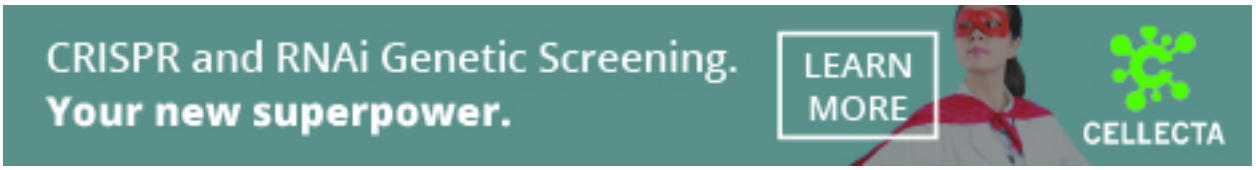
The *ggRibo* single-gene viewer reveals insights into translatome and other nucleotide-resolution omics data

Hsin-Yen Larry Wu, Isaiah D. Kaufman and Polly Yingshan Hsu

Genome Res. published online July 22, 2025
Access the most recent version at doi:10.1101/gr.280480.125



Resource

The *ggRibo* single-gene viewer reveals insights into translatome and other nucleotide-resolution omics data

Hsin-Yen Larry Wu, Isaiah D. Kaufman, and Polly Yingshan Hsu

Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

Visualizing Ribo-seq and other sequencing data within genes of interest is a powerful approach to studying gene expression, but its application is limited by a lack of robust tools. Here, we introduce *ggRibo*, a user-friendly R package for visualizing individual gene expression, integrating Ribo-seq, RNA-seq, and other genome-wide data sets with flexible scaling options. *ggRibo* visualizes 3 nt periodicity, a hallmark of translating ribosomes, within a gene-structure context, including introns and untranslated regions, enabling the study of novel open reading frames (ORFs), translation of different isoforms, and mechanisms of translational regulation. *ggRibo* can plot multiple Ribo-seq/RNA-seq data sets from different conditions for comparison. It also contains functions for plotting single-transcript views, reading-frame decomposition, and RNA-seq coverage alone. Importantly, *ggRibo* supports the visualization of other omics data sets that could also be presented with single-nucleotide resolution, such as RNA degradome, transcription start sites, translation initiation sites, and epitranscriptomic modifications. We demonstrate its utility with examples of upstream ORFs, downstream ORFs, nested ORFs, and differential isoform translation in humans, *Arabidopsis*, tomatos, and rice. We also provide examples of multiomic comparisons that reveal insights that connect the transcriptome, translatome, and degradome. In summary, *ggRibo* is an advanced single-gene viewer that offers a valuable resource for studying gene expression regulation through its intuitive and flexible platform.

[Supplemental material is available for this article.]

Visualization is emphasized in cell biology research to generate hypotheses and results. However, in omics studies, the visualization of sequencing data within genes of interest remains underutilized owing to a lack of appropriate tools. Recent advancements in plotting tools in the R programming language (R Core Team 2025) now offer opportunities to develop specialized software packages for multiomic data visualization.

Ribo-seq is a widely used technique for studying noncanonical open reading frames (ORFs) and mechanisms of mRNA translation at single-nucleotide resolution (Ingolia et al. 2009; Brar and Weissman 2015; Wu et al. 2024b). By optimizing RNase digestion to obtain ribosome-protected mRNA fragments (ribosome footprints), Ribo-seq captures the codon-by-codon, 3 nucleotide (nt) movement of ribosomes during translation, a feature known as 3 nt periodicity (Ingolia et al. 2009). During data analysis, specific single-nucleotide positions can be assigned to each ribosome footprint and color-coded by reading frame, enabling visualization of 3 nt periodicity along translated ORFs (Calviello et al. 2016; Hsu et al. 2016). Detecting 3 nt periodicity is crucial for identifying noncanonical ORFs and investigating the mechanism of mRNA translation (Guydosh and Green 2014; Chothani et al. 2023; Wu et al. 2024a).

Although several Ribo-seq software packages generate aggregate plots to examine 3 nt periodicity at a global level (Ji et al. 2015; Calviello et al. 2016, 2020; Zhang et al. 2017; Xiao et al. 2018; Choudhary et al. 2020; Harnett et al. 2021), visualizing 3 nt periodicity within individual genes remains challenging owing to the complexity of gene structure and isoforms. Many studies simply present Ribo-seq data as coverage plots or single-color lines for genes of interest, disregarding the rich information offered by 3 nt periodicity. Others may depict periodicity within a single transcript isoform (Kiniry et al. 2019), assuming the selected isoform is the main expressed/translated isoform, thereby missing the opportunity to examine other annotated or unannotated isoforms. Although some studies display Ribo-seq in a genome browser, providing gene-structure and isoform context, the reads are still shown in one color, preventing the intuitive visualization of 3 nt periodicity (Michel et al. 2018). Additionally, many available Ribo-seq visualization tools plot one sample at a time, making comparisons between samples challenging.

We previously developed *RiboPlotR* to visualize Ribo-seq reads for all annotated isoforms of individual genes within their gene structures, including 5′ and 3′ untranslated regions (UTRs) and introns. It displays both RNA-seq coverage and Ribo-seq periodicity by color-coding reads from each reading frame (Wu and Hsu 2021). Although *RiboPlotR* enables the visualization of isoform translation and many unannotated translation features, it has several limitations: First, it only accepts a specific annotation format commonly used for plant gene annotation, in which transcript IDs are assigned according to gene IDs plus a number. Therefore, *RiboPlotR* cannot plot data from human and many other non-plant species. Second, it only plots annotated ORFs and upstream ORFs (uORFs) in the 5′ UTR, not other unannotated ORFs, such as downstream ORFs (dORFs) in the 3′ UTR or novel small ORFs within annotated noncoding RNAs. Third, it only displays a maximum of two Ribo-seq samples per plot, preventing large-scale comparisons. Fourth, it only accepts a tabular input file for Ribo-seq and BAM files for RNA-seq, but not other standard input files, such

Corresponding author: pollyhsu@msu.edu

Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.280480.125. Freely available online through the *Genome Research* Open Access option.

as bedGraph or bigWig. Fifth, it does not display DNA and amino acid sequences, which aid in identifying ORFs or specific sequence features. Sixth, it relies on base R plotting capabilities, producing figures that are less refined than those generated with *ggplot2* (Wickham 2016).

To enhance the flexibility and quality of Ribo-seq data visualization, here we introduce *ggRibo*, which extends the advantages of *RiboPlotR* while leveraging the capabilities of *ggplot2*. *ggRibo* aims to support broader species and input formats, visualize more types of unannotated ORFs, accommodate as many Ribo-seq/RNA-seq data sets as the computer's RAM allows, display DNA and amino acid sequences when necessary, and, most importantly, integrate other sequencing data types with single-nucleotide resolution, in which specific nucleotide positions convey sequence read information. Together, *ggRibo* enables parallel comparisons of multiomic data sets, facilitating the study of gene regulation at multiple levels and revealing novel biological insights. We provide examples of using *ggRibo* to visualize different types of noncanonical ORFs in humans, *Arabidopsis*, rice, and tomatos. Additionally, we present examples that demonstrate how *ggRibo* can be applied to gain an in-depth understanding of gene regulation using multiomics data in *Arabidopsis*.

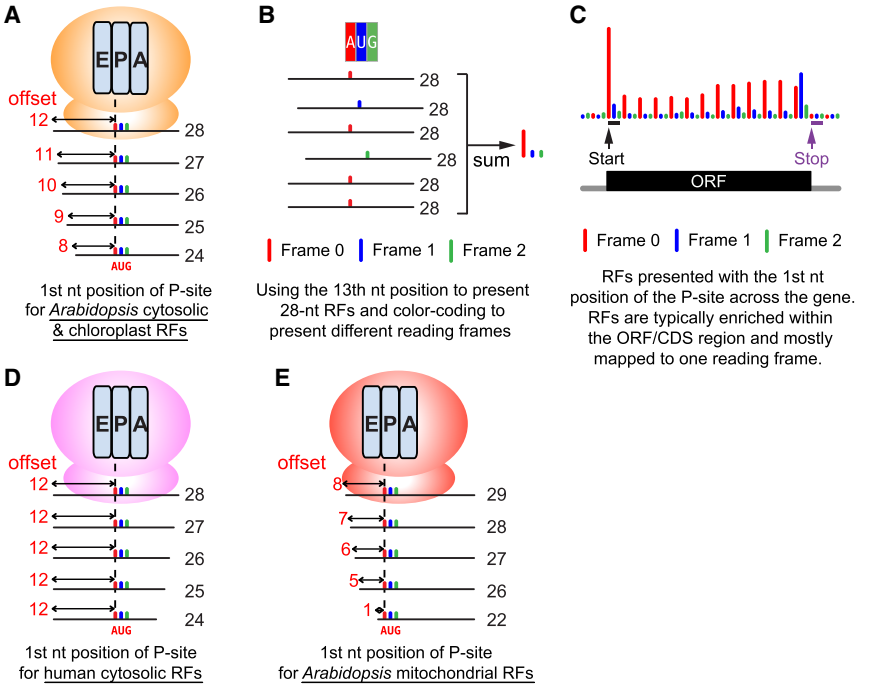


Figure 1. An illustration of how ribosome footprints (RFs) can be represented by a single nucleotide (nt) and color-coded to highlight 3 nt periodicity. (*A*) Metagene analysis by aligning all RFs mapped to the annotated start codon (AUG) can reveal the position of the P-site within ribosomes for each length of RFs. For *Arabidopsis* cytosolic and chloroplast RFs, the 13th nucleotide (offset by 12 nt from 5′) is typically the first nucleotide position of the P-site for a 28 nt RF. Similarly, the 12th nucleotide (offset by 11 nt from 5′) is the first nucleotide position of the P-site for a 27 nt RF. Note the P-site offsets could vary in different organisms and organelles (Ingolia et al. 2011; Hsu et al. 2016; Wu et al. 2024a). For high-quality *Arabidopsis* Ribo-seq data, ≤28 nt footprints exhibit excellent frame enrichment, making their P-site assignments reliable. In contrast, ≥29 nt footprints exhibit poorer frame enrichment (Wu et al. 2024a). Accordingly, for these *Arabidopsis* data, only ≤28 nt footprints are used for visualization, although longer footprints are still useful for other analyses. (*B*) An illustration showing how 28 nt RFs are mapped to a codon. The 28 nt RFs are presented by their 13th nucleotide position and color-coded by reading frame (annotated frame = frame 0, red; frame 1, blue; frame 2, green). Summing the reads mapped to each of the three nucleotides reveals the enrichment in frame 0 (red), which is expected in high-quality data sets. (*C*) Combining all RFs mapped to a particular gene produces a P-site plot, which visualizes the distribution of the RFs and colored 3 nt periodicity across the gene. At the codon prior to the stop codon, ribosome conformational change at the translation termination causes RFs to have a different length and results in an enrichment of frame 1 (blue) (Wu et al. 2024a). (*D*) The P-site offsets for human "cytosolic" RFs from induced pluripotent stem cells (iPSCs) and cardiomyocytes (Chen et al. 2020). Note that RNase digestion for different lengths of RFs occurs at the 3′ of the RFs in humans compared with the 5′ of the RFs in *Arabidopsis* (see *A*). (*E*) The P-site offsets for *Arabidopsis* "mitochondrial" RFs. The P-site offsets are shorter than the cytosolic and chloroplast RFs in *Arabidopsis* (see *A*).

Results

Design of *ggRibo*

Figure 1, A–C, illustrates how ribosome footprints can be represented by single-nucleotide positions (often the first nucleotide of the peptidyl site [P-site] within ribosomes) and color-coded by reading frames, enabling visualization of 3 nt periodicity along translated ORFs. P-site positions can be assigned using tools such as RiboTaper or *Ribo-seQC* (Calviello et al. 2016, 2019). Note that the P-sites for different ribosome footprint lengths may vary across organisms and organelles (Fig. 1A,D,E) and depend on the footprinting conditions (Ingolia et al. 2012; Bazzini et al. 2014; Hsu et al. 2016; Chen et al. 2020; Wu et al. 2024a).

To enhance the visual interpretation of periodic Ribo-seq reads, we developed *ggRibo* (https://github.com/hsinyenwu/ggRibo), a single-gene viewer that displays Ribo-seq periodicity alongside RNA-seq coverage in the context of the gene structure, including 5′/3′ UTRs and introns from multiple isoforms (Fig. 2A). *ggRibo* supports all GTF and GFF annotation files compatible with the *GenomicRanges* package (Lawrence et al. 2013). *ggRibo* accepts common standard input files, such as bedGraph, bigWig, and BAM. Built on the *ggplot2* framework, *ggRibo* generates publication-quality plots. To simplify interpretation, *ggRibo* displays genes from left to right, regardless of whether they are on the Watson or Crick strand. Ribo-seq reads are plotted by the first nucleotide of the P-site and color-coded by reading frames (Figs. 1, 2A). Together, these features facilitate the detection and the detailed visualization of unannotated ORFs and noncanonical translation features.

Notably, *ggRibo* visualizes both annotated ORFs and extra ORFs (called "eORFs" in the *ggRibo* package) (Fig. 2B). It can also extend the main ORF reading frame into the 5′ UTR and 3′ UTR, allowing the visualization of alternative start sites (AUG or non-AUG) and stop codon readthrough, respectively (Fig. 2B). Additionally, *ggRibo* can plot Ribo-seq reads on annotated noncoding RNAs to visually inspect whether a potential small ORF is translated with strong 3 nt periodicity (Fig. 2C). A summary of *ggRibo*'s major improvements over its prototype, *RiboPlotR*, is described in Supplemental Figure S1.

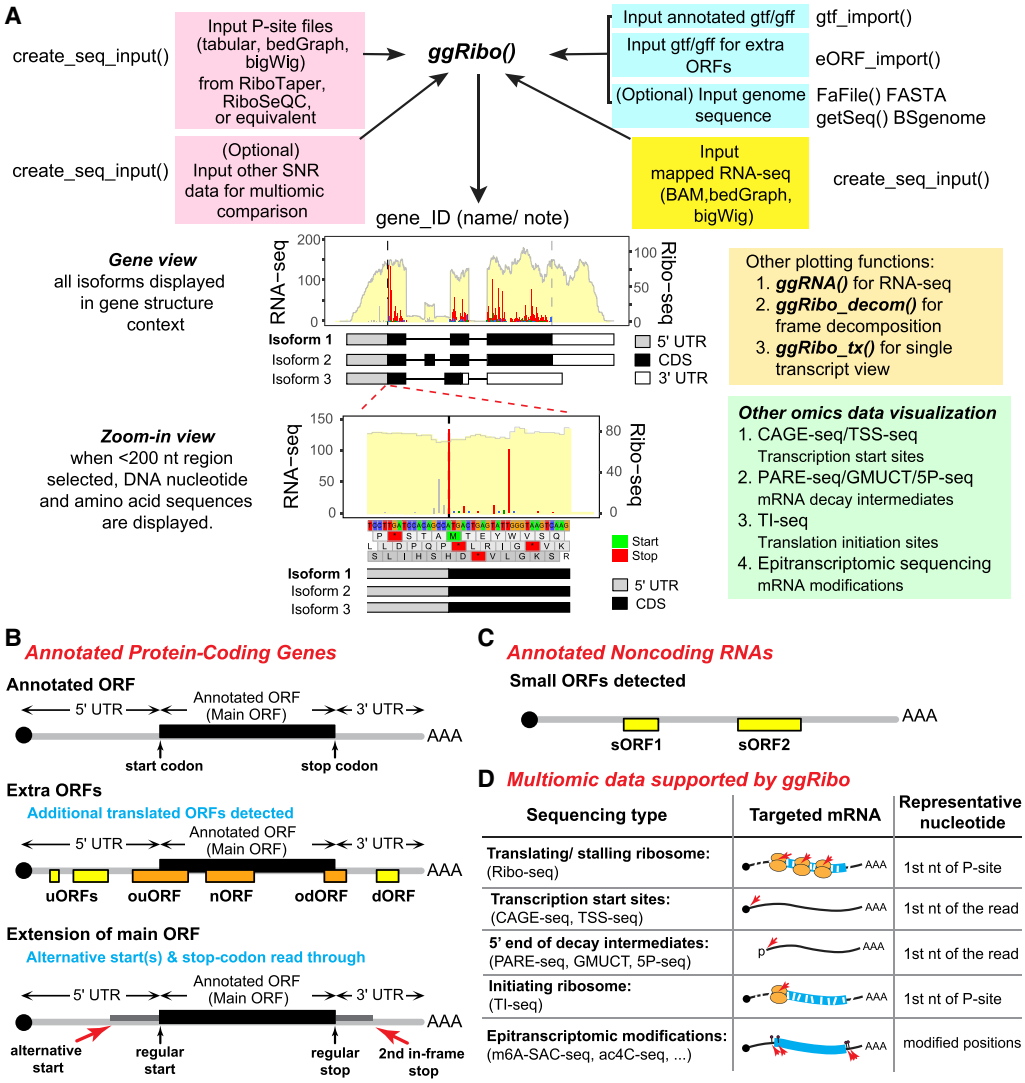


Figure 2. Overview of *ggRibo* and the ORF/data types supported. (*A*) Input/output and the functions of *ggRibo*. *ggRibo* requires Ribo-seq P-site information, genome annotation for annotated ORFs and (optional) extra ORFs, and mapped RNA-seq file(s). Other omics data presented with single-nucleotide resolution (SNR) can be input and visualized in parallel or independently. To visualize DNA and/or amino acid sequences, users can optionally provide the corresponding genome sequence file. Output plots can be in various scales, such as gene view or zoomed-in view. (*B*,*C*) ORF types can be visualized by *ggRibo*. Annotated genes are typically grouped into either coding (*B*) or noncoding (*C*), and *ggRibo* can plot ORFs in both gene types. For annotated protein-coding genes, typically only one ORF (main ORF) is annotated, whereas for annotated noncoding RNA genes, no ORF is annotated by default. By providing GTF/GFF files that include extra ORFs, users can visualize any ORF beyond those already annotated. "Extra ORFs" may be translated using a frame different from the main ORF and/or separate from the main ORF, including upstream ORFs (uORFs) within the 5′ UTR and downstream ORFs (dORFs) within the 3′ UTR. Some of the uORFs and dORFs may partially overlap with the main ORF (ouORFs and odORFs). Some overlapping ORFs may be nested within the main ORF (nORF). In contrast, "extension of main ORFs" uses the same reading frame as the main ORF and extends from 5′ or 3′. The fExtend and tExtend parameters allow extending the main ORF frames into the 5′ UTR and 3′ UTR, respectively. This enables investigating translation from alternative AUG or non-AUG start codons in the 5′ UTR or stop-codon readthrough events in the 3′ UTR. For annotated noncoding RNAs, some small ORFs (sORFs) may be translated. (*D*) Example omics data types with SNR that can be plotted by *ggRibo*.

ggRibo is a user-friendly R package that accepts a wide range of standard input files, is applicable to any species with a genome annotation, and offers improved flexibility for customization and easy automation compared with other existing tools, like GWIPS-viz, Trips-Viz, and RiboCrypt (Michel et al. 2018; Kiniry et al. 2019; https://ribocrypt.org/). GWIPS-viz and Trips-Viz are online genome browsers that display a collection of data for specific species and genome annotations, whereas RiboCrypt offers both online and user-installed genome browsers when used in conjunction with the ORFik package in R (Tjeldnes et al. 2021). In contrast, *ggRibo* can be easily integrated into any pipeline by offering simple functions for loading and visualizing data. Importantly, *ggRibo* provides an efficient presentation of gene structure, isoform expression/translation, and color-coded reading frame information simultaneously in high-quality figures (Supplemental Fig. S2A). A single transcript view option is also available using *ggRibo_tx* (Supplemental Fig. S3A; cf. other tools in Supplemental Fig. S3B,C). In contrast to *ggRibo*, GWIPS-viz plots Ribo-seq reads in a single color, which hinders the intuitive interpretation of reading frames (Supplemental Fig. S2B). Although

RiboCrypt offers color-coded reading frames in the "gene" view, the reading frame is assigned based on the genomic sequence, and therefore, the colors change as the exon sequences shift from one reading frame to another (Supplemental Fig. S2C). The change in color because of a reading frame shift makes it challenging to identify the ORF visually, which is crucial for studying multiple or overlapping ORFs in one transcript (see below) (Figs. 4E,F, 6A). Notably, *ggRibo* displays a consistent color for an ORF as the reading frame is assigned based on the coding sequence (CDS) (Supplemental Fig. S2A). Taken together, compared with existing tools, *ggRibo* offers intuitive visualization and superior flexibility for customization.

In addition to Ribo-seq/RNA-seq, *ggRibo* supports the visualization of other sequencing data that are presented with single-nucleotide resolution (Fig. 2D). Examples include CAGE-seq and TSS-seq for studying transcription start sites (TSSs) (Wakaguri et al. 2008; Valen et al. 2009); PARE-seq, GMUCT, and 5P-seq for studying mRNA decay intermediates (German et al. 2008; Gregory et al. 2008; Pelechano et al. 2015); TI-seq for studying translation initiation sites (Ingolia et al. 2011; Lee et al. 2012; Willems et al. 2017; Li and Liu 2020); and m6A-SAC-seq and ac4C-seq for studying epitranscriptomic modifications (Sas-Chen et al. 2020; Hu et al. 2022).

ggRibo package extends its capabilities beyond the *ggRibo* function by including three specialized tools, each tailored for specific visualization tasks: *ggRNA*, *ggRibo_decom*, and *ggRibo_tx*.

ggRNA: RNA-seq data visualization

ggRNA plots RNA-seq data alone when only RNA-seq coverage is needed (Supplemental Fig. S4A). Like the *ggRibo* function, it supports multiple samples and accommodates both paired-end and single-end RNA-seq data, offering flexibility for various experimental designs.

ggRibo_decom: reading frame decomposition

ggRibo_decom plots Ribo-seq reads from different reading frames separately for analyzing frame-specific distributions (Supplemental Fig. S4B). One data set is processed at a time for in-depth examination.

ggRibo_tx: single-transcript visualization

ggRibo_tx visualizes RNA-seq and Ribo-seq data in the context of mature RNA for a selected transcript isoform (Supplemental Fig. S4C), which is useful for genes with long or numerous introns. To prevent misinterpretation from incorrect isoforms, users should first identify expressed isoforms using *ggRibo* or *ggRNA* before employing *ggRibo_tx*. Similar to the *ggRibo* function, *ggRibo_tx* displays both RNA-seq and Ribo-seq data from multiple samples in a single plot.

Applications and further information

Below, we present examples demonstrating how *ggRibo* and its specialized functions can be used to explore unannotated ORFs and noncanonical translation features within the translatome, as well as how it can integrate multiomics data sets to enhance our understanding of gene regulation. A tutorial detailing the setup, requirements, and example code of the *ggRibo* package is available on the *ggRibo* GitHub site.

Visualizing translated ORFs using human and rice data

To demonstrate *ggRibo*'s basic usage, we first provide an example in the human *MRPL11* gene (Fig. 3), which contains a uORF in the 5′ UTR that represses downstream main ORF translation (Calvo et al. 2009). Figure 3A presents the Ribo-seq and RNA-seq data of *MRPL11* in human embryonic stem cells (Chothani et al. 2022), displaying all annotated RNA isoforms, with the uORF highlighted in yellow in the gene models. Figure 3B adds an amino acid track with start and stop codon positions highlighted. Figure 3C zooms in on the *MRPL11* uORF, displaying both DNA nucleotide and amino acid sequence tracks.

To focus on the expressed isoforms, we next use *ggRibo_tx* to visualize *MRPL11*. Figure 3A suggests that the main expressed isoform is ENST00000310999. Therefore, we plot this isoform with *ggRibo_tx* in Figure 3D. To demonstrate the importance of visualizing the data within the gene context, we also plot the isoform ENST00000329819, which is expressed and translated at low levels relative to ENST00000310999, as shown by the small number of periodic reads in exon 5 of ENST00000329819 (Fig. 3E).

Next, we examine the human immune-related gene interferon induced transmembrane protein 3 (*IFITM3*), which contains isoforms with and without an annotated CDS, using human brain data (Chothani et al. 2022). *ggRibo* plot reveals that ENST00000399808 is the major expressed and translated isoform (for a gene view displaying all RNA isoforms, see Supplemental Fig. S5A; for a zoom-in view of a subset of isoforms, see Supplemental Fig. S5B). This example again highlights how visualizing RNA-seq and Ribo-seq reads in the context of gene structure helps identify expressed and translated isoforms.

Similarly, we provide an example in rice to visualize a uORF encoding a conserved peptide in the *AdoMetDC* gene (Franceschetti et al. 2001; Hayden and Jorgensen 2007) using *ggRibo* and *ggRibo_tx plots* (Supplemental Fig. S6) with the data reported by Yang et al. (2021). The striking difference in Ribo-seq levels between the uORF and the main ORF suggests that the uORF almost completely suppresses the translation of the main ORF (Supplemental Fig. S6).

Visualizing translation of unannotated ORFs and within noncoding RNAs in *Arabidopsis* and tomato

We next present four examples to showcase how *ggRibo* facilitates the identification and visualization of novel isoforms, unannotated ORFs, as well as translation detected within noncoding RNAs using *Arabidopsis* and tomato data (Wu et al. 2019, 2024a).

First example: translation of an unannotated isoform

PAE7 encodes a pectin acetylesterase and has two annotated RNA isoforms in *Arabidopsis*. RNA-seq coverage indicates that both isoforms are expressed, and the Ribo-seq periodicity supports that both isoforms are translated (Fig. 4A). Additionally, cross-referencing the RNA-seq profile with gene models reveals that exon 14 of isoform 2 is longer than the annotated range (Fig. 4A, blue dashed box). This finding further demonstrates how gene-structure-based visualization aids in interpreting isoform translation and refining gene annotations.

Second example: translated minimum uORF (AUG-STOP)

SnRK2.2 encodes a key regulator of plant abiotic stress responses, and previous research identified a minimal uORF (AUG-STOP, the shortest possible ORF) in this gene (Wu et al. 2024a). Here,

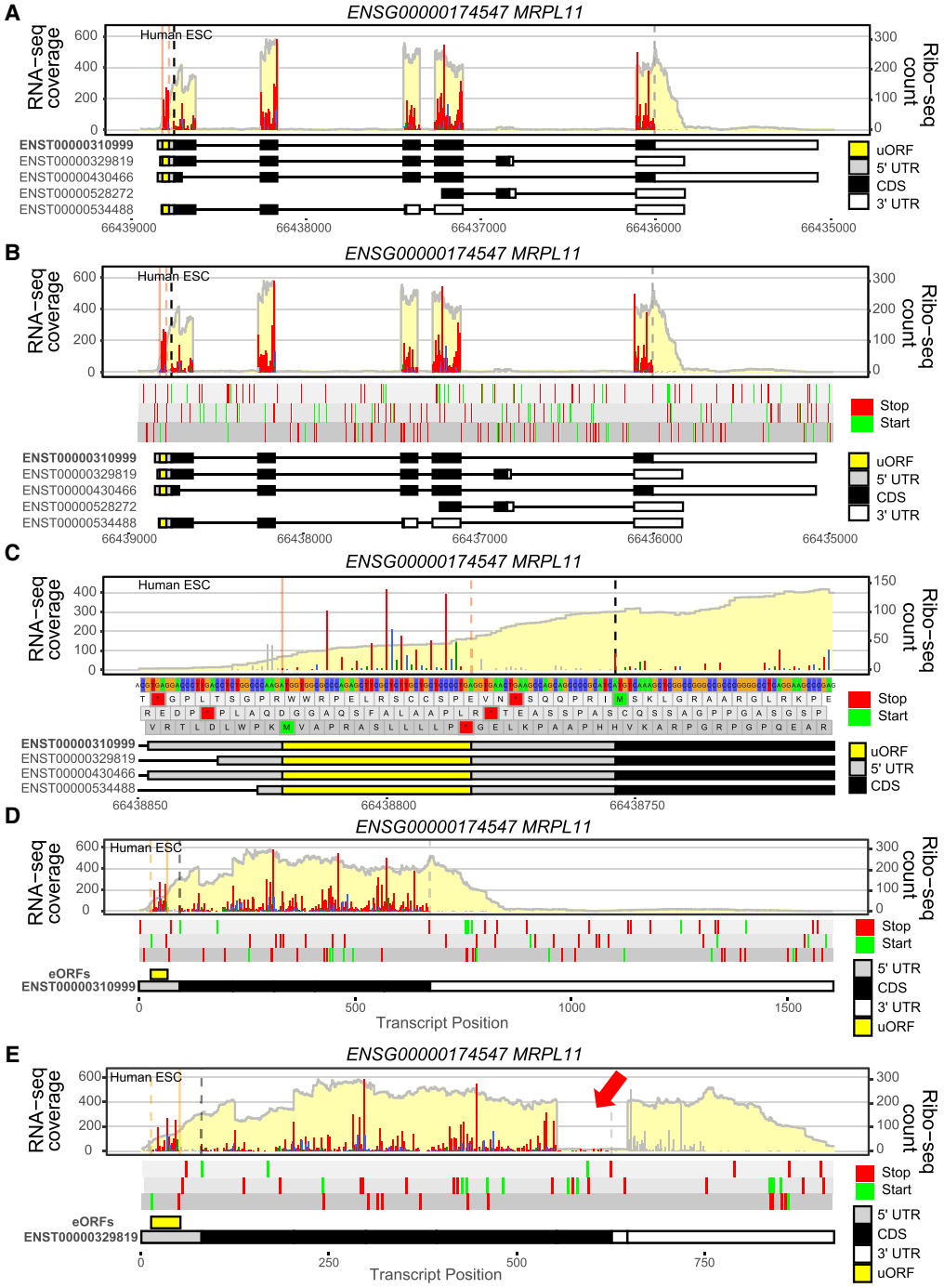


Figure 3. Examples of *ggRibo* and *ggRibo_tx* plots for the human *MRPL11* gene. Different display options using *ggRibo* to present RNA-seq and Ribo-seq profiles of the human nuclear-encoded *MRPL11* gene. The human embryonic stem cell (ESC) Ribo-seq and RNA-seq data are from Chothani et al. (2022). RNA-seq coverage is shown with a light-yellow background. Ribo-seq reads are presented with their first nucleotide of the P-site, and they are color-coded in red, blue, and green to indicate they map to reading frames 0 (expected), 1, and 2, respectively. *MRPL11* possesses five annotated transcript isoforms, and both a uORF and the main ORF are translated. The user can specify which transcript isoform to focus on and can assign the reading frame according to its gene structure; here, ENST00000310999 is selected (bolded). Within the gene models, black boxes indicate the coding sequence (CDS) (i.e., the annotated/main ORF); yellow box(es) indicate the uORF region; and gray and white boxes indicate 5′ UTR and 3′ UTR, respectively. Black and gray vertical dashed lines represent the start and stop codons, respectively, for the annotated mORF. Orange solid and dashed vertical lines represent start and stop codons, respectively, for the uORF. Reads outside of the ORF range are shown in gray. Genomic or transcript coordinates on the chromosomes are indicated *below* the gene models. (*A*) Only Ribo-seq and RNA-seq profiles are presented on *top* of the annotated gene models. (*B*) An additional DNA sequence track was added, but only the start and stop codons are displayed owing to the large span (>201 nt). (*C*) The view is further zoomed in on the uORF region, showing both the DNA nucleotide and amino acid sequence tracks in detail (when length <201 nt). Note that the line width of the Ribo-seq reads could be modified by the "*ribo_linewidth*" parameter, and the sequence color may be modified by the "*nucleotide_color_scheme*" parameter. (*D*) Single-transcript plot by *ggRibo_tx* to show the RNA-seq and Ribo-seq reads that map to the main expressed isoform ENST00000310999. (*E*) Similar to *D*, but for a very lowly expressed isoform ENST00000329819 in this data set (see *A*). The arrow highlights exon 5, which contains a small number of Ribo-seq reads with 3 nt periodicity.

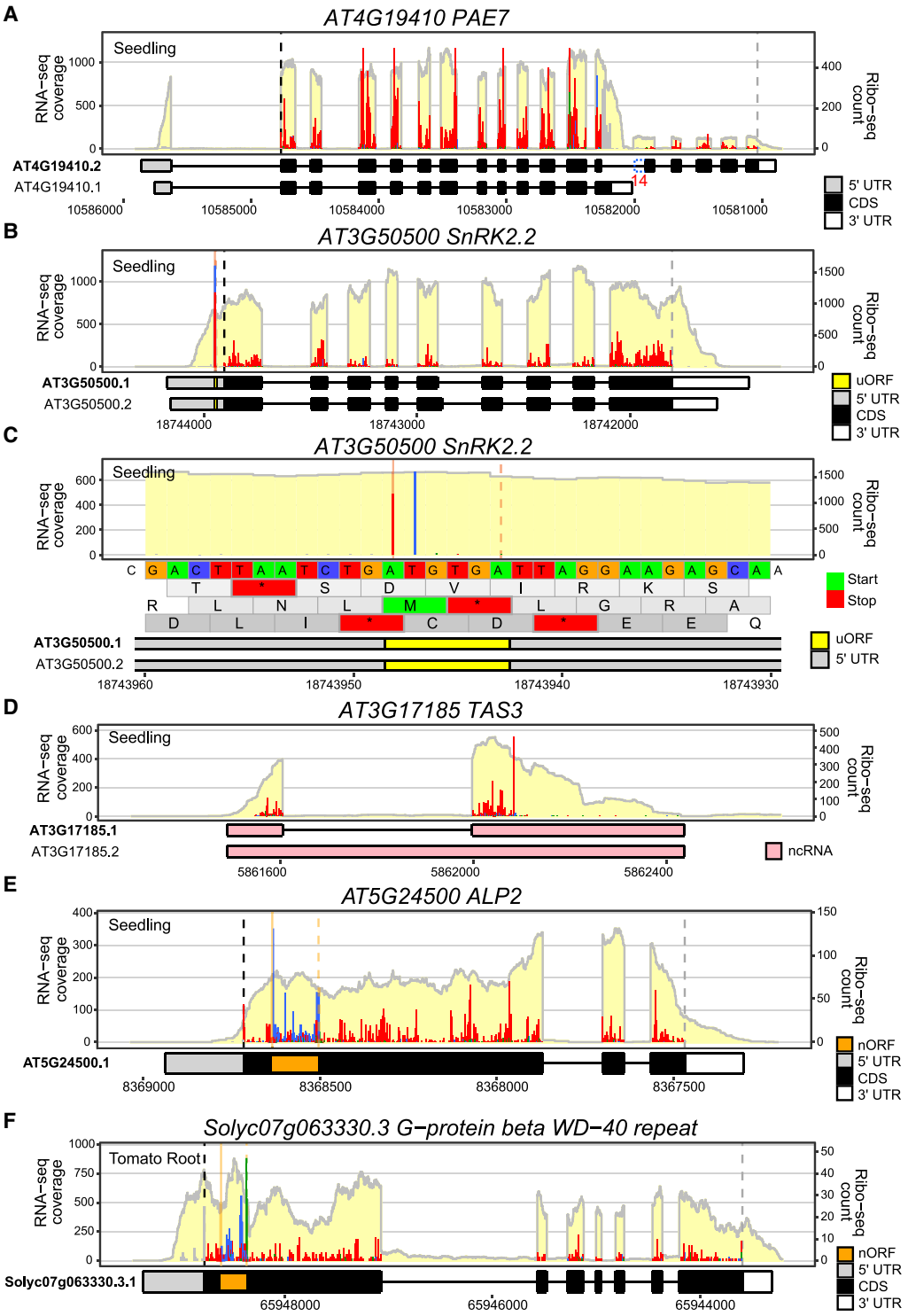


Figure 4. Examples of *ggRibo* plots with *Arabidopsis* and tomato genes. The *Arabidopsis* seedling Ribo-seq and RNA-seq data are from Wu et al. (2024a). The tomato Ribo-seq and RNA-seq data are from Wu et al. (2019). Data presentations are as described in the Figure 3 legend. (*A*) Gene view *of PAE7* shows that both isoforms are translated. Note that the RNA-seq coverage reveals that the 14th exon (highlighted) is longer than the annotated exon range for this isoform. The inferred exon 14 extension is shown with a blue dashed box (added with Adobe Illustrator). (*B*,*C*) A gene view and zoom-in view show the strong minimum uORF (i.e., AUG-stop) in *SnRK2.2*. In *C*, the Ribo-seq reads precisely map to the "A" and "T" of the start codon. Orange solid and dashed vertical lines represent the translation start and stop, respectively, for the uORF. (*D*) A small ORF is translated within *TAS3*, an annotated noncoding RNA (ncRNA). Robust 3 nt periodicity is detected in the 50-aa small ORF region. (*E*) A nested ORF (nORF) using the blue reading frame is detected in the *Arabidopsis ALP2* gene. The nORF range is highlighted with an orange box in the gene model. (*F*) A nORF using the blue reading frame is detected in a tomato G protein-coupled receptor beta subunit gene. The nORF range is highlighted with an orange box in the gene model.

we plot the entire *SnRK2.2* gene (Fig. 4B) and a zoomed-in view with the DNA nucleotide and amino acid sequence tracks centered on the uORF (Fig. 4C). The Ribo-seq reads for this minimal uORF are predominantly mapped to the first two nucleotides, A (frame 0, red) and U (frame 1, blue) (Fig. 4C). This read distribution, which is uniquely associated with minimal uORFs, results from the altered ribosome footprint length at the codon preceding the stop codon (Wu et al. 2024a).

Third example: translation of noncoding RNA

TAS3 is an annotated noncoding RNA, encoding the primary transcript of *trans*-acting short-interfering RNA (*tasiRNA*). A small ORF is known to be translated within *TAS3* primary transcript, a process that regulates tasiRNA biogenesis (Hsu et al. 2016; Li et al. 2016; Bazin et al. 2017; Hsu and Benfey 2018; Iwakawa et al. 2021; Wu et al. 2024a). Using *ggRibo*, we can observe that Ribo-seq reads exhibit strong 3 nt periodicity at the region corresponding to this small ORF (Fig. 4D; for zoom-in view showing the sequences near the start and stop codons, see Supplemental Fig. S7). Further analysis using *ggRibo_decom*, which plots three-frame reads separately, confirms that the Ribo-seq reads are predominantly enriched in frame 0 (Supplemental Fig. S8).

Notably, for noncoding RNAs, *ggRibo* assigns the reading frame from the first nucleotide of the annotated RNA sequence, rather than from the start of an ORF, as it does for coding RNAs. As a result, a translated ORF in a noncoding RNA may use one of the reading frames in red, blue, or green. Supplemental Figure S9 shows an example of a small ORF enriched in the "blue" reading frame from an annotated long noncoding RNA (lncRNA) in *Arabidopsis*.

Fourth example: ORFs within ORFs, visualizing overlapping/nested translation events

We next show two examples of nested ORFs, in which two ORFs using different reading frames are both translated, and the smaller ORF is embedded within the longer one. Although this phenomenon was discovered decades ago in bacterial phage ϕX174 (Barrell et al. 1976; Sanger et al. 1977), it is challenging to detect such nested ORFs in eukaryotes (Wright et al. 2022). Here, we attempt to visualize such events using high-coverage superresolution Ribo-seq data with *ggRibo*. In the *Arabidopsis ALP2* gene, a nested ORF using the blue reading frame (corresponding to the orange box in the transcript model) is clearly visible (Fig. 4E). Similarly, in the tomato *Solyc07g063330.3* gene, using data from Wu et al. (2019), a nested ORF using the blue reading frame is also clearly visible (Fig. 4F). Strong 3 nt periodicity in the nested ORF regions suggests that they are actively translated (Fig. 4E,F). Frame decomposition using *ggRibo_decom* further supports these findings (Supplemental Fig. S10A,B). The *ggRibo_decom* function is particularly useful for illustrating such frame enrichment for complex translation events.

Comparing differentially translated isoforms across data sets using *ggRibo*

ggRibo can visualize multiple Ribo-seq/RNA-seq data sets to compare differential translated isoforms and/or translation efficiency of genes of interest. Here, we present an example of tissue-specific isoform translation in *Arabidopsis* BETA CARBONIC ANHYDRASE 4 (*BCA4*), comparing the root, shoot, and whole seedling. *BCA4* plays a crucial role in gas exchange between plant leaves and the atmosphere, catalyzing the conversion of CO_2 to HCO_3^- and contributing to the stomatal CO_2 response (Hu et al. 2010, 2015). *BCA4* has three annotated isoforms; the *ggRibo* plot suggests that isoforms 2 and 3 are expressed and translated in the tissues examined (Supplemental Fig. S11). Notably, although isoform 2 is highly expressed and translated in the shoot and whole seedling, it is nearly undetectable in the root. Instead, the root mainly expresses the isoform 3 (Supplemental Fig. S11). Recent work has shown that the isoform 2 protein (the longer isoform) localizes to the plasma membrane, whereas the isoform 3 protein (the shorter isoform) localizes to the cytosol (Weerasooriya et al. 2024). This example highlights *ggRibo*'s capacity to compare multiple data sets, enabling investigations into differential translation or isoform-specific translation across tissues/conditions and providing valuable insights for future functional characterization.

Multiomic data comparisons using *ggRibo*

Besides visualizing Ribo-seq/RNA-seq profiles, *ggRibo* can integrate other omics data sets that offer single-nucleotide resolution (see Fig. 2D). To illustrate its utility in multiomic data comparison, we provide three examples below to showcase *ggRibo*'s capacity to reveal novel biological insights for complex gene expression regulation. Note that other single-nucleotide resolution data can be displayed either with a single color (for an example, see Fig. 5A) or with three colors to view the reading frame information (for examples, see Fig. 5B,C) by setting the "sample_color" parameter in the *ggRibo* function. Other omics data sets should be formatted similarly to Ribo-seq and input using the *create_seq_input* function.

First example: transcriptional start sites of LITTLE ZIPPER 1 affect its translation

LITTLE ZIPPERS (ZPR1 to ZP4) are the first microproteins studied in *Arabidopsis* (Wenkel et al. 2007; Kim et al. 2008). Microproteins are single-domain proteins that usually bind to their multidomain homologs and act as dominant negative regulators (Eguen et al. 2015; Bhati et al. 2021). Although ZPR1 was predicted to be a 93-aa peptide (Wenkel et al. 2007), it is annotated as 136 aa in both the TAIR10 and Araport11 annotations. To verify its actual translated peptide length, we examined *ZPR1* Ribo-seq profile using *ggRibo*. We found that most Ribo-seq reads for *ZPR1* locate downstream from the annotated AUG start (Fig. 5A, see Ribo-seq panel). A closer inspection revealed that the Ribo-seq reads mostly start at the next in-frame AUG, that is, from the ORF in the *AT2G45450.x* isoform we defined (Fig. 5A, see black arrowheads in Ribo-seq panel and the start/stop codon track), supporting that *ZPR1* is predominantly translated into a 93 aa peptide.

Why does *ZPR1* primarily use the second AUG? Our recent study revealed that TSSs in *Arabidopsis* can be highly variable, affecting ORF translation initiation sites (Wu and Hsu 2024). Accordingly, some transcripts may begin transcription downstream from the annotated start codon (AUG), thereby excluding the first AUG from the mature mRNA. To test this possibility, we examined available TSS-seq data, which identify TSSs (Nielsen et al. 2019), and compared them with the Ribo-seq data and the gene model (Fig. 5A). Consistent with our hypothesis, the TSS-seq signals for *ZPR1* occur primarily downstream from the annotated start codon, explaining why *ZPR1* translation preferentially initiates at the second AUG (Fig. 5A, TSS-seq panel). This demonstrates how *ggRibo*'s capacity to integrate multiomic data sets facilitates a deeper understanding of gene regulation.

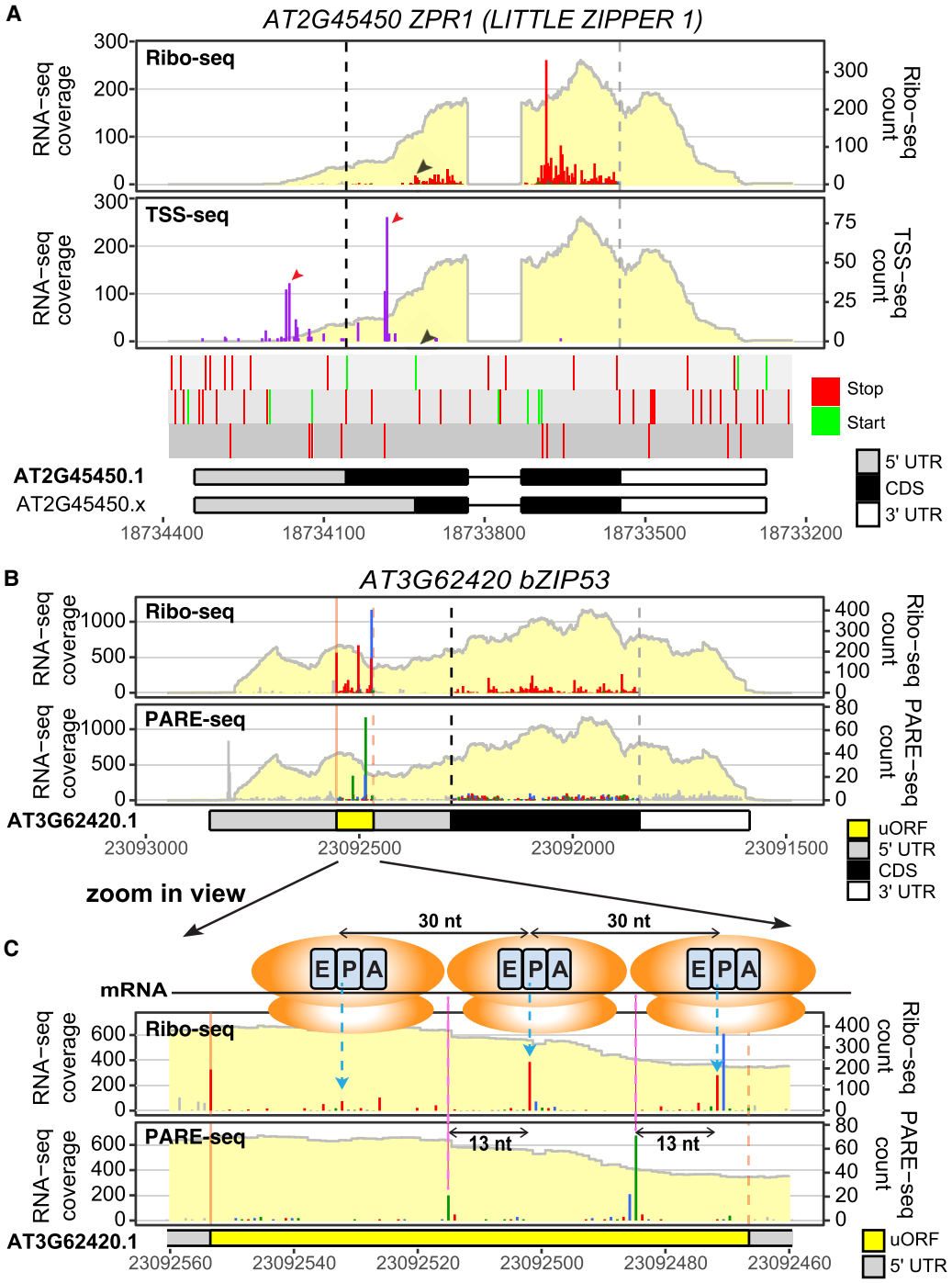


Figure 5. Multiomic comparisons reveal novel insights connecting the transcriptome, translatome, and degradome. (*A*) A gene view showing Ribo-seq/RNA-seq (Wu et al. 2024a) and TSS-seq (Nielsen et al. 2019) profiles for AT2G45450 *LITTLE ZIPPER 1 (ZPR1)*. The Ribo-seq panel confirmed the translation of *ZPR1*, but the translation does not initiate at the annotated start codon (black dashed line); rather, reads mostly begin at the next in-frame AUG (black arrowhead in the Ribo-seq panel and start/stop codon track). TSS-seq (purple lines) reveals two major transcription start site (TSS) clusters (red arrowheads): one upstream of and one downstream from the annotated start codon. The downstream TSS cluster will cause the translation to start at the second AUG and result in the 93-aa peptide. The color for TSS-seq data can be modified using the *sample_color* parameter in *ggRibo*. Ribo-seq and RNA seq data are reported by Wu et al. (2024a), and TSS-seq data are reported by Nagarajan et al. (2019). (*B*) A gene view showing Ribo-seq/RNA-seq (Wu et al. 2024a) and PARE-seq (Nagarajan et al. 2019) in parallel for the *bZIP53* gene; RNA-seq coverage is shown in a light-yellow background for each panel. The PARE-seq are also color-coded to compare the reading frame relationship with the Ribo-seq data. (*C*) A zoomed-in view of the conserved peptide uORF region. Accumulation of Ribo-seq reads at a 30 nt interval (highlighted by blue dashed arrows) upstream of the uORF stop codon (orange dashed line) suggests that multiple ribosomes stack in this region. Accumulation of the PARE-seq reads (highlighted by pink dashed lines) occurs at 13 nt upstream of the corresponding Ribo-seq peaks, supporting cotranslational mRNA decay in this region. The shift between the red/blue reading frames prior to the stop codon in Ribo-seq is also observed in PARE-seq (blue/green). Ribosomes, blue dashed arrows, and pink dashed lines were added in Adobe Illustrator.

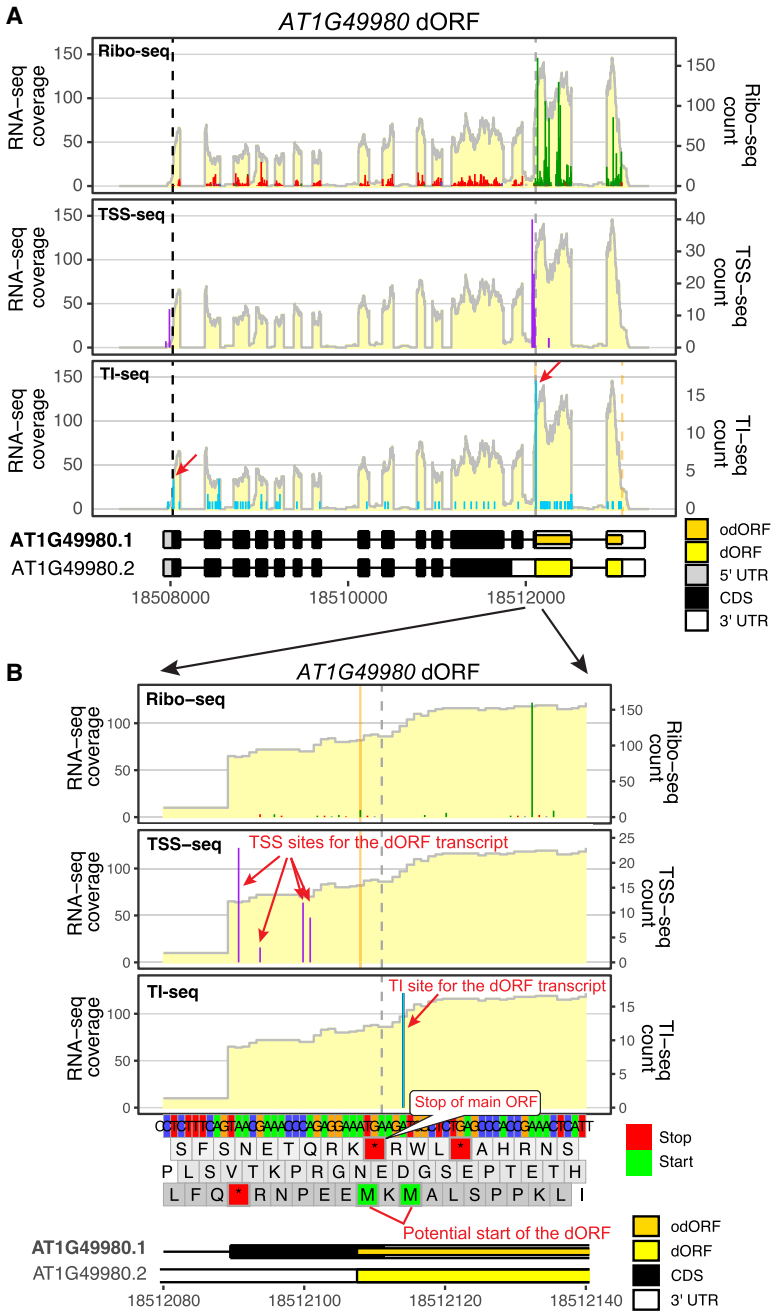


Figure 6. Multiomic comparison reveals the nature of a translated "dORF." Ribo-seq and RNA-seq data are reported by Wu et al. (2024a); TSS-seq data are reported by Nielsen et al. (2019); and TI-seq data are reported by Willems et al. (2017). (*A*) A gene view showing Ribo-seq, TSS-seq, and TI-seq in parallel for *AT1G49980*; RNA-seq coverage is shown in the background for each panel. The Ribo-seq panel confirms a potential dORF translated in the 3′ UTR using frame 2 (green), which was previously identified by RiboTaper (Calviello et al. 2016). TSS-seq panel reveals two major TSS clusters (purple lines); one corresponds to the annotated TSS for *AT1G49980*, and the other is upstream of this dORF. The existence of a TSS upstream of the dORF supports this dORF arising from an unannotated gene/transcript, which has higher RNA levels than *AT1G49980*. (*B*) A zoomed-in view focusing on the 5′ of the dORF using the same data above. The gray vertical dashed line indicates the stop codon of the main ORF, and the orange solid line indicates the RiboTaper-predicted start codon for the dORF, which corresponds to the first of the two adjacent ATG starts (indicated by two methionine amino acids). The TI-seq data reveal that this dORF exclusively uses the second start codon, and therefore, this ORF does not overlap with the main ORF stop codon. Note that RiboTaper predicted this dORF to overlap with the main ORF of isoform 1 but not the main ORF of isoform 2. Therefore, in isoform 1, this dORF is considered an overlapping dORF (colored orange in the isoform 1 gene model), whereas in isoform 2, it is considered a distinct dORF (colored yellow in the isoform 2 gene model). The red arrows and comments were added in Adobe Illustrator.

Second example: cotranslational mRNA decay observed in a conserved peptide uORF

Cotranslational mRNA decay is a process in which mRNA is degraded while still being translated or occupied by ribosomes. It is often detected at conserved peptide uORFs, in which significant ribosome stalling occurs upstream of the stop codon (Hou et al. 2016; Yu et al. 2016; Guo et al. 2023). Here, we present an example in the *bZIP53* transcription factor gene, comparing Ribo-seq/RNA-seq profiles with PARE-seq data (Nagarajan et al. 2019) to evaluate potential mRNA decay associated with ribosome stalling (Fig. 5B). PARE-seq detects the 5′-monophosphate end, which is a common mRNA decay intermediate (German et al. 2008). As noted in previous studies, Ribo-seq and PARE-seq suggest that ribosomes stall at ~30 nt intervals (the estimated width of a ribosome footprint) from the end of the uORF (Fig. 5C). Meanwhile, PARE-seq data display corresponding peaks 13 nt upstream of the Ribo-seq peaks (Fig. 5C). This pattern aligns with the expectations for cotranslational decay, in which ribosomes act as a protective barrier against nucleolytic degradation, causing mRNA decay to stop at the 5′ boundary of the ribosome (Hou et al. 2016; Yu et al. 2016). Additionally, the shift between red and blue reading frames in the Ribo-seq data prior to the stop codon (Fig. 5C, Ribo panel) (Wu et al. 2024a) is also reflected in the PARE-seq data (blue/green) (Fig. 5C, PARE panel). Thus, visualization of multiomic data using *ggRibo* allows in-depth analysis of gene regulatory mechanisms.

Third example: a dORF that potentially overlaps with the main ORF

We previously reported that *AT1G49980*, which encodes a DNA/RNA polymerase family protein, harbors a strongly translated dORF in the 3′ UTR, using an ORF discovery tool, RiboTaper (Calviello et al. 2016; Wu et al. 2024a). RiboTaper classified this dORF as an overlapping dORF in the *AT1G49980.1* isoform, overlapping with the stop codon of the main ORF (for gene models, see Fig. 6A). Plotting *AT1G49980* by *ggRibo* confirmed a highly translated ORF within the annotated 3′ UTR (Fig. 6A, Ribo-seq panel, green reading frame). However, the higher RNA coverage in the dORF region compared with the main ORF region suggests that this dORF arises from

a separate, unannotated gene or transcript. To test this possibility, we examined TSSs reported in TSS-seq data (Nielsen et al. 2019) by *ggRibo*. We found two distinct TSS clusters: one near the annotated TSS and the other upstream of the dORF (Fig. 6A,B, TSS-seq panel, purple lines). This pattern suggests the presence of an unannotated gene/transcript downstream from *AT1G49980*, which is supported by an independent study (Thieffry et al. 2020). Therefore, this apparent dORF is more likely a novel ORF from an unannotated gene, rather than a dORF of *AT1G49980*.

Further analysis of the dORF's DNA sequence reveals two potential ATG start codons: one overlapping with the main ORF's stop codon and the other located two codons downstream (Fig. 6B, bottom panel). To determine which start codon is used by this dORF, we examined available TI-seq data (Willems et al. 2017), a technique similar to Ribo-seq but enriching for initiating ribosomes, which assists in the identification of translation initiation sites (Ingolia et al. 2011; Lee et al. 2012; Willems et al. 2017; Li and Liu 2020). The TI-seq data show two major peaks: the first one near the annotated start codon of *AT1G49980* and the second one at the dORF region (Fig. 6A, TI-seq panel, arrows). Zooming in on the 5′ region of the dORF revealed that this dORF exclusively uses the second ATG (Fig. 6B, TI-seq panel), indicating that this translated dORF does not overlap with the *AT1G49980* main ORF. This example again highlights *ggRibo*'s capacity to integrate multiomic data with unparalleled resolution, which helps reveal novel gene regulatory mechanisms for future studies.

Discussion

In omics studies, the visualization of sequencing data within genes of interest remains underutilized owing to a lack of appropriate tools. To address this gap, we developed *ggRibo*, which enables in-depth exploration of translational landscapes under various experimental conditions and facilitates multiomics comparisons, providing deeper insights into translational regulation and different levels of gene regulation.

ggRibo and its associated plotting functions offer several advantages over web-based tools. First is flexible visualization: Users can input and visualize RNA-seq, Ribo-seq (which exhibits 3-nt periodicity), and any other single-nucleotide resolution data together for genes of interest with flexible scaling. Other single-nucleotide resolution sequencing data and RNA-seq data can also be plotted alone, independent of Ribo-seq. Second is batch plotting: Users can plot hundreds of genes at once to visualize their genes of interest. For example, one could quickly compare uORF translation under different conditions across hundreds of uORF-containing genes. The same task might take days using genome or transcript browsers, which require manual copying and pasting for each gene. Third is publication-ready output: The *ggRibo* output can be saved as in PDF, SVG, PNG, or any other file format compatible with *ggplot2*'s ggsave function for easy data management and editing of figures for publication. Fourth is custom ORF input: Users can input a GTF file to include genomic coordinates of additional ORFs (such as uORFs and dORFs) to facilitate the visualization of noncanonical ORFs. Fifth is isoform confirmation: The gene view in *ggRibo* allows users to identify translated isoforms, whether annotated or unannotated. Relying solely on annotated isoforms in single-transcript viewers might overlook translation occurring on unannotated transcripts. Sixth is availability: Online tools may be restrained by server capacity and only maintained as funding permits, but users can always use R and *ggRibo*.

In summary, with its flexibility and simple workflow, *ggRibo* is a powerful tool for multiomics analysis. It helps researchers efficiently visualize individual gene expression from diverse sequencing data, produce publication-quality plots, synthesize hypotheses for future studies, and develop a more comprehensive understanding of gene regulation.

Methods

The *ggRibo* package provides a suite of functions for processing and visualizing Ribo-seq and RNA-seq data, facilitating ORF analysis within the gene-structure or transcript context. All data files are imported and stored as R6 objects, leveraging object-oriented principles for efficient data handling. The package now supports a broader range of input formats, including bedGraph and bigWig, enhancing its compatibility with diverse sequencing data sets. The workflow and example code are available on the *ggRibo* GitHub page. The design, function, and basic steps for *ggRibo* visualization are summarized below.

Annotation import

Annotation files are imported using the gtf_import() and eORF_import() functions:

- gtf_import()—imports GTF/GFF annotation files, creating a Range_info object that extracts genomic features such as exons, transcripts, CDSs, and UTRs. This object serves as a foundational data set for subsequent analyses.
- eORF_import()—imports user-defined extra ORF (eORF) annotations, creating an eORF_Range_info object to accommodate nonstandard ORFs not included in standard annotations.

Data import

For data import, the *create_seq_input()* function is introduced alongside the existing *Ribo_data()* function:

- *create_seq_input()*—constructs input lists for RNA-seq and Ribo-seq data by automatically detecting file formats (BAM, bigWig, bedGraph) and organizing the data into a structured format for downstream analysis. It supports both paired-end and single-end RNA-seq data and accommodates various Ribo-seq file types, enhancing the package's flexibility.
- *Ribo_data()*—reads Ribo-seq data files (or other single-nucleotide resolution data) and organizes them into a list of data frames, each containing read counts, chromosome numbers, positions, and strand information, enabling efficient manipulation and analysis.

Visualization

The visualization capabilities are expanded with the following plotting functions:

- *ggRibo()*—is the core plotting function, visualizing Ribo-seq periodicity alongside RNA-seq coverage within the gene-structure context. It integrates transcript isoforms and genome annotations to highlight unannotated ORFs and unconventional translation events, now supporting bedGraph and bigWig input formats alongside BAM. It allows customization of aesthetics, scaling, and inclusion of additional sequencing data (e.g., TI-seq, degradome sequencing, CAGE-seq).
- *ggRNA()*—plots RNA-seq data alone, offering visualization of transcription levels independent of Ribo-seq data.
- *ggRibo_decom()*—generates a plot that decomposes Ribo-seq signals into three separate plots, one for each reading frame, providing insights into translation dynamics.

- *ggRibo_tx()*—plots RNA-seq and Ribo-seq coverage in transcript coordinates, offering a detailed view of translation events at the transcript level by mapping data to spliced exon sequences.

Basic steps for running *ggRibo*

The Basic steps for running *ggRibo* are as follows:

1. *gtf_import*—input genome annotation. Optional is using *eORF_import* to import the transcript and ORF ranges for eORFs recorded in GTF or GFF3 format.
2. *create_seq_input*—input RNA-seq files with *rna_files*, Ribo-seq files with *ribo_files*, and sample names with *sample_names*. *ggRNA* only takes in RNA-seq files in this step.
3. *ggRibo*—input a transcript id with *tx_id*. Alternatively, input a gene ID with *gene_id*, and the first transcript sorted by R will be used for *tx_id*. Users can input an eORF.tx_id that exists in the eORF GTF/GFF3 file. Similar parameters work for *ggRNA*, *ggRibo_decom*, and *ggRibo_tx*.

These functions utilize the *ggplot2* framework for high-quality, customizable plots, seamlessly integrating with the package's object-oriented design. Two other sets of functions work behind the scenes for frame assignment and track plotting.

Frame assignment

Several frame assignment functions are provided to assign translated reading frames accurately:

- *assign_frames()*—assigns reading frames to Ribo-seq reads based on provided CDS ranges, enabling precise determination of translated codons.
- *assign_frames_extended()*—extends annotated frame assignments into UTRs, accommodating overlapping eORFs for accurate frame calculations in complex regions.
- *assign_frames_with_extension()*—allows custom extensions into the 5′ and 3′ UTRs (via fExtend and tExtend parameters), offering flexibility for analyzing alternative translation start sites or stop codon readthrough events.
- *exclude_eORF_reads()*—filters out Ribo-seq reads overlapping with eORF regions from the main data set, preventing redundancy in visualization.

Functions for plotting the DNA sequence and gene model tracks

Additional functions enhance the visualization of DNA sequences and gene models:

- *plotDNAandAA()*—for genomic regions ≤201 nucleotides (via the plot_range parameter in *ggRibo(), ggRibo_decom(),* and *ggRNA()*)), generates plots displaying DNA nucleotides and corresponding amino acids, aiding in codon usage and reading frame analysis. For regions >201 nt, it highlights start codons (green) and stop codons (red) in each frame.
- *plotGeneTxModel()*—creates gene models illustrating exons, UTRs, CDS, and eORFs for specified genes and isoforms, adjusting for overlapping features and enhancing clarity with customized colors and labels.
- Two similar functions—*plotDNAandAA_tx()* and *plotGeneTx Model_tx()* perform identical functions as their parent functions for the *ggRibo_tx*() function.

The coding for the *ggRibo* package and parts of this Method section was assisted by the ChatGPT model o1 and Grok v3, with initial code adapted from a modified version of *RiboPlotR* (using base R plots) and transformed into *ggplot2* by AI tools and HLW.

Software availability

ggRibo is available on GitHub (https://github.com/hsinyenwu/ggRibo). The source code is also provided in Supplemental Code. The example data are stored on Mendeley Data: *Arabidopsis* (https://data.mendeley.com/datasets/wm6cs4zbtw/1), rice (https://data.mendeley.com/datasets/gtvnp8228c/1), and human (https://data.mendeley.com/datasets/m3t293k4wr/1).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work used computational resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University. This work was supported by a predoctoral training award under grant number T32-GM110523 from the National Institute of General Medical Sciences of the National Institutes of Health (NIH) to I.D.K. and research grants from the National Science Foundation (NSF) under award numbers 2425390 and 2051885 and the National Institute of General Medical Sciences of the NIH under award number R35GM155375 to P.Y.H. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or NIH.

Author contributions: H.L.W. developed the *ggRibo* package and tested it with *Arabidopsis* and tomato data. I.D.K. tested *ggRibo* with human and rice data and fixed several bugs. H.L.W. and I.D.K. prepared the GitHub tutorial. H.L.W. and P.Y.H. wrote the paper with input from I.D.K.

References

Barrell BG, Air GM, Hutchison CA. 1976. Overlapping genes in bacteriophage phiX174. *Nature* **264:** 34–41. doi:10.1038/264034a0

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. *Proc Natl Acad Sci* **114:** E10018–E10027. doi:10.1073/pnas.1708433114

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. *EMBO Journal* **33:** 981–993. doi:10.1002/embj.201488411

Bhati KK, Dolde U, Wenkel S. 2021. Microproteins: expanding functions and novel modes of regulation. *Mol Plant* **14:** 705–707. doi:10.1016/j.molp.2021.01.006

Brar GA, Weissman JS. 2015. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* **16:** 651–664. doi:10.1038/nrm4069

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. *Nat Methods* **13:** 165–170. doi:10.1038/nmeth.3688

Calviello L, Sydow D, Harnett D, Ohler U. 2019. Ribo-seQC: comprehensive analysis of cytoplasmic and organellar ribosome profiling data. bioRxiv doi:10.1101/601468

Calviello L, Hirsekorn A, Ohler U. 2020. Quantification of translation uncovers the functions of the alternative transcriptome. *Nat Struct Mol Biol* **27:** 717–725. doi:10.1038/s41594-020-0450-4

Calvo SE, Pagliarini DJ, Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci* **106:** 7507–7512. doi:10.1073/pnas.0810916106

Chen J, Brunner A-D, Cogan JZ, Nuñez JK, Fields AP, Adamson B, Itzhak DN, Li JY, Mann M, Leonetti MD, et al. 2020. Pervasive functional translation of noncanonical human open reading frames. *Science* **367:** 1140–1146. doi:10.1126/science.aay0262

Chothani SP, Adami E, Widjaja AA, Langley SR, Viswanathan S, Pua CJ, Zhihao NT, Harmston N, D'Agostino G, Whiffin N, et al. 2022. A

high-resolution map of human RNA translation. *Mol Cell* **82:** 2885–2899.e8. doi:10.1016/j.molcel.2022.06.023
Chothani S, Ho L, Schafer S, Rackham O. 2023. Discovering microproteins: making the most of ribosome profiling data. *RNA Biol* **20:** 943–954. doi:10.1080/15476286.2023.2279845
Choudhary S, Li W, D Smith A. 2020. Accurate detection of short and long active ORFs using Ribo-seq data. *Bioinformatics* **36:** 2053–2059. doi:10.1093/bioinformatics/btz878
Eguen T, Straub D, Graeff M, Wenkel S. 2015. Microproteins: small size—big impact. *Trends Plant Sci* **20:** 477–482. doi:10.1016/j.tplants.2015.05.011
Franceschetti M, Hanfrey C, Scaramagli S, Torrigiani P, Bagni N, Burtin D, Michael AJ. 2001. Characterization of monocot and dicot plant S-adenosyl-l-methionine decarboxylase gene families including identification in the mRNA of a highly conserved pair of upstream overlapping open reading frames. *Biochem J* **353:** 403–409. doi:10.1042/0264-6021:3530403
German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, et al. 2008. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotechnol* **26:** 941–946. doi:10.1038/nbt1417
Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, Ecker JR. 2008. A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev Cell* **14:** 854–866. doi:10.1016/j.devcel.2008.04.005
Guo R, Yu X, Gregory BD. 2023. The identification of conserved sequence features of co-translationally decayed mRNAs and upstream open reading frames in angiosperm transcriptomes. *Plant Direct* **7:** e479. doi:10.1002/pld3.479
Guydosh NR, Green R. 2014. Dom34 rescues ribosomes in 3′ untranslated regions. *Cell* **156:** 950–962. doi:10.1016/j.cell.2014.02.006
Harnett D, Meerdink E, Calviello L, Sydow D, Ohler U. 2021. Genome-wide analysis of actively translated open reading frames using RiboTaper/ORFquant. *Methods Mol Biol* **2252:** 331–346. doi:10.1007/978-1-0716-1150-0_16
Hayden CA, Jorgensen RA. 2007. Identification of novel conserved peptide uORF homology groups in *Arabidopsis* and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. *BMC Biol* **5:** 32. doi:10.1186/1741-7007-5-32
Hou CY, Lee WC, Chou HC, Chen AP, Chou SJ, Chen HM. 2016. Global analysis of truncated RNA ends reveals new insights into ribosome stalling in plants. *Plant Cell* **28:** 2398–2416. doi:10.1105/tpc.16.00295
Hsu PY, Benfey PN. 2018. Small but mighty: functional peptides encoded by small ORFs in plants. *Proteomics* **18:** 1700038. doi:10.1002/pmic.201700038
Hsu PY, Calviello L, Wu HYL, Li FW, Rothfels CJ, Ohler U, Benfey PN. 2016. Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proc Natl Acad Sci* **113:** E7126–E7135. doi:10.1073/pnas.1614788113
Hu H, Boisson-Dernier A, Israelsson-Nordström M, Böhmer M, Xue S, Ries A, Godoski J, Kuhn JM, Schroeder JI. 2010. Carbonic anhydrases are upstream regulators of CO2-controlled stomatal movements in guard cells. *Nat Cell Biol* **12:** 87–93; sup pp. 1–18. doi:10.1038/ncb2009
Hu H, Rappel W-J, Occhipinti R, Ries A, Böhmer M, You L, Xiao C, Engineer CB, Boron WF, Schroeder JI. 2015. Distinct cellular locations of carbonic anhydrases mediate carbon dioxide control of stomatal movements. *Plant Physiol* **169:** 1168–1178. doi:10.1104/pp.15.00646
Hu L, Liu S, Peng Y, Ge R, Su R, Senevirathne C, Harada BT, Dai Q, Wei J, Zhang L, et al. 2022. M6a RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat Biotechnol* **40:** 1210–1219. doi:10.1038/s41587-022-01243-z
Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223. doi:10.1126/science.1168978
Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:** 789–802. doi:10.1016/j.cell.2011.10.002
Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7:** 1534–1550. doi:10.1038/nprot.2012.086
Iwakawa H, Lam AYW, Mine A, Fujita T, Kiyokawa K, Yoshikawa M, Takeda A, Iwasaki S, Tomari Y. 2021. Ribosome stalling caused by the argonaute-microRNA-SGS3 complex regulates the production of secondary siRNAs in plants. *Cell Rep* **35:** 109300. doi:10.1016/j.celrep.2021.109300
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. doi:10.7554/eLife.08890
Kim Y-S, Kim S-G, Lee M, Lee I, Park H-Y, Seo PJ, Jung J-H, Kwon E-J, Suh SW, Paek K-H, et al. 2008. HD-ZIP III activity is modulated by competitive inhibitors via a feedback loop in *Arabidopsis* shoot apical meristem development. *Plant Cell* **20:** 920–933. doi:10.1105/tpc.107.057448
Kiniry SJ, O'Connor PBF, Michel AM, Baranov PV. 2019. Trips-Viz: a transcriptome browser for exploring Ribo-seq data. *Nucleic Acids Res* **47:** D847–D852. doi:10.1093/nar/gky842
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 Comput Biol* **9:** e1003118. doi:10.1371/journal.pcbi.1003118
Lee S, Liu B, Lee S, Huang SX, Shen B, Qian SB. 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci* **109:** E2424–E2432. doi:10.1073/pnas.1207846109
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. doi:10.7554/elife.22750
Li YR, Liu MJ. 2020. Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants. *Genome Res* **30:** 1418–1433. doi:10.1101/GR.261834.120
Michel AM, Kiniry SJ, O'Connor PBF, Mullan JP, Baranov PV. 2018. GWIPS-viz: 2018 update. *Nucleic Acids Res* **46:** D823–D830. doi:10.1093/nar/gkx790
Nagarajan VK, Kukulich PM, Von Hagel B, Green PJ. 2019. RNA degradomes reveal substrates and importance for dark and nitrogen stress responses of *Arabidopsis* XRN4. *Nucleic Acids Res* **47:** 9216–9230. doi:10.1093/nar/gkz712
Nielsen M, Ard R, Leng X, Ivanov M, Kindgren P, Pelechano V, Marquardt S. 2019. Transcription-driven chromatin repression of intragenic transcription start sites. *PLoS Genet* **15:** e1007969. doi:10.1371/journal.pgen.1007969
Pelechano V, Wei W, Steinmetz LM. 2015. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* **161:** 1400–1412. doi:10.1016/j.cell.2015.05.008
R Core Team. 2025. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265:** 687–695. doi:10.1038/265687a0
Sas-Chen A, Thomas JM, Matzov D, Taoka M, Nance KD, Nir R, Bryson KM, Shachar R, Liman GLS, Burkhart BW, et al. 2020. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature* **583:** 638–643. doi:10.1038/s41586-020-2418-2
Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, Sandelin A. 2020. Characterization of *Arabidopsis thaliana* promoter bidirectionality and antisense RNAs by inactivation of nuclear RNA decay pathways. *Plant Cell* **32:** 1845–1867. doi:10.1105/tpc.19.00815
Tjeldnes H, Labun K, Torres Cleuren Y, Chyżyńska K, Świrski M, Valen E. 2021. ORFik: a comprehensive R toolkit for the analysis of translation. *BMC Bioinformatics* **22:** 336. doi:10.1186/s12859-021-04254-w
Valen E, Pascarella G, Chalk A, Maeda N, Kojima M, Kawazu C, Murata M, Nishiyori H, Lazarevic D, Motti D, et al. 2009. Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res* **19:** 255–265. doi:10.1101/gr.084541.108
Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K. 2008. DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res* **36:** D97–D101. doi:10.1093/nar/gkm901
Weerasooriya HN, Longstreth DJ, DiMario RJ, Rosati VC, Cassel BA, Moroney JV. 2024. Carbonic anhydrases in the cell wall and plasma membrane of *Arabidopsis thaliana* are required for optimal plant growth on low CO_2. *Front Mol Biosci* **11:** 1267046. doi:10.3389/fmolb.2024.1267046
Wenkel S, Emery J, Hou B-H, Evans MMS, Barton MK. 2007. A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes. *Plant Cell* **19:** 3379–3390. doi:10.1105/tpc.107.055772
Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Springer, Cham, Switzerland. https://link.springer.com/book/10.1007/978-3-319-24277-4
Willems P, Ndah E, Jonckheere V, Stael S, Sticker A, Martens L, Van Breusegem F, Gevaert K, Van Damme P. 2017. N-terminal proteomics assisted profiling of the unexplored translation initiation landscape in *Arabidopsis thaliana*. *Mol Cell Proteomics* **16:** 1064–1080. doi:10.1074/mcp.M116.066662
Wright BW, Yi Z, Weissman JS, Chen J. 2022. The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol* **32:** 243–258. doi:10.1016/j.tcb.2021.10.010
Wu HYL, Hsu PY. 2021. *RiboPlotR*: a visualization tool for periodic Ribo-seq reads. *Plant Methods* **17:** 124. doi:10.1186/s13007-021-00824-4
Wu H-YL, Hsu PY. 2024. Dispersed transcription start sites modulate uORF-mediated regulation in *Arabidopsis*. bioRxiv doi:10.1101/2024.04.25.591216
Wu HYL, Song G, Walley JW, Hsu PY. 2019. The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiol* **181:** 367–380. doi:10.1104/pp.19.00541

Wu H-YL, Ai Q, Teixeira RT, Nguyen PHT, Song G, Montes C, Elmore JM, Walley JW, Hsu PY. 2024a. Improved super-resolution ribosome profiling reveals prevalent translation of upstream ORFs and small ORFs in *Arabidopsis*. *Plant Cell* **36:** 510–539. doi:10.1093/plcell/koad290

Wu H-YL, Jen J, Hsu PY. 2024b. What, where, and how: regulation of translation and the translational landscape in plants. *Plant Cell* **36:** 1540–1564. doi:10.1093/plcell/koad197

Xiao Z, Huang R, Xing X, Chen Y, Deng H, Yang X. 2018. De novo annotation and characterization of the translatome with ribosome profiling data. *Nucleic Acids Res* **46:** e61. doi:10.1093/nar/gky179

Yang X, Song B, Cui J, Wang L, Wang S, Luo L, Gao L, Mo B, Yu Y, Liu L. 2021. Comparative ribosome profiling reveals distinct translational landscapes of salt-sensitive and -tolerant rice. *BMC Genomics* **22:** 612. doi:10.1186/s12864-021-07922-6

Yu X, Willmann MR, Anderson SJ, Gregory BD. 2016. Genome-wide mapping of uncapped and cleaved transcripts reveals a role for the nuclear mRNA cap-binding complex in cotranslational RNA decay in *Arabidopsis*. *Plant Cell* **28:** 2385–2397. doi:10.1105/tpc.16.00456

Zhang P, He D, Xu Y, Hou J, Pan B-F, Wang Y, Liu T, Davis CM, Ehli EA, Tan L, et al. 2017. Genome-wide identification and differential analysis of translational initiation. *Nat Commun* **8:** 1749. doi:10.1038/s41467-017-01981-8

Received February 10, 2025; accepted in revised form July 17, 2025.