Bazzini et al., 2014; Guydosh and Green, 2014; Chung et al., 2015; Schafer et al., 2015; Hsu et al., 2016).

Next, we performed reference-guided *de novo* transcriptome assembly for the RNA-seq data using *stringtie*, a transcript assembler (Pertea et al., 2015). Then, the newly assembled

transcriptomes from the replicates were merged and compared to the ITAG3.2 annotations using *gffcompare* software (Pertea et al., 2016) (Figure 1C). In total, we uncovered 2263 unannotated transcripts that could potentially encode for novel proteins. These transcripts could be classified into six groups based on their strands and genomic positions relative to existing gene features, such as intergenic (class-u), cis-natural antisense transcripts (cis-NAT, class-x), intronic (class-i) and others (class-y and class-o) (Figure 3A and 3C, the nomenclature and descriptions of these discovered transcripts are adapted based on the *gffcompare* software; Pertea et al., 2016). Class-s is expected to result from mapping errors (Pertea et al., 2016) and was included in our downstream analysis as a negative control. The most abundant classes of uncharacterized transcript in our data were intergenic transcripts (class-u; 1260) and cis-NATs (class-x; 568). All six classes of uncharacterized transcripts, along with the annotated genes in ITAG3.2, were used to find translated ORFs.

Translational landscape of tomato roots as defined by ribosome profiling

After collecting the transcript information, we used RiboTaper (Calviello et al., 2016) to interrogate both the annotated transcripts in ITAG3.2 and the newly assembled transcripts to search for all possible ORFs in the transcriptome. RiboTaper examines the P-site signals within each possible ORF and tests whether the signals display a statistically significant 3-nt periodicity (Calviello et al., 2016). As a quality control, we first examined translated ORFs detected at annotated coding regions. In total, 20659 annotated ORFs were identified as translated in our dataset (Figure 3B and Supplemental Dataset S1A). Among 20285 annotated protein-coding transcripts that have reasonable transcript levels (transcript per million (TPM) > 0.5 in RNA-seq), 18626 (92%) have translated ORFs identified. This indicates our approach to identifying translated ORFs is efficient and robust. In addition to annotated ORFs, there were 1329 unannotated uORFs translated from the 5′ UTR of annotated genes (Figure 3B, Supplemental Dataset S1B, Supplemental Dataset S2). Notably, since only approximately half

of the transcripts in ITAG3.2 (17684 out of 35768) have an annotated 5′ UTR and because RiboTaper can only identify ORFs in defined transcript ranges, the total number of uORFs in tomato root is clearly an underestimate.

Excitingly, we identified 354 unannotated translated ORFs from the newly assembled transcripts (Figure 3B, Supplemental Dataset S1C and Supplemental Dataset S3). These unannotated ORFs were found in different classes of transcripts, but none were detected in the negative control, class-s (Figure 3C). As expected, most of the newly discovered ORFs were relatively small; ~71% of them (250) encode proteins of less than 100 amino acids (Figure 3D). Due to their relatively small size, hereafter, we call them small ORFs (sORFs). The average lengths of the uORFs, sORFs, and annotated ORFs are 31, 95 and 422 amino acids, respectively. Among the 354 sORFs, 87 have a predicted signal peptide and are expected to be secreted proteins/peptides (Figure 3E and Supplemental Dataset S1D). To test if the sORFs and annotated ORFs have similar translational properties, we compared their translation efficiency (see the definition in the Methods) and found that they were statistically indistinguishable (Figure 3F). This result supports the newly identified sORFs are genuine protein-coding genes in the tomato genome.

The majority of the identified ORFs have high fractions of P-site signals mapped to the expected reading frame (Supplemental Figure S4). Visualizing the profiles of individual transcripts confirmed that both the sORFs and numerous annotated ORFs display strong 3-nt periodicity within the identified coding regions (Figure 3G-H). Therefore, by combining the high-quality Ribo-seq data with RiboTaper analysis, we not only validated many of the annotated gene models but also discovered new ORFs. These previously unannotated translated regions have been compiled and ready to be incorporated into the official tomato annotation (Supplemental Dataset S1A-C, Supplemental Datasets S2 and S3).

Evolutionarily conserved and Solanaceae-specific sORFs

Previously, we identified 27 sORFs in Arabidopsis by applying RiboTaper on Ribo-seq data (Hsu et al., 2016). Eight of the Arabidopsis sORFs have known tomato homologs. Our tomato root data showed that seven of the conserved sORFs were both transcribed and translated (Supplemental Figure S5A-D). Since Arabidopsis and tomato diverged approximately 100 million years ago (Ku et al., 2000), our data support that some sORFs are conserved across evolution.

If the newly-identified tomato sORFs encode proteins for conserved biological processes, we would expect them to be preserved during evolution. We performed tblastn using 157 single-exon sORFs that were 16–100 amino acids long on 10 diverse plant genomes, including a wild tomato (*S. pennellii*), potato (*Solanum tuberosum*, which belongs to the same family as tomato, the Solanaceae), four dicots in other families, two monocots, a lycophyte and a moss (Supplemental Figure S6). In total, we found 96 Solanaceae-specific sORFs, including 18 sORFs unique to tomato and 78 sORFs shared by tomato and either wild tomato or potato. Out of 157 sORFs analyzed, 139 of them have homologs in at least one other plant genome. Some of the sORFs are highly conserved across these 10 genomes (Supplemental Figure S6), suggesting the functional significance of these sORFs throughout evolution. Importantly, the conserved patterns among the homologs correlate well with their phylogenic relationships, indicating that these sORF homologs are unlikely to be false positives that randomly occurred in the blast search. While some sORFs are widely conserved, 96 sORFs are unique to Solanaceae, highlighting our approach to study translatomes directly in tomato revealed translational events that was impossible to learn by studying Arabidopsis alone. Taken together, our results reveal both evolutionarily conserved and Solanaceae-specific sORFs.

Some sORFs and uORFs generate stable proteins *in planta*

To evaluate whether the previously unknown ORFs, including sORFs and uORFs, accumulate stable proteins *in planta* and to validate our Ribo-seq results, we performed a

proteogenomic analysis (Walley and Briggs, 2015) to identify “novel” peptides arising from these unannotated ORFs. Because the sORFs and uORFs are quite small, their protein products do not always generate peptides with ideal size and/or mass-to-charge ratios that are suitable for detection by mass spectrometry (MS). To increase the diversity of peptides for MS analysis, we extracted proteins from the roots and shoots of tomato seedlings and digested the proteins into peptides using trypsin or GluC, independently, prior to two-dimensional liquid chromatography-tandem mass spectrometry (2D-LC-MS/MS). As the sORFs and uORFs are currently missing from the tomato annotation, we created a custom protein database (Supplemental Dataset S4) derived from our Ribo-seq data to assist in identifying these unannotated proteins. In addition, we used our custom protein database to search publicly available proteomic data from the tomato fruit (ProteomeXchange PXD004887) and pericarp (ProteomeXchange PXD004947) (Mata et al., 2017; Szymanski et al., 2017). In total, we identified 12172 proteins, including 29 sORFs and 30 uORFs, with at least one unique peptide from these six proteomic datasets (Figure 3B, Supplemental Dataset S1E-F, Supplemental Dataset S5A-C). The MS detection rates (at least one unique peptide) for sORFs below 100 amino acids, 100–200 amino acids, and higher than 200 amino acids in our data are 4.8%, 16.3% and 35.3%, respectively, suggesting proteins with a larger size have better chances to be detected by MS. Despite the limitations of MS in small protein identification, our results support that some uORFs and sORFs accumulate stable proteins *in planta*.

Ribo-seq fine-tunes and improves genome annotation

Comparing the RiboTaper output and the annotated gene models, we found cases in which the translated ORFs were dramatically different from the predicted gene models. For example, translation may occur in a different reading frame or at a distinct region on the transcript (Figure 3I, Supplemental Figure S7A-F). Thus, Ribo-seq provides a high-throughput experimental approach to validate and improve genome annotation. Furthermore, in several

cases, using visual inspection, we found regions that appear to contain a short ORF that overlaps with the long annotated ORF but uses a different reading frame (e.g., Figure 3J). These overlapping ORFs are similar to non-upstream coding ORFs identified in human genome (Michel et al., 2012), and their functional importance is still unknown.

The translation start sites in the genome annotation are typically defined computationally, and often the most upstream AUG is predicted to be the start codon. Unexpectedly, in 64 genes, the RiboTaper-defined translation start sites were actually upstream of the annotated start sites (e.g., Figure 4A and Supplemental Dataset S1G). In contrast, some ORFs appeared to use start sites downstream of the annotated start sites (e.g., Figure 4B, Supplemental Figure S5B). Currently, ITAG3.2 contains only one isoform per gene, and hence only one transcription start site is predicted per gene. It is possible that in some cases, translation starts downstream of the annotated site because transcription initiates downstream of the annotated transcription start site. Nonetheless, it appears that the most upstream AUG is not always used as the translation start site.

Non-AUG translation initiation has been discovered in animals and plants (Simpson et al., 2010; Laing et al., 2015; Kearse and Wilusz, 2017; Spealman et al., 2018). Twelve evolutionarily conserved noncanonical translation starts upstream of the most likely AUG have been predicted in Arabidopsis (Simpson et al., 2010), and we previously showed that at least one of them, in *AT3G10985*, has high Ribo-seq coverage using a CUG codon (Hsu et al., 2016). The profile of the tomato homolog of *AT3G10985* confirmed the possible usage of the CUG start site (Figure 4C). Next, we identified tomato homologs of all twelve predicted noncanonical-start genes and systematically checked their Ribo-seq coverage upstream of the annotated AUG start sites. We selected genes that met the following criteria: 1) the Ribo-seq reads cover at least 7 in-frame P-site positions within the first 20 codons upstream of the AUG; 2) there is no stop codon within the first 20 codons upstream of the AUG. We found eight tomato genes that met the above criteria contain abundant reads upstream of the annotated AUG, suggesting that

285 they use non-AUG start sites (Figure 4D). Thus, despite the evolutionary distance between
286 Arabidopsis and tomato, the usage of noncanonical translation initiation remains conserved in
287 these homologs.
288
289 **uORFs regulate translation efficiency**
290 Using RiboTaper, we identified 1329 translated uORFs based on their significant 3-nt
291 periodicity (Figure 3B, Supplemental Dataset S1B, Supplemental Dataset S2). These uORFs
292 included previously predicted conserved uORFs in the tomato *SAC51* homolog (Figure 5A)
293 (Imai et al., 2006), as well as previously unknown uORFs in numerous coding genes (e.g.
294 Figure 5B). Manual inspection of these transcripts suggested that the high stringency of
295 RiboTaper might miss uORFs with lower periodicity, overlapping uORFs and non-AUG-start
296 uORFs. For example, the second of the three uORFs in the tomato *SAC51* transcript (Figure 5A)
297 was not identified as coding by RiboTaper, presumably due to the imperfect periodicity in this
298 area. Nevertheless, those identified are high-confidence translated uORFs.
299 Global analyses have reported that translated uORFs repress the translation of their
300 downstream main ORFs (Liu et al., 2013; Lei et al., 2015; Chew et al., 2016; Johnstone et al.,
301 2016). Consistent with these reports, we found that globally, transcripts containing uORFs have
302 lower translation efficiency than those without uORFs (Figure 5C). In addition, more uORFs in a
303 transcript correlate with stronger translational repression (Figure 5C). To investigate which
304 physiological pathways might be regulated by uORFs, we checked the gene ontology (GO)
305 terms of the uORF-containing genes. Intriguingly, uORF-containing genes were enriched for
306 protein kinases and phosphatases, as well as signal transduction (Figure 5D). This is similar to
307 previous prediction in Arabidopsis (Kim et al., 2007), except transcription factors are not
308 enriched in our data. Our results imply that a substantial portion of the protein
309 phosphorylation/dephosphorylation and signal transduction pathways in tomato are likely
310 translationally regulated through uORFs.

311 Translation start sites have a well-defined Kozak consensus sequence in different
312 organisms (Kozak, 1987; Lütcke et al., 1987). For example, the conserved nucleotides at
313 positions -3 and +4 of the Kozak sequence in plants are purines (A/G) and G, respectively
314 (Lütcke et al., 1987). As expected, we observed this conserved pattern among the annotated
315 ORFs (Figure 5E). Next, we examined the Kozak consensus sequences of the translated
316 uORFs and their downstream main ORFs. Whereas the downstream main ORFs also favor the
317 conserved nucleotides at -3 and +4 of the Kozak sequence, this pattern is missing in the uORFs
318 (Figure 5E and 5F). Similar results were observed in the Kozak sequences of the uORFs and
319 downstream main ORFs in Arabidopsis (Liu et al., 2013). The poorly conserved Kozak
320 sequences might allow for more leaky scanning, a phenomenon that a weak initiation context is
321 sometimes skipped by ribosome during translation initiation, so the downstream main ORFs
322 could still have some chances to be translated.
323
324 **Regulation of gene expression by microRNAs**
325 MicroRNAs regulate gene expression through mRNA cleavage and translational
326 repression (Yu et al., 2017; Li et al., 2018). The roles of microRNAs in tomato are less well
327 understood than in Arabidopsis. We first predicted 6312 microRNA target genes in tomato
328 (Supplemental Dataset S1H-I) using psRNATarget (Dai et al., 2018). Next, we compared their
329 RNA-seq and Ribo-seq levels and the translation efficiency of the microRNA targets and other
330 coding genes globally. The transcript levels of the miRNA targets were slightly but significantly
331 reduced, consistent with the possibility that microRNAs regulate gene expression through
332 mRNA cleavage (Figure 6A). In addition, both the Ribo-seq levels and translation efficiency of
333 the microRNA target genes were reduced (Figure 6B and 6C), consistent with prior observations
334 of translational repression mediated by microRNAs (Faghihi and Wahlestedt, 2009). Thus, our
335 results suggest that globally, microRNAs regulate gene expression at both the transcript and
336 translational levels in tomato.

DISCUSSION

Most of the plant research on mRNA translation was performed in Arabidopsis, and the knowledge has been transferred into several crops to improve crop performance. However, on a genome-wide level, it is unclear how well the Arabidopsis translatome compares to other species. In this study, we combined *de novo* transcriptome assembly and ribosome profiling to study the tomato translatome. We found that despite Arabidopsis and tomato diverging over 100 million years ago, many translational features are well conserved. Overall, we observed shared features between our Arabidopsis and tomato Ribo-seq data, including the most abundant ribosome footprint size and the inferred P-site within ribosome footprints. We found that previously unannotated translation events, such as uORFs and sORFs, are also widespread in tomato. In addition, we observed that usage of non-AUG translation start sites is shared between Arabidopsis and tomato. Finally, translational regulatory mechanisms, including uORFs on their downstream main ORFs and miRNAs on their target genes, are also well-conserved in these two species.

Interestingly, we discovered 96 previously unknown sORFs only present in Solanaceae, including 78 shared by tomato and either wild tomato or potato, and 18 sORFs uniquely found in tomato. These family-specific sORFs may provide functions unique to Solanaceae. The idea of family-specific regulatory molecules was proposed based on systemin, the first peptide hormone identified in plants. Systemin is only present in Solaneae, a subtribe of the Solanaceae (Pearce et al., 1991; Constabel et al., 1998). Such family- or sub-family-specific regulatory molecules may evolve during evolution for a specific lineage of plants. Even species-specific sORFs have been proposed to be important (Andrews and Rothnagel, 2014). The functions of the widely conserved and Solanaceae-specific sORFs require further studies.

Peptide signaling is crucial for cell-cell communication in numerous aspects of plant development and stress responses (Tavormina et al., 2015; Hsu and Benfey, 2018). We found

87 sORFs that encode potential secreted peptides. However, as about 50% of secreted proteins in plants lack a well-defined signal peptide (Agrawal et al., 2010), some sORFs without a predicted signal peptide may still be secreted. In addition, sORF products without a signal peptide have been found to play an important role in a wide range of physiological processes in plants, such as vegetative and reproductive development, siRNA biogenesis, and stress tolerance (Casson, 2002; Blanvillain et al., 2011; Ikeuchi et al., 2011; Valdivia et al., 2012; De Coninck et al., 2013). Therefore, the identification of sORFs using ribosome profiling facilitates potential applications of these peptides in improving crop performance.

Several studies have illustrated the power of altering mRNA translation via uORFs to improve agriculture (Sagor et al., 2016; Xu et al., 2017b; Zhang et al., 2018). For example, engineering rice that specifically induces defense proteins when a uORF is repressed by pathogen attack enables immediate plant resistance without compromising plant growth in the absence of pathogens (Xu et al., 2017b). The identification of translated ORFs provides new possibilities to fine-tune the synthesis of proteins involved in diverse physiological pathways. Notably, the number of uORFs in tomato is still an underestimate. Approximately half of the tomato genes still lack annotated 5′ UTRs, and RiboTaper only searches for potential translated ORFs in defined transcript regions. Thus, uORFs could be an even more widespread mechanism to control translation in tomato. Future studies using a combination of CAGE-seq or PEAT-seq with the long-read sequencing could facilitate defining the 5′ UTRs associated with specific isoforms (Ozsolak and Milos, 2011) and enable identification of missing uORFs.

Ribo-seq has been integrated into proteomic research to achieve deeper proteome coverage (Menschaert et al., 2013; Crappe et al., 2014; Van Damme et al., 2014; Calviello et al., 2016). Unlike DNA or RNA molecules, which can be sequenced using genomic technologies, proteins are typically identified by matching MS spectra to theoretical spectra from candidate peptides in a reference protein database. Before ribosome profiling became available, to include potential protein sequences, the conventional proteogenomics approach exploited either three-

frame-translation using transcriptome data or six-frame-translation using genomic sequences (Walley and Briggs, 2015; Ruggles et al., 2017). Integrating Ribo-seq data into the construction of protein databases for proteogenomic studies has several advantages: 1) Ribo-seq discovers unannotated translation events and thus enables the identification of unknown proteins that were previously missed in the annotation; 2) compared with three-frame or six-frame translation, Ribo-seq reduces the search space and false positives. Therefore, our custom protein database, built based on the Ribo-seq data, may aid in proteomic research in tomato.

CONCLUSIONS

In summary, our approach combining transcriptome assembly and ribosome profiling enabled identification of translated ORFs genome-wide in tomato and revealed conserved and unique translational features across evolution. Our results not only provide valuable information to the plant community but also present a practical strategy to study translatomes in other less-well annotated organisms.

MATERIALS AND METHODS

Plant materials and preparation of lysates for RNA-seq and Ribo-seq

Tomato seeds (*Solanum lycopersicum*, 'Heinz 1706' cultivar) were obtained from the C.M. Rick Tomato Genetics Resource Center (Accession: LA4345) and bulked. For each replicate, ~300 tomato seeds were surface-sterilized in 70% (v/v) ethanol 5 min followed by bleach solution (2.4% (v/v) NaHClO, 0.3% (v/v) Tween-20) for 30 min with shaking. The seeds were then washed with sterile water 5 times. Next, the seeds were stratified on 1x Murashige and Skoog media (4.3 g/L Murashige and Skoog salt, 1% (w/v) sucrose, 0.5 g/L MES, pH 5.7, 1% (w/v) agar), and kept at 22°C in the dark for 3 days before being grown under 16-h light/8-h dark conditions at 22°C for 4 days. Seedlings that germinated at approximately the same time and of similar size were selected for the experiments. Roots (~3 cm from the tip) from ~180

plants were harvested at ZT 3 (3 h after lights on) in batches and immediately frozen in liquid nitrogen. The frozen tissues were pooled and pulverized in liquid nitrogen using a mortar and pestle. Approximately 0.4 g of tissue powder was resuspended in 1.2 mL lysis buffer (100 mM Tris·HCl (pH 8), 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 DTT, 100 µg/mL cycloheximide, and 10 unit/mL DNase I (Epicenter D9905K)) as described in Hsu et al. (2016). After incubation on ice with gentle shaking for 10 min, the lysate was spun at 4°C at 20,000 *g* for 10 min. The supernatant was transferred to a new tube and divided into 100-µL aliquots. The aliquoted lysates were flash frozen in liquid nitrogen and stored at -80°C until processing.

RNA purification and RNA-seq library construction

For RNA-seq samples, 10 µL 10% (w/v) SDS was added to the 100-µL lysate aliquots described above. RNA greater than 200 nt was extracted using a Zymo RNA Clean & Concentrator kit (Zymo Research R1017). The obtained RNA was checked with a Bioanalyzer (Agilent) RNA pico chip to access the RNA integrity, and a RIN value ranging from 9.2 to 9.4 was obtained for each replicate. Ribosomal RNAs (rRNAs) were depleted using a RiboZero Plant Leaf kit (Illumina MRZPL1224). Next, 100 ng of the rRNA-depleted RNA was used as the starting material, fragmented to ~200 nt based on the RIN reported by the Bioanalyzer, and processed using an NEBNext Ultra Directional RNA Library Prep Kit (NEB E7420S) to create strand-specific libraries. The libraries were barcoded and enriched using 11 cycles of PCR amplification. The libraries were brought to equal molarity, pooled and sequenced on one lane of a Hi-Seq 4000 using PE-100 sequencing.

Ribosome footprinting and Ribo-seq library construction

The Ribo-seq samples were prepared based on Hsu et al. 2016 (Hsu et al., 2016) with modifications described as follows, which optimize the method for tomato. Briefly, the RNA

441 concentration of each lysate was first determined using a Qubit RNA HS assay (Invitrogen
442 Q32852) using a 10-fold dilution. Next, 100 µL of the lysate described above was treated with
443 100 units of nuclease (provided in the TruSeq Mammalian Ribo Profile Kit, Illumina
444 RPHMR12126) per 40 µg of RNA with gentle shaking at room temperature for 1 h. The
445 nuclease reaction was stopped by immediately transferring to ice and adding 15 µL of
446 SUPERase-IN (Invitrogen AM2696). The ribosomes were isolated using illustra MicroSpin S-
447 400 HR columns (GE Healthcare 27514001). RNA greater than 17 nt was purified first (Zymo
448 Research R1017), and then RNA smaller than 200 nt was enriched (Zymo Research R1015).
449 Next, the rRNAs were depleted using a RiboZero Plant Leaf kit (Illumina MRZPL1224). The
450 rRNA-depleted RNA was then separated via 15% (w/v) TBE-urea PAGE (Invitrogen
451 EC68852BOX), and gel slices ranging from 28 to 30 nt were excised. Ribosome footprints were
452 recovered from the excised gel slices using the overnight elution method, and the sequencing
453 libraries were constructed according to the TruSeq Mammalian Ribo Profile Kit manual. The
454 final libraries were amplified via 9 cycles of PCR. The libraries were brought to equal molarity,
455 pooled and sequenced on two lanes of a Hi-Seq 4000 using SE-50 sequencing.
456
457 **RNA-seq and Ribo-seq data analysis**
458 The raw RNA-seq and Ribo-seq data and detailed mapping parameters have been
459 deposited in the Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo) under
460 accession no. GSE124962. The tomato reference genome sequence and annotation files used
461 in this study were downloaded from the Sol Genomics Network (Fernandez-Pozo et al., 2015).
462 The adaptor sequence AGATCGGAAGAGCACACGTCT was first removed from the Ribo-seq
463 data using FASTX_clipper v0.0.14 (http://hannonlab.cshl.edu/fastx-toolkit). For both RNA-seq
464 and Ribo-seq, the rRNA, tRNA, snRNA, snoRNA and repeat sequences were removed using
465 Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012). The rRNA, tRNA, snRNA, and snoRNA
466 sequences were extracted from the SL2.5 genome assembly with the ITAG2.4 annotation

(Fernandez-Pozo et al., 2015), and the repeat sequences were extracted from SL3.0 genome assembly with the ITAG3.2 annotation. After these contaminating sequences were removed using Bowtie2, the pre-processed RNA-seq and Ribo-seq files were used to calculate the read distribution in different gene features (Figure 2B) using the featureCounts function of the Subread package v1.5.3 (Liao et al., 2014).

Next, the pre-processed RNA-seq and Ribo-seq reads were mapped to the tomato reference genome sequence SL3.0 with the ITAG3.2 annotation using the STAR v2.6.0.c (Dobin et al., 2013). The reference-guided *de novo* assembly of the mapped RNA-seq reads was performed with stringtie v1.3.3b (Pertea et al., 2015), and the newly assembled gtf files were compared to ITAG3.2 using gffcompare v0.10.1 (Pertea et al., 2016). The i, x, y, o, u, s classes of new transcripts (see Figure 3A for details) and their descriptions were extracted from the gffcompare output gtf and concatenated with ITAG3.2. This combined gtf (referred as "Tomato_Root_ixyous+ITAG3.2.gtf"; submitted to GEO as a processed file within GSE124962) was used to map the RNA-seq and Ribo-seq reads again with STAR. Notably, all six classes of uncharacterized transcripts in Tomato_Root_ixyous+ITAG3.2.gtf were assigned as ncRNAs, and this gtf was used for downstream RiboTaper analysis. The three biological replicates of the mapped bam files for RNA-seq were merged into one large bam file with SAMtools v1.8 (Li et al., 2009). The three mapped Ribo-seq bam files were also merged. The two merged bam files above were then used for ORF discovery with RiboTaper v1.3 (Calviello et al., 2016).

For RiboTaper analysis, the RiboTaper annotation files and the offset parameters (i.e., the inferred P-site position for each footprint length) were first obtained. The RiboTaper annotation files were generated using the create_annotations_files.bash function in the RiboTaper package using SL3.0 assembly and the Tomato_Root_ixyous+ITAG3.2.gtf. To obtain the offset parameters, the create_metaplots.bash and metag.R functions in the RiboTaper package were used to generate meta-gene plots. The offset parameters were identified through the meta-gene plots. For 24-, 25-, 26-, 27-, 28-nt footprints, the offset values

were 8, 9, 10, 11, and 12, respectively (Supplemental Figure S3). Next, we performed RiboTaper analysis using the RiboTaper annotation, offset parameters, and RNA-seq and Ribo-seq bam files. The coding sequences identified by RiboTaper from the newly assembled transcripts were extracted from the translated_ORFs_filtered_sorted.bed file and integrated with Tomato_Root_ixyous+ITAG3.2.gtf to generate Supplemental Dataset S2_uORF.gtf and Supplemental Dataset S3_sORF.gtf.

We then mapped the Ribo-seq and RNA-seq data again to the CDS ranges with STAR, and the transcripts per million (TPM) for the CDS of each transcript was quantified via RSEM v1.3.0 (Li and Dewey, 2011). The formula to calculate translation efficiency is “TE = (the TPM_{CDS} of Ribo-seq)/(the TPM_{CDS} of RNA-seq)”. To avoid inflation due to a small denominator, only genes with an RNA-seq TPM greater than 0.5 were used in the statistical analysis of translation efficiency. The plotting of 3-nt periodicity of the Ribo-seq and coverage of RNA-seq was generated by incorporating the plot function in R v3.4.3 (R Core Team (2013), 2017) with functions from GenomicRanges v1.30.3, GenomicFeatures v1.30.3, and GenomicAlignments v1.14.2 libraries (Lawrence et al., 2013) to read in the gtf file and RNA-seq bam file. The merged RNA-seq bam file from STAR and the processed "P_sites_all" file from RiboTaper were used to plot the RNA-seq coverage and P-sites of Ribo-seq, respectively. The Linux command line code to preprocess the “P_sites_all” file before used for plotting was “cut -f 1,3,6 P_sites_all | sort | uniq -c | sed -r 's/^(*[^]+) +/\1\t/' > name_output_file”. For plotting the CUG/non-AUG start gene, the CDS range of the gene in the gtf file was manually modified before plotting.

Statistical analysis

The statistical analysis in the paper was performed in R (R Core Team (2013), 2017). The chisq.test and ks.test functions of the "stats" package in R were used for the Chi-squared analysis and the Kolmogorov-Smirnov test, respectively. The Pearson and Spearman correlation coefficients were calculated using the "cor" function. Pairwise comparisons were

519 performed using the "corrplot" function in the corrplot v0.84 package (Wei, 2013). The empirical
520 cumulative probabilities of translation efficiency were calculated using the "ecdf" function (in the
521 “stats” package) and plotted with the base R plot function.
522
523 **Protein extraction and digestion**
524 Roots (~3 cm near the tip) and shoots (shoot tip including ~1 cm hypocotyl) of four-day-
525 old tomato seedlings were harvested at ZT3 (3 h after light on). The proteomics experiments
526 were carried out based established methods as follows (Castellana et al., 2014; Song et al.,
527 2018b; Song et al., 2018a). Five volumes (v:w) of Tris buffered phenol pH 8 was added to 150
528 mg of ground tissue, vortexed 1 min, then mixed with 5 volumes (buffer:tissue, v:w) of extraction
529 buffer (50 mM Tris pH 7.5, 1 mM EDTA pH 8, 0.9 M sucrose), and then centrifuge at 13,000 *g*,
530 for 10 min at 4°C. The phenol phase was transferred to a new tube and a second phenol
531 extraction was performed on the aqueous phase. The two phenol phase extractions were
532 combined and 5 volume of prechilled methanol with 0.1 M ammonium acetate was added. This
533 was mixed well and keep at -80°C for 1 h prior to centrifugation at 4,500 *g*, for 10 min at 4°C.
534 Precipitation with 0.1 M ammonium acetate in methanol was performed twice with incubation at
535 -20°C for 30 min. The sample was resuspended in 70% (v/v) methanol at kept at -20°C for 30
536 min prior to centrifuging at 4,500 *g*, for 10 min at 4°C. The supernatant was discarded and the
537 pellet was placed in a vacuum concentrator till near dry. Two volumes (buffer:pellet, v:v) of
538 protein digestion buffer (8 M urea, 50 mM Tris pH 7, 5 mM Tris(2-carboxyethyl)phosphine
539 hydrochloride (TCEP)) was added to the pellet. The samples were then probe sonicated to aid
540 in resuspension of the pellet. The protein concentration was then determined using the Bradford
541 assay (Thermo Scientific).
542 The solubilized protein (~ 1 mg) was added to an Amicon Ultracel – 30K centrifugal filter
543 (Cat # UFC803008) and centrifuged at 4,000 *g* for 20–40 min. This step was repeated once.
544 Then 4 mL of urea solution with 2 mM TCEP was added to the filter unit and centrifuged at

4,000 *g* for 20–40 min. Next, 2 mL iodoacetamide (IAM) solution (50 mM IAM in 8 M urea) was added and incubated without mixing at room temperature for 30 min in the dark prior to centrifuging at 4,000 *g* for 20–40 min. Two mL of urea solution was added to the filter unit, which was then centrifuged at 4,000 *g* for 20–40 min. This step was repeated once. Two mL of 0.05 M NH_4HCO_3 was added to the filter unit and centrifuged at 4,000 *g* for 20–40 min. This step was repeated once. Then 2 mL 0.05M NH_4HCO_3 with trypsin (enzyme to protein ratio 1:100) or GluC (enzyme to protein ratio 1:20) was added. Samples were incubated at 37°C overnight. Undigested protein was estimated using Bradford assays then trypsin (1 µg/µL) was added to a ratio of 1:100 and an equal volume of Lys-C (0.1 µg/uL) were added to the trypsin/Lys-C digested sample and GluC was added at a ratio of 1:20 to the sample digested with GluC. The digests were incubated for an additional 4 h at 37°C. The filter unit was added to a new collection tube and centrifuged at 4,000 *g* for 20–40 min. One mL 0.05M NH_4HCO_3 was added and centrifuged at 4,000 *g* for 20–40 min. The samples were acidified to pH 2–3 with 99% (v/v) formic acid and centrifuged at 21,000 *g* for 20 min. Finally, samples were desalted using 50 mg Sep-Pak C18 cartridges (Waters). Eluted peptides were dried using a vacuum centrifuge (Thermo) and resuspended in 0.1% (v/v) formic acid. Peptide amount was quantified using the Pierce BCA Protein assay kit.

LC/MS-MS

An Agilent 1260 quaternary HPLC was used to deliver a flow rate of ~600 nL min-1 via a splitter. All columns were packed in house using a Next Advance pressure cell and the nanospray tips were fabricated using fused silica capillary that was pulled to a sharp tip using a laser puller (Sutter P-2000). 25 µg of peptides were loaded unto 20 cm capillary columns packed with 5 µM Zorbax SB-C18 (Agilent), which was connected using a zero dead volume 1 µm filter (Upchurch, M548) to a 5 cm long strong cation exchange (SCX) column packed with 5 µm PolySulfoethyl (PolyLC). The SCX column was then connected to a 20 cm nanospray tip

packed with 2.5 µM C18 (Waters). The 3 sections were joined and mounted on a custom electrospray source for on-line nested peptide elution. A new set of columns was used for every sample. Peptides were eluted from the loading column unto the SCX column using a 0 to 80% acetonitrile (ACN) gradient over 60 min. Peptides were then fractionated from the SCX column using a series of ammonium acetate salt steps as following:10, 30, 32.5, 35, 37.5, 40, 42.5, 45, 50, 55, 65, 75, 85, 90, 95, 100, 150, and 1000 mM. For these analyses, buffers A (99.9% H_2O, 0.1% formic acid), B (99.9% ACN, 0.1% formic acid), C (100 mM ammonium acetate, 2% formic acid), and D (1 M ammonium acetate, 2% formic acid) were utilized. For each salt step, a 150-minute gradient program comprised of a 0–5 minute increase to the specified ammonium acetate concentration (using buffers C or D), 5–10 min hold, 10–14 min at 100% buffer A, 15–120 min 5–35% buffer B, 120–140 min 35–80% buffer B, 140–145 min 80% buffer B, and 145–150 min buffer A was employed.

Eluted peptides were analyzed using a Thermo Scientific Q-Exactive Plus high-resolution quadrupole Orbitrap mass spectrometer, which was directly coupled to the HPLC. Data dependent acquisition was obtained using Xcalibur 4.0 software in positive ion mode with a spray voltage of 2.00 kV and a capillary temperature of 275 °C and an RF of 60. MS1 spectra were measured at a resolution of 70,000, an automatic gain control (AGC) of 3e6 with a maximum ion time of 100 ms and a mass range of 400-2000 m/z. Up to 15 MS2 were triggered at a resolution of 17,500. An AGC of 1e5 with a maximum ion time of 50 ms, an isolation window of 1.5 m/z, and a normalized collision energy of 28. Charge exclusion was set to unassigned, 1, 5–8, and >8. MS1 that triggered MS2 scans were dynamically excluded for 25s.

Database search and FDR filtering

The raw data were analyzed using MaxQuant version 1.6.3.3 (Tyanova et al., 2016). A customized protein database containing 22513 proteins (Supplemental Dataset S4) was generated from the RiboTaper output file “ORFs_max_filt.” Spectra were searched against the

customized protein database 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. Digestion parameters were set to “specific” and “Trypsin/P;LysC” or “GluC”. Up to two missed cleavages were allowed. A false discovery rate less than 0.01 and protein identification level was required. The “second peptide” option was used to identify co-fragmented peptides. The “match between runs” feature of MaxQuant was not utilized. Raw data files and MaxQuant Search results have been deposited in the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository: https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp with dataset identifier: MSV000083363.

Prediction of the subcellular localization of sORFs

A fasta file containing the sORF amino acid sequences was uploaded to the TargetP website (Emanuelsson et al., 2000). We selected "Plant" as the organism group and ">0.90" as the specificity cutoff and then submitted for analysis.

Evolutionary analysis

The "tblastn" function for BLAST v2.7.1 (OS Linux_x86_64)(Camacho et al., 2009) was used for the homology search. Because several plant genomes still lack exon-intron junction information in their annotations, we only selected single-exon tomato sORFs that encoded 16–100 amino acid residues for this analysis, and the reference genomes (Athaliana_167_TAIR9.fa, Atrichopoda_291_v1.0.fa, Csinensis_154_v1.fa, Mtruncatula_285_Mt4.0.fa, Osativa_323_v7.0.fa, Ppatens_318_v3.fa, S_lycopersicum_chromosomes.3.00.fa, Sitalica_312_v2.fa, Smoellendorffii_91_v1.fa, Stuberosum_448_v4.03.fa) were downloaded from Phytozome v12 (Goodstein et al., 2012).The fa (fasta) files for each genome were used to generate blast databases with the following code: “makeblastdb -in genome.fa -parse_seqids -

dbtype nucl", where genome.fa was replaced with the fasta file for each genome. Next, the code "tblastn -query input.fa -db species_database -out species_blast_result.txt -evalue 0.001 -outfmt '6 qseqid sseqid length qlen qstart qend sstart send pident gapopen mismatch evalue bitscore' -num_threads 10" was used to search for sequence homologs in the target genomes. The names of "species_database" and "species_blast_result.txt" were changed correspondingly. The final heatmap for amino acid identity was plotted in R using the pheatmap v1.0.10 (Kolde, 2015) and RColorBrewer v1.1.2 libraries (Neuwirth, 2014).

miRNA target identification

The tomato miRNA sequences were extracted from Kaur et al. (2017) and Liu et al. (2017). Next, we used psRNATarget (Dai et al., 2018) against ITAG3.2 mRNA sequences to identify potential miRNA targets. We used "Schema V2 (2017 release)" (Dai et al., 2018) and selected "calculate target accessibility" as the analysis parameters.

GO term analysis

agriGO v2.0 (Tian et al., 2017) was used for the GO analysis of uORF-containing genes.

ACCESSION NUMBERS

- The raw RNA-seq and Ribo-seq data have been deposited in the Gene Expression Omnibus (GEO) database under accession no. GSE124962.
- Proteomics raw data files and MaxQuant Search results have been deposited at the MassIVE repository with dataset identifier: MSV000083363.

SUPPLEMENTAL DATA

Supplemental Figure S1. Correlation between RNA-seq and Ribo-seq data.

Supplemental Figure S2. Meta-gene analysis and inference of the P-site for ribosome footprints of different lengths.

Supplemental Figure S3. Summary of the inferred P-site position for each footprint length.

Supplemental Figure S4. Fractions of in-frame P-sites for different groups of translated ORFs.

Supplemental Figure S5. Translation of tomato homologs of Arabidopsis sORFs.

Supplemental Figure S6. Evolutionary conservation of sORFs.

Supplemental Figure S7. Examples of conflicts between annotated gene models and translational profiles.

Supplemental Dataset S1_lists of ORFs_proteomics_miRNAs. xlsx, spreadsheets (A) to (I).

(A): ORF_ccds (annotated ORFs) from RiboTaper output "ORF_max_filt"

(B): uORFs from RiboTaper output "ORF_max_filt"

(C): sORFs from RiboTaper output "ORF_max_filt"

(D): TargetP results for sORFs

(E): sORF MassSpec spectra

(F): uORF MassSpec spectra

(G): 64 ORFs using an upstream start rather than annotated start identified by RiboTaper

(H): miRNAs used for psRNATarget prediction

(I): predicted miRNA-targets by psRNATarget

Supplemental Dataset S2. uORF.gtf (gtf for uORFs)

Supplemental Dataset S3. sORF.gtf (gtf for sORFs)

Supplemental Dataset S4_amino_acid_sequences_for_translated_ORFs. fa (amino acid sequences for all translated ORFs identified by RiboTaper in this study)

Supplemental Dataset S5_Proteogenomics. xlsx: Proteogenomics, spreadsheets (A) to (C).

(A) MaxQuant_proteinGroups

(B) MaxQuant_peptides

(C) MaxQuant_modificationSpecificPeptides

ACKNOWLEDGEMENTS

We thank Philip N. Benfey at Duke University for generous support to help initiate this project. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by an NIH S10 OD018174 Instrumentation Grant.

FIGURE LEGENDS

Figure 1: Experimental and data analysis procedures for ribosome profiling in tomato roots. A, Four-day-old tomato seedling roots (approximately 3 cm from the tip) were used in this study. B, Experimental workflow for RNA-seq and Ribo-seq and the schematics of their expected read distributions in the three reading frames. This figure was adapted from Hsu et al. (2016). C, Data analysis workflow for reference-guided de novo transcriptome assembly and ORF discovery using RiboTaper.

Figure 2: Ribosome footprints are enriched in coding sequences and display strong 3-nt periodicity.

A, The distribution of read length of the ribosome footprints. B, The distribution of the Ribo-seq and RNA-seq reads in different genomic features annotated in ITAG3.2. C, Meta-gene analysis of the 28-nt ribosome footprints near the annotated translation start and stop sites defined by ITAG3.2. The red, blue and green bars represent reads mapped to the first (expected), second and third reading frames, respectively. The majority of footprints were mapped to the CDS in the expected reading frame (85.5% in frame). For each read, only the first nucleotide in the P-site was plotted (for details, see Supplemental Figure S2 and Supplemental Figure S3). The A-site

700 (aminoacyl-tRNA entry site), P-site (peptidyl-tRNA formation site) and E-site (uncharged tRNA
701 exit site) within the ribosomes at translation initiation and termination, and the inferred P-site
702 (13^{th}–15^{th} nts) and A-site (16^{th}–18^{th} nts) are illustrated. The original meta-plots generated by
703 RiboTaper for all footprint lengths are shown in Supplemental Figure 2.
704
705 **Figure 3: The translational landscape of the tomato root.**
706 A, Classes of newly assembled transcripts identified by stringtie and gffcompare and used in
707 downstream ORF identification. This figure was adapted from the gffcompare website (Pertea et
708 al., 2016). B, Summary of translated ORFs identified by RiboTaper in our dataset and peptide
709 support from mass spectrometry data. The uORFs and annotated ORFs were identified from the
710 5′ UTRs and expected CDSs of annotated protein-coding genes in ITAG3.2, respectively. The
711 previously unknown ORFs were identified from the newly assembled transcripts. The bottom
712 row indicates the number of proteins in each category supported by mass spectrometry
713 datasets, either from our own proteomic analysis or searches against publicly available data. C,
714 Summary of newly assembled transcripts and ORFs identified in each class of newly assembled
715 transcripts. The total number of transcripts, number of transcripts identified as translated and
716 the total number of translated ORFs are listed. D, Size distribution of each class of sORFs,
717 uORFs and annotated ORFs (aORFs). E, Predicted subcellular localization of proteins encoded
718 by the sORFs. The prediction was performed using TargetP (Emanuelsson et al., 2000) with
719 specificity 0.9 as a cutoff. F, Translation efficiency of sORFs compared with annotated ORFs.
720 Only the coding regions were used to compute the TPM and translation efficiency of each
721 transcript. For the x-axis, only the range from 0 to 3 (arbitrary unit) is shown. A two-sample
722 Kolmogorov-Smirnov test was used to determine statistical significance. G to J, RNA-seq
723 coverage and Ribo-seq periodicity in different genes: an intergenic sORF on chromosome 4 (G);
724 an annotated coding gene that has good support from the Ribo-seq data for the predicted gene
725 model (H); a mis-annotated ORF (I), note the Ribo-seq reads do not match the CDS in the gene

model and a different reading frame is used; a transcript with a potentially overlapping ORF within the annotated ORF (J). In G to J, the x-axis indicates the genomic coordinate of the gene. The y-axis shows the normalized read count (counts per hundred million reads). Ribo-seq reads are shown by plotting the first nucleotide of their P-sites (denoted as the P-site signals). The black and gray dashed vertical lines mark the predicted translation start and stop sites, respectively. The red, blue and green lines in the Ribo-seq plot indicate the P-site signals mapped to the first (expected) reading frame and the second and third reading frames, respectively; the grey lines indicate the P-site signals mapped to outside of the annotated or identified coding regions. Hence, a higher ratio of red means better 3-nt periodicity. For the gene model beneath the Ribo-seq data, the gray, black and white areas indicate the 5′ UTR, CDS and 3′ UTR, respectively. In J, the yellow box above the gene model indicates the region with a potential ORF overlapping with the annotated ORF.

Figure 4: Upstream/downstream start sites and non-AUG start sites.

A and B, Examples of the usage of an upstream start site (A) or a downstream start site (B). The gene model and data presentation are the same as those described in the legend of Figure 3. The blue triangle marks the location of the annotated translation start site. The orange triangle marks the location of the RiboTaper-identified translation start site. C, A tomato homolog of an Arabidopsis gene that was predicted to use an upstream CUG start site (orange triangle). Note the abundant in-frame P-site signals upstream of the annotated AUG start (blue triangle) in the 5′ UTR. D, Conservation of potential CUG/non-AUG start sites. The Arabidopsis gene ID, tomato gene ID, percent amino acid identity, and number of in-frame P-site positions with Ribo-seq reads within the first 20 codons upstream of the AUG in our tomato root data are shown.

Figure 5: uORFs repress translation efficiency of their downstream main ORFs and

contain less-pronounced Kozak sequences.

A and B, Profiles of genes containing conserved uORFs (A) or a previously uncharacterized uORF (B). The gene model and data presentation are the same as those described in the legend of Figure 3. The uORFs are labeled with yellow and orange boxes in the gene models. For the uORFs, the orange and green dashed vertical lines mark the translation start and stop sites, respectively. C, The translation efficiency (TE) of the main ORFs for transcripts containing a different number of translated uORFs. Only the coding regions were used to compute the TPM and translation efficiency of each transcript. The colored bars before the p-values indicate the pairs of data used to determine statistical significance. The p-values were determined with two-sample Kolmogorov-Smirnov tests. D, Selected non-redundant GO categories for genes containing one or more uORFs. E and F, Kozak sequences of annotated ORFs, uORFs, and uORF-associated main ORFs. The statistical significance in F was determined using Chi-squared tests.

Figure 6: Regulation of gene expression by microRNAs (miRNAs). A to C, Cumulative distributions of RNA-seq (A), Ribo-seq (B) and translation efficiency (TE; C) of miRNA targets and non-miRNA target genes. For the x-axis in A and B, only the range from 0 and 50 (TPM) is shown. Only the coding regions were used to compute the TPM and translation efficiency of each transcript. The p-values were determined with two-sample Kolmogorov-Smirnov tests.

773 **LITERATURE CITED**

774 **Agrawal GK, Jwa N-S, Lebrun M-H, Job D, Rakwal R** (2010) Plant secretome: Unlocking
775 secrets of the secreted proteins. PROTEOMICS **10**: 799–827

776 **Andreev DE, O'Connor PBF, Loughran G, Dmitriev SE, Baranov P V., Shatsky IN** (2017)
777 Insights into the mechanisms of eukaryotic translation gained with ribosome profiling.
778 Nucleic Acids Research **45**: 513–526

779 **Andrews SJ, Rothnagel JA** (2014) Emerging evidence for functional peptides encoded by
780 short open reading frames. Nature reviews Genetics **15**: 193–204

781 **von Arnim AG, Jia Q, Vaughn JN** (2013) Regulation of plant translation by upstream open
782 reading frames. Plant Science **214**: 1–12

783 **Basrai MA, Hieter P, Boeke JD** (1997) Small Open Reading Frames: Beautiful Needles in the
784 Haystack. Genome Research **7**: 768–771

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

789 **Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES,**
790 **Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al** (2014) Identification of small ORFs in
791 vertebrates using ribosome footprinting and evolutionary conservation. The EMBO Journal
792 **33**: 981–993

793 **Blanvillain R, Young B, Cai Y, Hecht V, Varoquaux F, Delorme V, Lancelin J-M, Delseny M,**
794 **Gallois P** (2011) The Arabidopsis peptide kiss of death is an inducer of programmed cell
795 death. The EMBO Journal **30**: 1173–1183

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

798 **Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS** (2012) High-

799 resolution view of the yeast meiotic program revealed by ribosome profiling. Science **335**:
800 552–7
801 **Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M,**
802 **Obermayer B, Ohler U** (2016) Detecting actively translated open reading frames in
803 ribosome profiling data. Nature Methods **13**: 165–170
804 **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL** (2009)
805 BLAST+: architecture and applications. BMC Bioinformatics **10**: 421
806 **Casson SA** (2002) The POLARIS Gene of Arabidopsis Encodes a Predicted Peptide Required
807 for Correct Root Growth and Leaf Vascular Patterning. The Plant Cell **14**: 1705–1721
808 **Castellana NE, Shen Z, He Y, Walley JW, Cassidy CJ, Briggs SP, Bafna V** (2014) An
809 automated proteogenomic method uses mass spectrometry to reveal novel genes in Zea
810 mays. Molecular & Cellular Proteomics **13**: 157–67
811 **Chew G-L, Pauli A, Schier AF** (2016) Conservation of uORF repressiveness and sequence
812 features in mouse, human and zebrafish. Nature Communications **7**: 11663
813 **Chotewutmontri P, Barkan A** (2016) Dynamics of Chloroplast Translation during Chloroplast
814 Differentiation in Maize. PLoS Genetics **12**: e1006106
815 **Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley I** (2015)
816 The use of duplex-specific nuclease in ribosome profiling and a user-friendly software
817 package for Ribo-seq data analysis. RNA **10**: 1731–45
818 **Claverie J** (1997) Computational methods for the identification of genes in vertebrate genomic
819 sequences. Human Molecular Genetics **6**: 1735–1744
820 **De Coninck B, Carron D, Tavormina P, Willem L, Craik DJ, Vos C, Thevissen K, Mathys J,**
821 **Cammue BPA, SM. D, et al** (2013) Mining the genome of Arabidopsis thaliana as a basis
822 for the identification of novel bioactive peptides involved in oxidative stress tolerance.
823 Journal of Experimental Botany **64**: 5297–5307
824 **Constabel CP, Yip L, Ryan CA** (1998) Prosystemin from potato, black nightshade, and bell

pepper: primary structure and biological activity of predicted systemin polypeptides. Plant Molecular Biology **36**: 55–62

Crappe J, Ndah E, Koch A, Steyaert S, Gawron D, De Keulenaer S, De Meester E, De Meyer T, Van Criekinge W, Van Damme P, et al (2014) PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. Nucleic Acids Research **43**: 1–10

Dai X, Zhuang Z, Zhao PX (2018) psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Research **46**: W49–W54

Van Damme P, Gawron D, Van Criekinge W, Menschaert G (2014) N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. Molecular & Cellular Proteomics **13**: 1245–61

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

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. Journal of Molecular Biology **300**: 1005–1016

Faghihi MA, Wahlestedt C (2009) Regulatory roles of natural antisense transcripts. Nature Reviews Molecular Cell Biology **10**: 637–643

Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, et al (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Research **43**: D1036–D1041

Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, et al (2015) A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. Molecular Cell **60**: 816–827

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W,

851 **Hellsten U, Putnam N, et al** (2012) Phytozome: a comparative platform for green plant
852 genomics. Nucleic Acids Research **40**: D1178-86
853 **Guydosh NR, Green R** (2014) Dom34 rescues ribosomes in 3' untranslated regions. Cell **156**:
854 950–62
855 **Hsu PY, Benfey PN** (2018) Small but Mighty: Functional Peptides Encoded by Small ORFs in
856 Plants. PROTEOMICS **18**: 1700038
857 **Hsu PY, Calviello L, Wu H-YL, Li F-W, Rothfels CJ, Ohler U, Benfey PN** (2016) Super-
858 resolution ribosome profiling reveals unannotated translation events in Arabidopsis.
859 Proceedings of the National Academy of Sciences of the United States of America **113**:
860 E7126–E7135
861 **Ikeuchi M, Yamaguchi T, Kazama T, Ito T, Horiguchi G, Tsukaya H** (2011) ROTUNDIFOLIA4
862 Regulates Cell Proliferation Along the Body Axis in Arabidopsis Shoot. Plant and Cell
863 Physiology **52**: 59–69
864 **Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T** (2006) The dwarf
865 phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an upstream
866 ORF of a bHLH gene. Development **133**: 3575–85
867 **Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS** (2009) Genome-wide analysis
868 in vivo of translation with nucleotide resolution using ribosome profiling. Science **324**: 218–
869 23
870 **Ji Z, Song R, Regev A, Struhl K** (2015) Many lncRNAs, 5'UTRs, and pseudogenes are
871 translated and some are likely to express functional proteins. eLife **4**: e08890
872 **Johnstone TG, Bazzini AA, Giraldez AJ** (2016) Upstream ORFs are prevalent translational
873 repressors in vertebrates. The EMBO Journal **35**: 706–723
874 **Juntawong P, Girke T, Bazin J, Bailey-Serres J** (2014) Translational dynamics revealed by
875 genome-wide profiling of ribosome footprints in Arabidopsis. Proceedings of the National
876 Academy of Sciences of the United States of America **111**: E203-12

Kaur P, Shukla N, Joshi G, VijayaKumar C, Jagannath A, Agarwal M, Goel S, Kumar A (2017) Genome-wide identification and characterization of miRNAome from tomato (Solanum lycopersicum) roots and root-knot nematode (Meloidogyne incognita) during susceptible interaction. PLoS ONE **12**: e0175178

Kearse MG, Wilusz JE (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes & Development **31**: 1717–1731

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

Kolde R (2015) pheatmap: Pretty Heatmaps. https://cran.r-project.org/web/packages/pheatmap/index.html

Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Research **15**: 8125–48

Ku HM, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proceedings of the National Academy of Sciences of the United States of America **97**: 9121–6

Laing WA, Martínez-Sánchez M, Wright MA, Bulley SM, Brewster D, Dare AP, Rassam M, Wang D, Storey R, Macknight RC, et al (2015) An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in Arabidopsis. The Plant Cell **27**: 772–86

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

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ (2013) Software for computing and annotating genomic ranges. PLoS Computational Biology **9**: e1003118

Lei L, Shi J, Chen J, Zhang M, Sun S, Xie S, Li X, Zeng B, Peng L, Hauck A, et al (2015)

903 Ribosome profiling reveals dynamic translational landscape in maize seedlings under
904 drought stress. The Plant Journal **84**: 1206–18
905 **Li B, Dewey CN** (2011) RSEM: accurate transcript quantification from RNA-Seq data with or
906 without a reference genome. BMC Bioinformatics **12**: 323
907 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin**
908 **R, 1000 Genome Project Data Processing Subgroup** (2009) The Sequence
909 Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079
910 **Li S, Le B, Ma X, Li S, You C, Yu Y, Zhang B, Liu L, Gao L, Shi T, et al** (2016) Biogenesis of
911 phased siRNAs on membrane-bound polysomes in Arabidopsis. eLife **5**: e22750
912 **Li Z, Xu R, Li N** (2018) MicroRNAs from plants to animals, do they define a new messenger for
913 communication? Nutrition & Metabolism **15**: 68
914 **Liao Y, Smyth GK, Shi W** (2014) featureCounts: an efficient general purpose program for
915 assigning sequence reads to genomic features. Bioinformatics **30**: 923–930
916 **Liu M-J, Wu S-H, Wu J-F, Lin W-D, Wu Y-C, Tsai T-Y, Tsai H-L, Wu S-H** (2013) Translational
917 landscape of photomorphogenic Arabidopsis. The Plant Cell **25**: 3699–710
918 **Liu M, Yu H, Zhao G, Huang Q, Lu Y, Ouyang B** (2017) Profiling of drought-responsive
919 microRNA and mRNA in tomato using high-throughput sequencing. BMC Genomics **18**:
920 481
921 **Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA** (1987) Selection of AUG
922 initiation codons differs in plants and animals. The EMBO Journal **6**: 43–8
923 **Mata CI, Fabre B, Hertog MLATM, Parsons HT, Deery MJ, Lilley KS, Nicolaï BM** (2017) In-
924 depth characterization of the tomato fruit pericarp proteome. PROTEOMICS **17**: 1600406
925 **Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappé J, Gevaert K, Van Damme P**
926 (2013) Deep proteome coverage based on ribosome profiling aids mass spectrometry-
927 based protein and peptide discovery and provides evidence of alternative translation
928 products and near-cognate translation initiation events. Molecular & Cellular Proteomics **12**:

1780–90

Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, Binder BM, Heber S, Stepanova AN, Alonso JM (2015) Gene-Specific Translation Regulation Mediated by the Hormone-Signaling Molecule EIN2. Cell **163**: 684–697

Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov P V (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Research **22**: 2219–29

Neuwirth E (2014) RColorBrewer: ColorBrewer Palettes.

Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nature reviews Genetics **12**: 87–98

Pearce G, Strydom D, Johnson S, Ryan CA (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science (New York, NY) **253**: 895–7

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols **11**: 1650–1667

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology **33**: 290–295

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

Ruggles K V, Krug K, Wang X, Clauser KR, Wang J, Payne SH, Fenyö D, Zhang B, Mani DR (2017) Methods, Tools and Current Perspectives in Proteogenomics. Molecular & Cellular Proteomics **16**: 959–981

Ruiz-Orera J, Albà MM (2019) Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. Trends in Genetics **35**: 186–198

Sagor GHM, Berberich T, Tanaka S, Nishiyama M, Kanayama Y, Kojima S, Muramoto K,

Kusano T (2016) A novel strategy to produce sweeter tomato fruits with high sugar contents by fruit-specific expression of a single bZIP transcription factor gene. Plant Biotechnology Journal **14**: 1116–1126

Schafer S, Adami E, Heinig M, Rodrigues KEC, Kreuchwig F, Silhavy J, van Heesch S, Simaite D, Rajewsky N, Cuppen E, et al (2015) Translational regulation shapes the molecular landscape of complex disease phenotypes. Nature Communications **6**: 7200

Schwarz D, Thompson AJ, Kläring H-P (2014) Guidelines to use tomato in experiments with a controlled environment. Frontiers in Plant Science **5**: 625

Shamimuzzaman M, Vodkin L (2018) Ribosome profiling reveals changes in translational status of soybean transcripts during immature cotyledon development. PLoS ONE **13**: e0194596

Simpson GG, Laurie RE, Dijkwel PP, Quesada V, Stockwell PA, Dean C, Macknight RC (2010) Noncanonical translation initiation of the Arabidopsis flowering time and alternative polyadenylation regulator FCA. The Plant Cell **22**: 3764–77

Song G, Brachova L, Nikolau BJ, Jones AM, Walley JW (2018a) Heterotrimeric G-Protein-Dependent Proteome and Phosphoproteome in Unstimulated Arabidopsis Roots. PROTEOMICS **18**: 1800323

Song G, Hsu PY, Walley JW (2018b) Assessment and Refinement of Sample Preparation Methods for Deep and Quantitative Plant Proteome Profiling. PROTEOMICS **18**: 1800220

Spealman P, Naik AW, May GE, Kuersten S, Freeberg L, Murphy RF, McManus J (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. Genome Research **28**: 214–222

Szymanski J, Levin Y, Savidor A, Breitel D, Chappell-Maor L, Heinig U, Töpfer N, Aharoni A (2017) Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues development. The Plant Journal **90**: 396–417

Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammue BPA (2015) The Plant

Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. The Plant Cell **27**: 2095–118

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Research **45**: W122–W129

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

Valdivia ER, Chevalier D, Sampedro J, Taylor I, Niederhuth CE, Walker JC (2012) DVL genes play a role in the coordination of socket cell recruitment and differentiation. Journal of Experimental Botany **63**: 1405–1412

Walley JW, Briggs SP (2015) Dual use of peptide mass spectra: Protein atlas and genome annotation. Current Plant Biology **2**: 21–24

Wei T (2013) corrplot: Visualization of a correlation matrix. https://cran.r-project.org/web/packages/corrplot/index.html

Xu G, Greene GH, Yoo H, Liu L, Marqués J, Motley J, Dong X (2017a) Global translational reprogramming is a fundamental layer of immune regulation in plants. Nature **545**: 487–490

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

Yu Y, Jia T, Chen X (2017) The ‘how’ and ‘where’ of plant microRNAs. New Phytologist **216**: 1002–1017

Zhang H, Si X, Ji X, Fan R, Liu J, Chen K, Wang D, Gao C (2018) Genome editing of upstream open reading frames enables translational control in plants. Nature Biotechnology **36**: 894–898

Zoschke R, Watkins KP, Barkan A (2013) A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. The Plant Cell **25**: 2265–75

1007

Figure 1

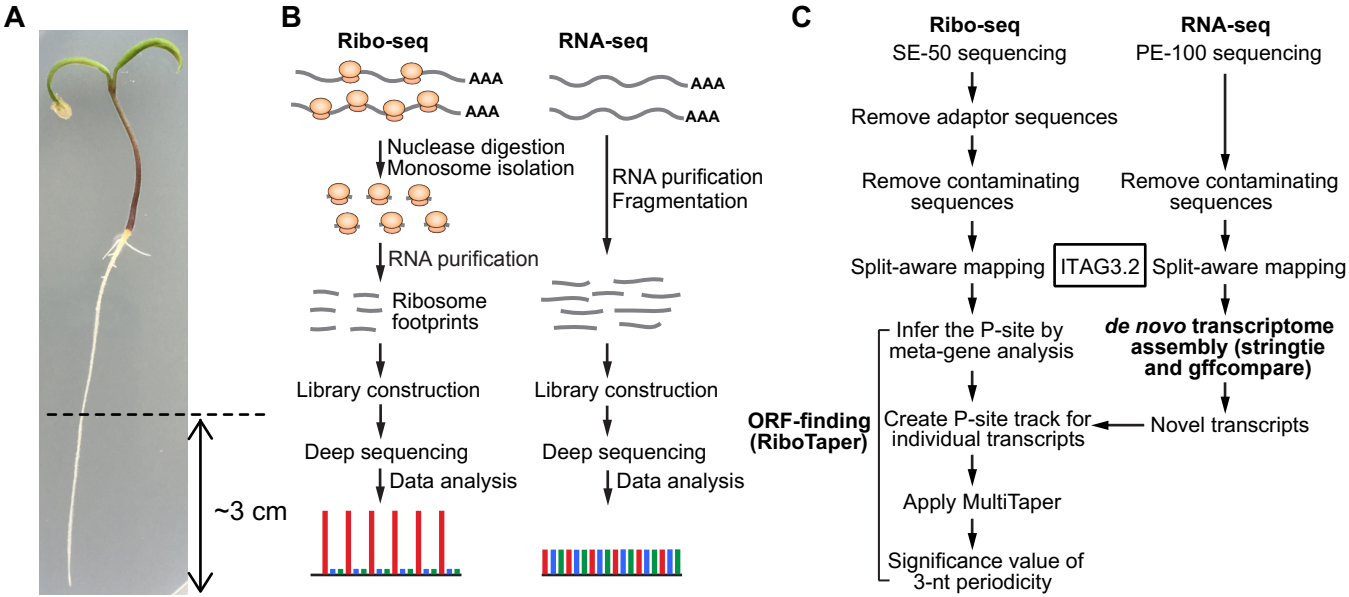


Figure 1: Experimental and data analysis procedures for ribosome profiling in tomato roots.
(A) Four-day-old tomato roots (approximately 3 cm from the tip) were used in this study.
(B) Experimental workflow for RNA-seq and Ribo-seq and the schematics of their expected read distributions in the three reading frames. This figure was adapted from Hsu *et al.* 2016.
(C) Data analysis workflow for reference-guided de novo transcriptome assembly and ORF discovery using RiboTaper.

Figure 2

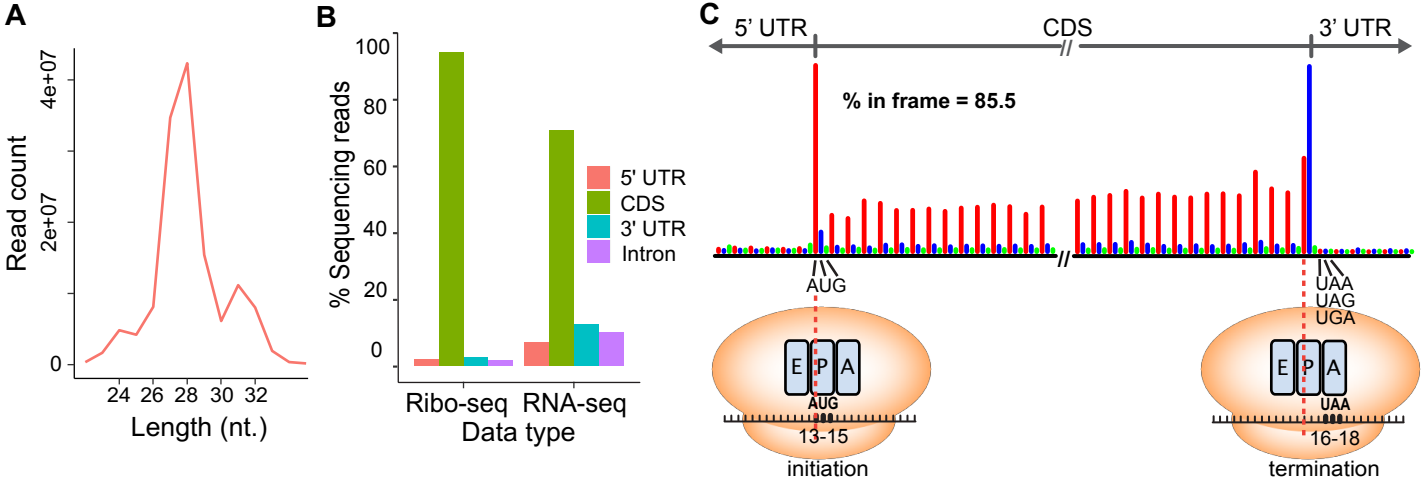


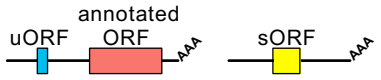
Figure 2: Ribosome footprints are enriched in coding sequences and display strong 3-nt periodicity.
(A) The distribution of read length of the ribosome footprints.
(B) The distribution of the Ribo-seq and RNA-seq reads in different genomic features annotated in ITAG3.2.
(C) Meta-gene analysis of the 28-nt ribosome footprints near the annotated translation start and stop sites defined by ITAG3.2. The red, blue and green bars represent reads mapped to the first (expected), second and third reading frames, respectively. The majority of footprints were mapped to the CDS in the expected reading frame (85.5% in frame). For each read, only the first nucleotide in the P-site was plotted (for details, see Supplemental Figures 2 and 3). The A-site (aminoacyl-tRNA entry site), P-site (peptidyl-tRNA formation site) and E-site (uncharged tRNA exit site) within the ribosomes at translation initiation and termination, and the inferred P-site (13th-15th nts) and A-site (16th-18th nts) are illustrated. The original meta-plots generated by RiboTaper for all footprint lengths are shown in Supplemental Figure 2.

Figure 3

A

o	On same strand and overlap with annotated exons
s	Intron match on the opposite strand (likely a mapping error)
i	New transcript fully contained in an annotated intron
x	Exon match on the opposite strand (a.k.a. cis-NATs)
y	New transcript contains an annotated gene within its intron(s)
u	Intergenic

B

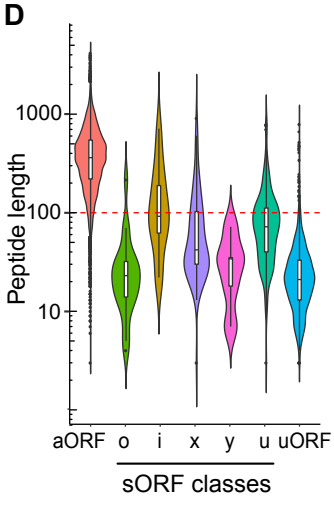


Method	Transcripts from ITAG3.2		Transcripts from *de novo* assembly
	uORFs	annotated ORFs	novel ORFs (sORFs)
Ribo-seq	1329	20659	354
MassSpec	29	12113	30

C

Class	o	s	i	x	y	u
No. transcripts	307	13	100	568	15	1260
No. translated transcripts	20	0	13	72	5	218
ORFs identified	23	0	14	82	5	230

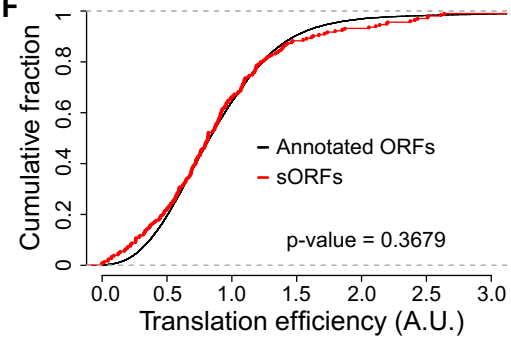
D



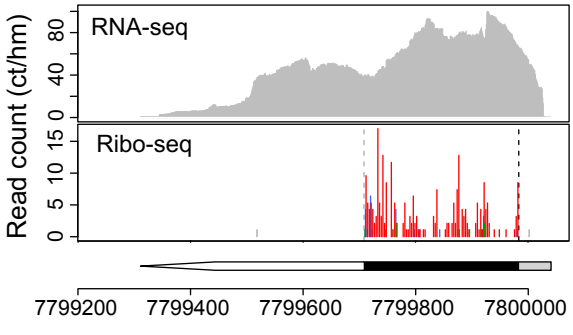
E

sORF Localization \ Specificity	0.90
Secreted	87
Chloroplast	19
Mitochondrion	7
Unknown	241

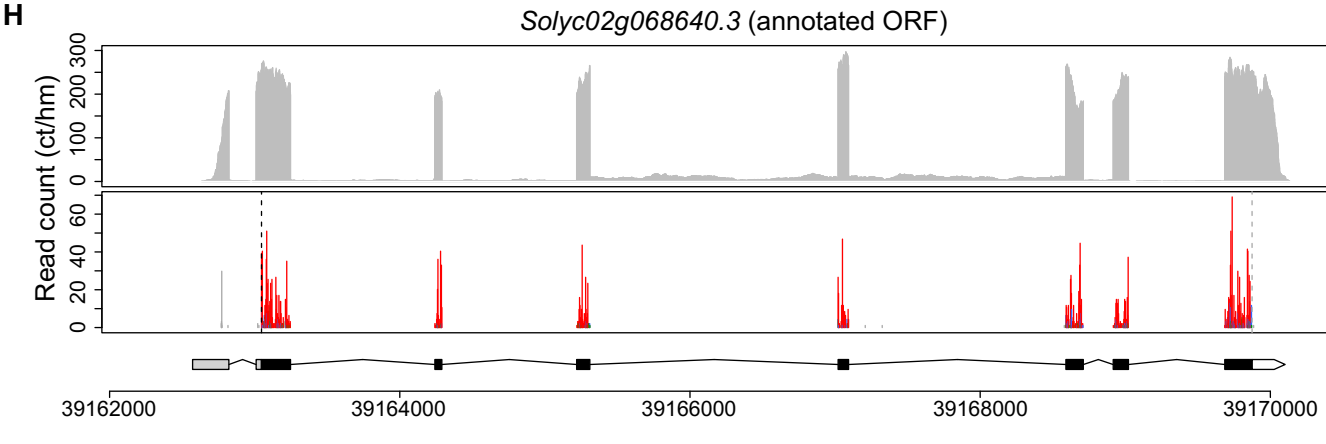
F



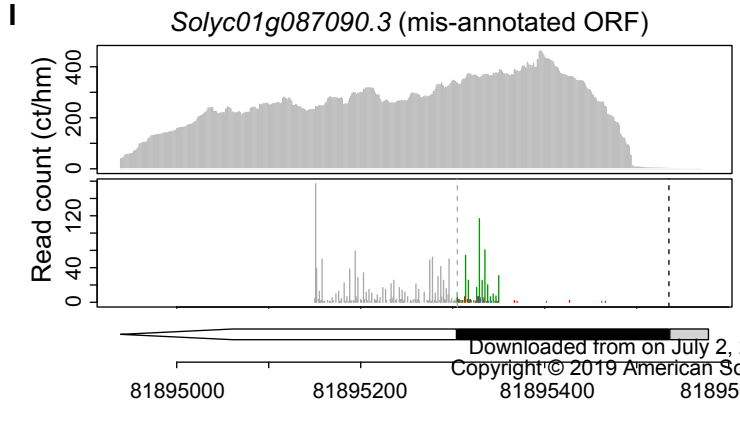
G



H



I



J

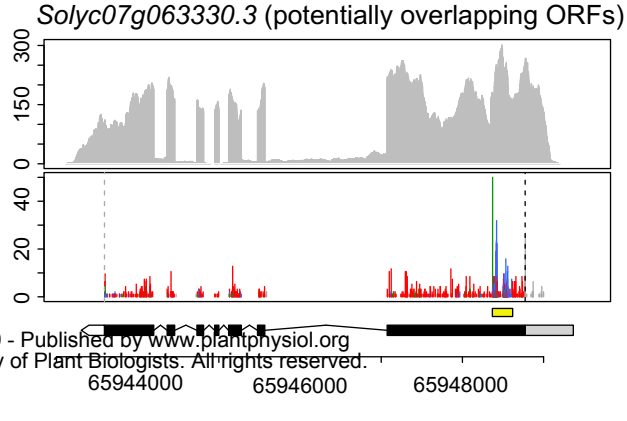
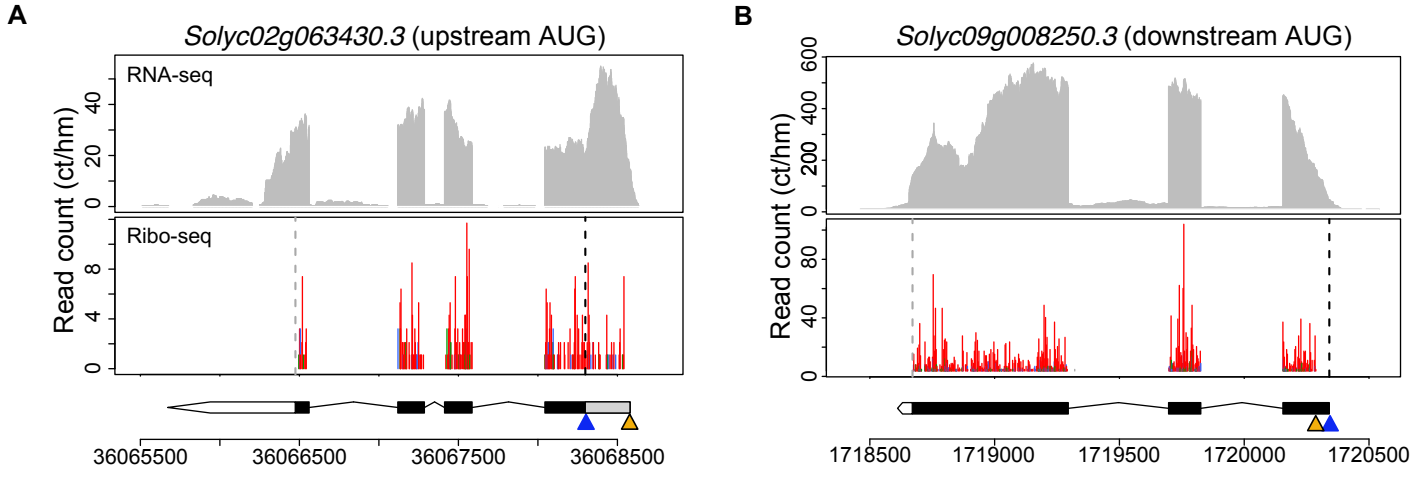
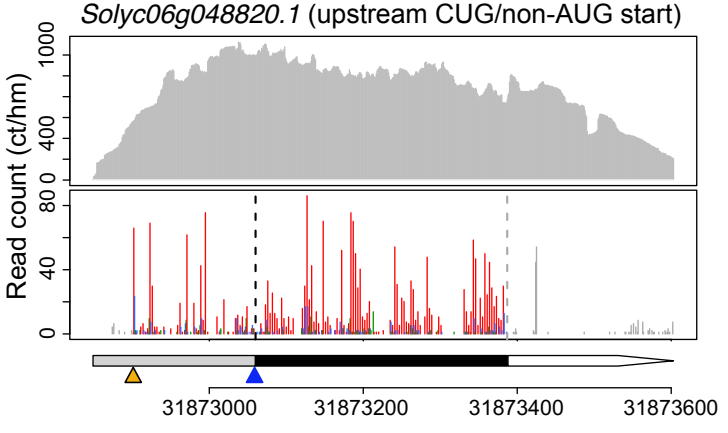


Figure 3: The translational landscape of the tomato root.
(A) Classes of newly assembled transcripts identified by stringtie and gffcompare and used in downstream ORF identification. This figure was adapted from the gffcompare website (Pertea et al., 2016).
(B) Summary of translated ORFs identified by RiboTaper in our dataset and peptide support from mass spectrometry data. The uORFs and annotated ORFs were identified from the 5' UTRs and expected CDSs of annotated protein-coding genes in ITAG3.2, respectively. The novel ORFs were identified from the newly assembled transcripts. The bottom row indicates the number of proteins in each category supported by mass spectrometry datasets, either from our own proteomic analysis or searches against publicly available data.
(C) Summary of newly assembled transcripts and ORFs identified in each class of newly assembled transcripts. The total number of transcripts, number of transcripts identified as translated and the total number of translated ORFs are listed.
(D) Size distribution of each class of sORFs, uORFs and annotated ORFs (aORFs).
(E) Predicted subcellular localization of proteins encoded by the sORFs. The prediction was performed using TargetP (Emanuelsson et al., 2000) with specificity 0.9 as a cutoff.
(F) Translation efficiency of sORFs compared with annotated ORFs. Only the coding regions were used to compute the TPM and translation efficiency of each transcript. For the x-axis, only the range from 0 to 3 (arbitrary unit) is shown. A two-sample Kolmogorov-Smirnov test was used to determine statistical significance.
(G-J) RNA-seq coverage and Ribo-seq periodicity in different genes: (G) an intergenic sORF on chromosome 4; (H) an annotated coding gene that has good support from the Ribo-seq data for the predicted gene model; (I) a mis-annotated ORF; note the Ribo-seq reads do not match the CDS in the gene model and a different reading frame is used; (J) a transcript with a potentially overlapping ORF within the annotated ORF. In (G-J), the x-axis indicates the genomic coordinate of the gene. The y-axis shows the normalized read count (counts per hundred million reads). Ribo-seq reads are shown by plotting the first nucleotide of their P-sites (denoted as the P-site signals). The black and gray dashed vertical lines mark the predicted translation start and stop sites, respectively. The red, blue and green lines in the Ribo-seq plot indicate the P-site signals mapped to the first (expected) reading frame and the second and third reading frames, respectively. Hence, a higher ratio of red means better 3-nt periodicity. For the gene model beneath the Ribo-seq data, the gray, black and white areas indicate the 5' UTR, CDS and 3' UTR, respectively. In (J), the yellow box above the gene model indicates the region with a potential ORF overlapping with the annotated ORF.

Figure 4



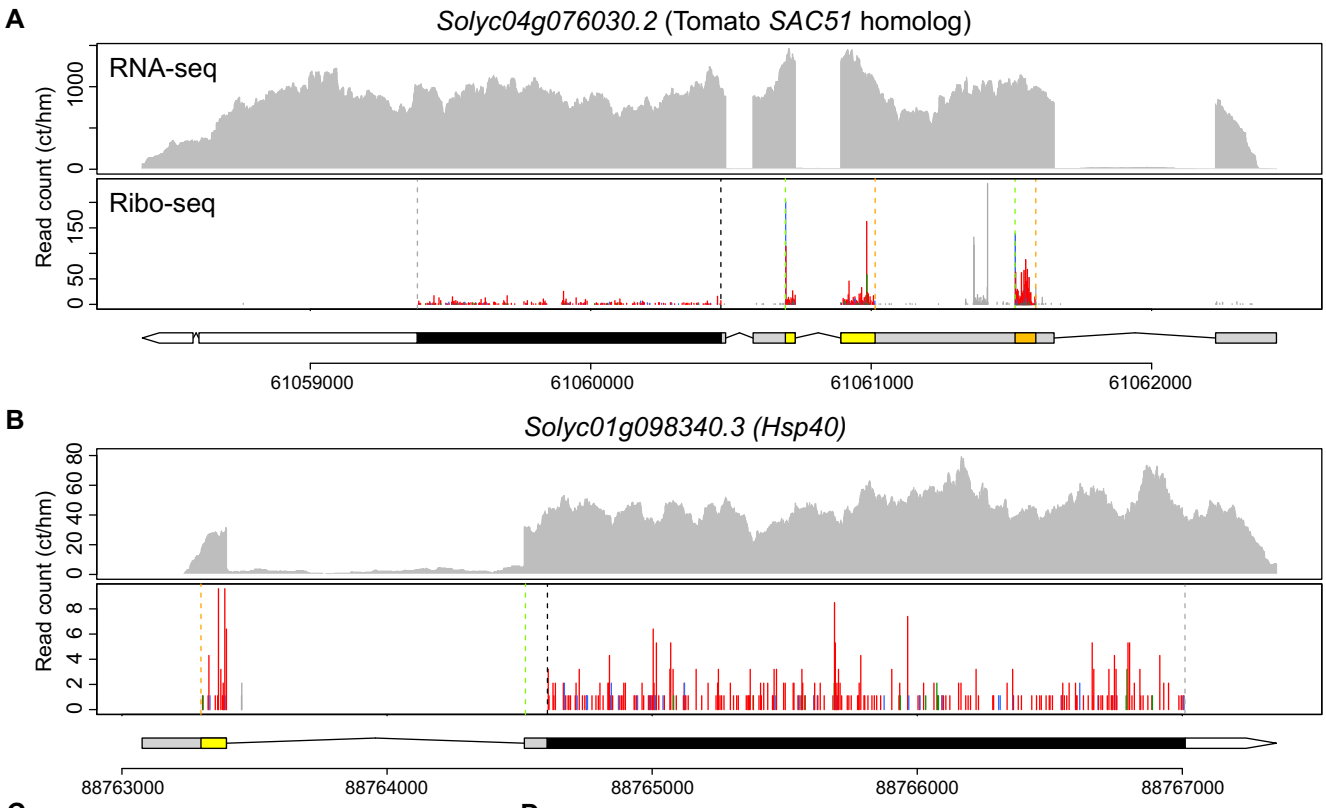


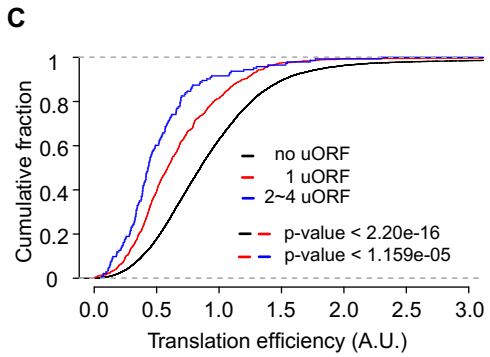
D

Arabidopsis gene ID	Tomato gene ID	% a.a. identity	In-frame P-sites/20
AT1G03260	*Solyc01g080110.3*	67.7	8/20
AT1G21000	*Solyc08g076860.3*	52.5	17/20
AT1G55760	*Solyc07g065930.3*	74.6	7/20
AT3G04080	*Solyc08g016010.3*	67.3	12/20
AT3G04080	*Solyc12g096560.2*	66.5	16/20
AT3G10985	*Solyc06g050760.1*	54.8	14/20
AT3G10985	*Solyc06g048820.1*	51.8	13/20
AT4G02680	*Solyc09g065640.3*	73.7	7/20

Figure 4: Upstream/downstream start sites and non-AUG start sites.
(A-B) Examples of the usage of an upstream start site (A) or a downstream start site (B). The gene model and data presentation are the same as those described in the legend of Figure 3. The blue triangle marks the location of the annotated translation start site. The orange triangle marks the location of the RiboTaper-identified translation start site.
(C) A tomato homolog of an Arabidopsis gene that was predicted to use an upstream CUG start site (orange triangle). Note the abundant in-frame P-site signals upstream of the annotated AUG start (blue triangle) in the 5' UTR.
(D) Conservation of potential CUG/non-AUG start sites. The Arabidopsis gene ID, tomato gene ID, percent amino acid identity, and number of in-frame P-site positions with Ribo-seq reads within the first 20 codons upstream of the AUG in our tomato root data are shown.

Figure 5





D

GO category	Description	FDR
0004672	protein kinase activity	6.5e-20
0006793	phosphorus metabolic process	1.1e-17
0043412	macromolecule modification	2.8e-16
0005524	ATP binding	1.4e-11
0004871	signal transducer activity	2.2e-05
0038023	signaling receptor activity	4.1e-05
0019888	protein phosphatase regulator activity	9.0e-05
0004721	phosphoprotein phosphatase activity	0.033

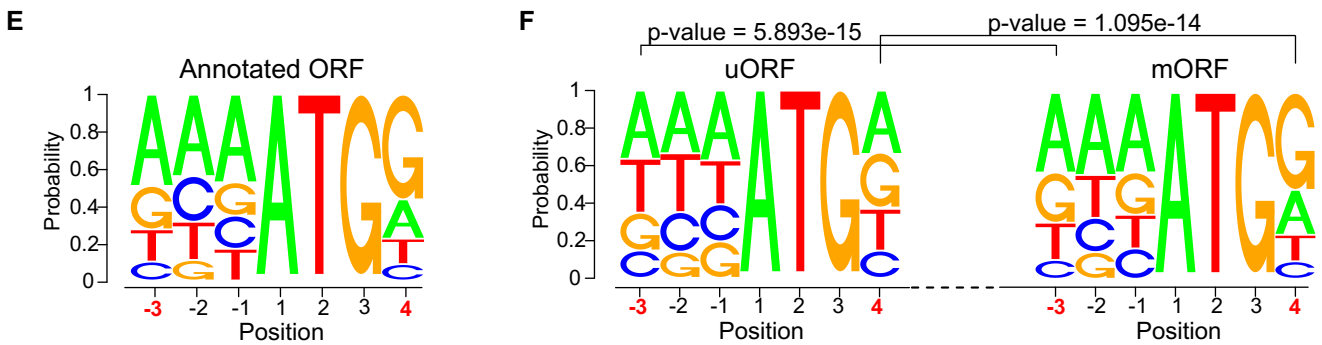


Figure 5: uORFs repress translation efficiency of their downstream main ORFs and contain less-pronounced Kozak sequences.

(A, B) Profiles of genes containing conserved uORFs (A) or a novel uORF (B). The gene model and data presentation are the same as those described in the legend of Figure 3. The uORFs are labeled with yellow and orange boxes in the gene models. For the uORFs, the orange and green dashed vertical lines mark the translation start and stop sites, respectively.

(C) The translation efficiency (TE) of the main ORFs for transcripts containing a different number of translated uORFs. Only the coding regions were used to compute the TPM and translation efficiency of each transcript. The colored bars before the p-values indicate the pairs of data used to determine statistical significance. The p-values were determined with two-sample Kolmogorov-Smirnov tests.

(D) Selected non-redundant GO categories for genes containing conserved uORFs.

(E, F) Kozak sequences of annotated ORFs, uORFs, and uORF-associated main ORFs. The statistical significance in (F) was determined using Chi-squared tests.

Figure 6

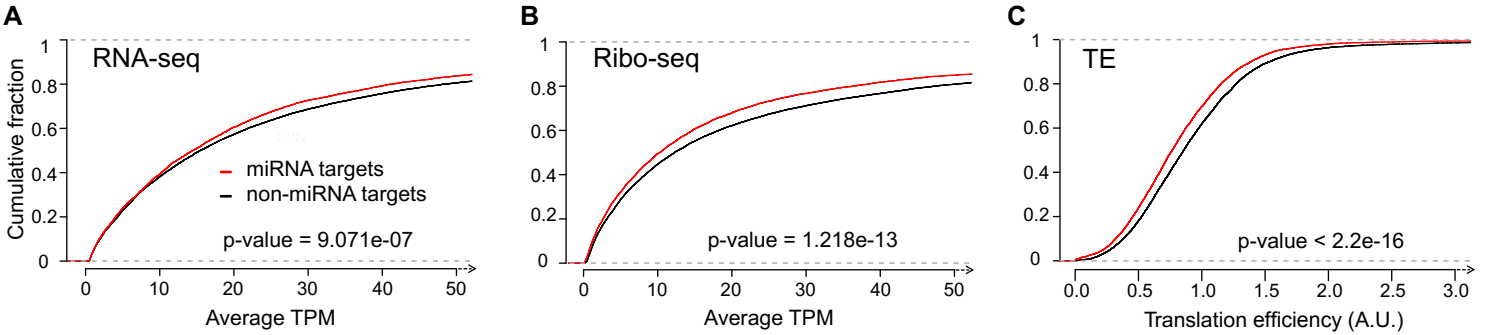


Figure 6: Regulation of gene expression by microRNAs (miRNAs).
Cumulative distributions of (A) RNA-seq, (B) Ribo-seq and (C) translation efficiency (TE) of miRNA targets and non-miRNA target genes. For the x-axis in (A) and (B), only the range from 0 and 50 (TPM) is shown. Only the coding regions were used to compute the TPM and translation efficiency of each transcript. The p-values were determined with two-sample Kolmogorov-Smirnov tests.

Parsed Citations

Agrawal GK, Jwa N-S, Lebrun M-H, Job D, Rakwal R (2010) Plant secretome: Unlocking secrets of the secreted proteins. PROTEOMICS 10: 799–827

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Andreev DE, O'Connor PBF, Loughran G, Dmitriev SE, Baranov P V., Shatsky IN (2017) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. Nucleic Acids Research 45: 513–526

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. Nature reviews Genetics 15: 193–204

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

von Arnim AG, Jia Q, Vaughn JN (2013) Regulation of plant translation by upstream open reading frames. Plant Science 214: 1–12

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Basrai MA, Hieter P, Boeke JD (1997) Small Open Reading Frames: Beautiful Needles in the Haystack. Genome Research 7: 768–771

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, 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

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. The EMBO Journal 33: 981–993

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Blanvillain R, Young B, Cai Y, Hecht V, Varoquaux F, Delorme V, Lancelin J-M, Delseny M, Gallois P (2011) The Arabidopsis peptide kiss of death is an inducer of programmed cell death. The EMBO Journal 30: 1173–1183

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

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

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335: 552–7

Pubmed: Author and Title
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, Ohler U (2016) Detecting actively translated open reading frames in ribosome profiling data. Nature Methods 13: 165–170

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10: 421

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Casson SA (2002) The POLARIS Gene of Arabidopsis Encodes a Predicted Peptide Required for Correct Root Growth and Leaf Vascular Patterning. The Plant Cell 14: 1705–1721

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Castellana NE, Shen Z, He Y, Walley JW, Cassidy CJ, Briggs SP, Bafna V (2014) An automated proteogenomic method uses mass spectrometry to reveal novel genes in Zea mays. Molecular & Cellular Proteomics 13: 157–67

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

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

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chotewutmontri P, Barkan A (2016) Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. PLoS Genetics 12: e1006106
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley I (2015) The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. RNA 10: 1731–45
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Claverie J (1997) Computational methods for the identification of genes in vertebrate genomic sequences. Human Molecular Genetics 6: 1735–1744
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

De Coninck B, Carron D, Tavormina P, Willem L, Craik DJ, Vos C, Thevissen K, Mathys J, Cammue BPA, SM. D, et al (2013) Mining the genome of Arabidopsis thaliana as a basis for the identification of novel bioactive peptides involved in oxidative stress tolerance. Journal of Experimental Botany 64: 5297–5307
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Constabel CP, Yip L, Ryan CA (1998) Prosystemin from potato, black nightshade, and bell pepper: primary structure and biological activity of predicted systemin polypeptides. Plant Molecular Biology 36: 55–62
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Crappe J, Ndah E, Koch A, Steyaert S, Gawron D, De Keulenaer S, De Meester E, De Meyer T, Van Criekinge W, Van Damme P, et al (2014) PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. Nucleic Acids Research 43: 1–10
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Dai X, Zhuang Z, Zhao PX (2018) psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Research 46: W49–W54
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Van Damme P, Gawron D, Van Criekinge W, Menschaert G (2014) N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. Molecular & Cellular Proteomics 13: 1245–61
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. Journal of Molecular Biology 300: 1005–1016
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Faghihi MA, Wahlestedt C (2009) Regulatory roles of natural antisense transcripts. Nature Reviews Molecular Cell Biology 10: 637–643
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, et al (2015) The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. Nucleic Acids Research 43: D1036–D1041
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, et al (2015) A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. Molecular Cell 60: 816–827
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40: D1178-86
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Guydosh NR, Green R (2014) Dom34 rescues ribosomes in 3′ untranslated regions. Cell 156: 950–62

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hsu PY, Benfey PN (2018) Small but Mighty: Functional Peptides Encoded by Small ORFs in Plants. PROTEOMICS 18: 1700038
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hsu PY, Calviello L, Wu H-YL, Li F-W, Rothfels CJ, Ohler U, Benfey PN (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
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ikeuchi M, Yamaguchi T, Kazama T, Ito T, Horiguchi G, Tsukaya H (2011) ROTUNDIFOLIA4 Regulates Cell Proliferation Along the Body Axis in Arabidopsis Shoot. Plant and Cell Physiology 52: 59–69
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T (2006) The dwarf phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. Development 133: 3575–85
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324: 218–23
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ji Z, Song R, Regev A, Struhl K (2015) Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. eLife 4: e08890
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Johnstone TG, Bazzini AA, Giraldez AJ (2016) Upstream ORFs are prevalent translational repressors in vertebrates. The EMBO Journal 35: 706–723
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Juntawong P, Girke T, Bazin J, 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: E203-12
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kaur P, Shukla N, Joshi G, VijayaKumar C, Jagannath A, Agarwal M, Goel S, Kumar A (2017) Genome-wide identification and characterization of miRNAome from tomato (Solanum lycopersicum) roots and root-knot nematode (Meloidogyne incognita) during susceptible interaction. PLoS ONE 12: e0175178
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kearse MG, Wilusz JE (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes & Development 31: 1717–1731
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kim B-H, Cai X, Vaughn JN, von Arnim AG (2007) On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. Genome biology 8: R60
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kolde R (2015) pheatmap: Pretty Heatmaps. https://cran.r-project.org/web/packages/pheatmap/index.html
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Research 15: 8125–48
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ku HM, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proceedings of the National Academy of Sciences of the United States of America 97: 9121–6
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Laing WA, Martínez-Sánchez M, Wright MA, Bulley SM, Brewster D, Dare AP, Rassam M, Wang D, Storey R, Macknight RC, et al (2015) An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in Arabidopsis. The Plant Cell 27: 772–86

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–9
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ (2013) Software for computing and annotating genomic ranges. PLoS Computational Biology 9: e1003118
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lei L, Shi J, Chen J, Zhang M, Sun S, Xie S, Li X, Zeng B, Peng L, Hauck A, et al (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. The Plant Journal 84: 1206–18
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li S, Le B, Ma X, Li S, You C, Yu Y, Zhang B, Liu L, Gao L, Shi T, et al (2016) Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. eLife 5: e22750
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li Z, Xu R, Li N (2018) MicroRNAs from plants to animals, do they define a new messenger for communication? Nutrition & Metabolism 15: 68

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923–930
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Liu M-J, Wu S-H, Wu J-F, Lin W-D, Wu Y-C, Tsai T-Y, Tsai H-L, Wu S-H (2013) Translational landscape of photomorphogenic Arabidopsis. The Plant Cell 25: 3699–710
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Liu M, Yu H, Zhao G, Huang Q, Lu Y, Ouyang B (2017) Profiling of drought-responsive microRNA and mRNA in tomato using high-throughput sequencing. BMC Genomics 18: 481
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differs in plants and animals. The EMBO Journal 6: 43–8
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Mata CI, Fabre B, Hertog MLATM, Parsons HT, Deery MJ, Lilley KS, Nicolaï BM (2017) In-depth characterization of the tomato fruit pericarp proteome. PROTEOMICS 17: 1600406
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappé J, Gevaert K, Van Damme P (2013) Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Molecular & Cellular Proteomics 12: 1780–90
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, Binder BM, Heber S, Stepanova AN, Alonso JM (2015) Gene-Specific Translation Regulation Mediated by the Hormone-Signaling Molecule EIN2. Cell 163: 684–697
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov P V (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Research 22: 2219–29
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Neuwirth E (2014) RColorBrewer: ColorBrewer Palettes.

Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nature reviews Genetics 12: 87–98
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pearce G, Strydom D, Johnson S, Ryan CA (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science (New York, NY) 253: 895–7
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols 11: 1650–1667
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33: 290–295
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

R Core Team (2013) (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. doi: /S0103-64402004000300015
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ruggles K V, Krug K, Wang X, Clauser KR, Wang J, Payne SH, Fenyö D, Zhang B, Mani DR (2017) Methods, Tools and Current Perspectives in Proteogenomics. Molecular & Cellular Proteomics 16: 959–981
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ruiz-Orera J, Albà MM (2019) Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. Trends in Genetics 35: 186–198
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sagor GHM, Berberich T, Tanaka S, Nishiyama M, Kanayama Y, Kojima S, Muramoto K, Kusano T (2016) A novel strategy to produce sweeter tomato fruits with high sugar contents by fruit-specific expression of a single bZIP transcription factor gene. Plant Biotechnology Journal 14: 1116–1126
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schafer S, Adami E, Heinig M, Rodrigues KEC, Kreuchwig F, Silhavy J, van Heesch S, Simaite D, Rajewsky N, Cuppen E, et al (2015) Translational regulation shapes the molecular landscape of complex disease phenotypes. Nature Communications 6: 7200
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Schwarz D, Thompson AJ, Kläring H-P (2014) Guidelines to use tomato in experiments with a controlled environment. Frontiers in Plant Science 5: 625
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Shamimuzzaman M, Vodkin L (2018) Ribosome profiling reveals changes in translational status of soybean transcripts during immature cotyledon development. PLoS ONE 13: e0194596
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Simpson GG, Laurie RE, Dijkwel PP, Quesada V, Stockwell PA, Dean C, Macknight RC (2010) Noncanonical translation initiation of the Arabidopsis flowering time and alternative polyadenylation regulator FCA. The Plant Cell 22: 3764–77
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Song G, Brachova L, Nikolau BJ, Jones AM, Walley JW (2018a) Heterotrimeric G-Protein-Dependent Proteome and Phosphoproteome in Unstimulated Arabidopsis Roots. PROTEOMICS 18: 1800323
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Song G, Hsu PY, Walley JW (2018b) Assessment and Refinement of Sample Preparation Methods for Deep and Quantitative Plant Proteome Profiling. PROTEOMICS 18: 1800220
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Spealman P, Naik AW, May GE, Kuersten S, Freeberg L, Murphy RF, McManus J (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. Genome Research 28: 214–222

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Szymanski J, Levin Y, Savidor A, Breitel D, Chappell-Maor L, Heinig U, Töpfer N, Aharoni A (2017) Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues development. The Plant Journal 90: 396–417
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammue BPA (2015) The Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions. The Plant Cell 27: 2095–118
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Research 45: W122–W129
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

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

Valdivia ER, Chevalier D, Sampedro J, Taylor I, Niederhuth CE, Walker JC (2012) DVL genes play a role in the coordination of socket cell recruitment and differentiation. Journal of Experimental Botany 63: 1405–1412
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Walley JW, Briggs SP (2015) Dual use of peptide mass spectra: Protein atlas and genome annotation. Current Plant Biology 2: 21–24
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wei T (2013) corrplot: Visualization of a correlation matrix. https://cran.r-project.org/web/packages/corrplot/index.html
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xu G, Greene GH, Yoo H, Liu L, Marqués J, Motley J, Dong X (2017a) Global translational reprogramming is a fundamental layer of immune regulation in plants. Nature 545: 487–490
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Xu G, Yuan M, Ai C, Liu L, Zhuang E, Karapetyan S, Wang S, Dong X (2017b) uORF-mediated translation allows engineered plant disease resistance without fitness costs. Nature 545: 491–494
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Yu Y, Jia T, Chen X (2017) The 'how' and 'where' of plant microRNAs. New Phytologist 216: 1002–1017
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zhang H, Si X, Ji X, Fan R, Liu J, Chen K, Wang D, Gao C (2018) Genome editing of upstream open reading frames enables translational control in plants. Nature Biotechnology 36: 894–898
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zoschke R, Watkins KP, Barkan A (2013) A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. The Plant Cell 25: 2265–75
Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title