ggRibo: a ggplot-based single-gene viewer for visualizing Ribo-seq and related omics datasets

Running Title: *ggRibo* visualization tool

Author Affiliations: Hsin-Yen Larry Wu[1], Isaiah D. Kaufman[1], and Polly Yingshan Hsu[1, a]

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

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

ORCID:

0000-0001-9407-338X (H.L.W.), 0000-0001-7091-1887 (I.D.K.), 0000-0001-7071-5798 (P.Y.H.)

ABSTRACT:

Visualizing periodic Ribo-seq data within genes of interest is a powerful approach to studying mRNA translation, 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 datasets with flexible scaling options. *ggRibo* presents the 3-nucleotide periodicity, a hallmark of translating ribosomes, within a gene-structure context, including introns and untranslated regions, enabling the study of novel ORFs, isoform translation, and mechanisms of translational regulation. *ggRibo* can plot multiple Ribo-seq/RNA-seq datasets from different conditions for comparison. Additionally, it supports the visualization of other omics datasets that could also be presented with single-nucleotide resolution, such as RNA degradome, transcription start sites, and translation initiation sites. Through its intuitive and flexible platform, *ggRibo* enables parallel comparisons of multi-omic datasets, facilitating a comprehensive understanding of gene expression regulation and promoting hypothesis generation. We demonstrate its utility with examples of upstream ORFs, downstream ORFs, isoform translation, and multi-omic comparison in humans and Arabidopsis. In summary, *ggRibo* is an advanced single-gene viewer that enhances the interpretation of translatome and related genome-wide datasets, offering a valuable resource for studying gene expression regulation. *ggRibo* is available on GitHub (https://github.com/hsinyenwu/ggRibo).

Keywords: ribosome profiling, visualization, periodicity, single nucleotide resolution, translation

INTRODUCTION:

Ribo-seq has emerged as a widely used technique for studying the landscape and molecular mechanisms of mRNA translation across diverse organisms (Brar and Weissman 2015; Wu et al. 2024b). By optimizing RNase digestion to obtain precise ribosome-protected mRNA fragments, i.e., ribosome footprints, Ribo-seq captures the periodic 3-nucleotide (3-nt, codon-by-codon) movement of ribosomes during translation—a feature known as 3-nt periodicity (Ingolia et al. 2009; Hsu et al. 2016). This hallmark is crucial for identifying translated open reading frames (ORFs) and investigating the mechanism of mRNA translation (Hsu et al. 2016; Chothani et al. 2023; Wu et al. 2024a). **Figure 1A-C** illustrates how ribosome footprints can be presented by single-nucleotide positions and color-coded to indicate the reading frame, enabling visualization of 3-nt periodicity along translated ORFs. Note the single-nucleotide positions assigned to different lengths of ribosome footprints could vary across different organisms and organelles (**Figure 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).

3-nt periodicity has been integrated into mainstream Ribo-seq software for genome-wide ORF discovery (Ji et al. 2015; Calviello et al. 2016; Zhang et al. 2017; Xiao et al. 2018; Calviello et al. 2020; Choudhary et al. 2020; Harnett et al. 2021). We further leveraged 3-nt periodicity within individual genes as a powerful tool for exploration and preliminary validation, uncovering novel translational features through visualizing the distribution of periodic Ribo-seq reads (Hsu et al. 2016; Hsu and Benfey 2018; Wu et al. 2019, 2024a). However, visualizing 3-nt periodicity within individual genes remains challenging, partly due to a lack of suitable visualization tools. Many studies simply present Ribo-seq data as coverage plots or single-color lines, 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 primary expressed/translated isoform, thereby missing the opportunity to examine other annotated or unannotated isoforms (**Figure S1A, B**). 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 visualization of 3-nt periodicity (Michel et al. 2018).

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 (**Figure S1C**). 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 unannotated translation features, it has

several limitations: 1) It only accepts specific annotation formats, where transcript IDs are assigned according to gene IDs—a common nomenclature in plant genome annotations. This restricts its applicability to certain genomes, such as the human genome. 2) It only supports the visualization of the annotated ORFs and upstream ORFs (uORFs) in the 5′ UTR but does not accommodate other unannotated ORFs, such as downstream ORFs in the 3′ UTR or novel small ORFs within annotated noncoding RNAs. 3) It only displays a maximum of two Ribo-seq samples per plot, preventing large-scale comparisons. 4) It relies on base R plotting capabilities, producing figures that are less refined than those generated with *ggplot2* (Wickham 2016).

To improve the flexibility and the quality of Ribo-seq data visualization, we developed *ggRibo*, which extends the advantages of *RiboPlotR* while leveraging the capabilities of *ggplot2*. *ggRibo* accepts broader annotation formats, visualizes all types of unannotated ORFs, and accommodates as many Ribo-seq datasets as the computer's RAM allows. *ggRibo* can also zoom in to investigate translation in parts of a gene and plot DNA and amino acid sequences alongside the Ribo-seq plots. These improvements allow *ggRibo* to fully harness the potential of high-quality Ribo-seq data visualization.

In addition, *ggRibo* can visualize other sequencing data types with single nucleotide resolution, where each sequencing read could be represented by one nucleotide position. Examples of such data include, but are not limited to: Cap Analysis of Gene Expression sequencing (CAGE-seq) (Valen et al. 2009), Transcription Start Site sequencing (TSS-seq) (Wakaguri et al. 2008), Parallel Analysis of RNA Ends sequencing (PARE-seq) (German et al. 2008), Genome-wide Mapping of Uncapped and Cleaved Transcripts) (GMUCT) (Gregory et al. 2008), 5′ Phosphate sequencing (5P-seq) (Pelechano et al. 2015), Translation Initiation sequencing (TI-seq) (Ingolia et al. 2011; Lee et al. 2012; Willems et al. 2017; Li and Liu 2020), Ribo-seq (Ingolia et al. 2009), and sequencing techniques that detect the precise positions of epitranscriptomic modifications such as m6A-SAC-seq (Hu et al. 2022) and ac4C-seq (Sas-Chen et al. 2020). In CAGE-seq and TSS-seq data, the first nucleotide of a sequencing read corresponds to the transcription start site. In PARE-seq, GMUCT, and 5P-seq data, the first nucleotide indicates the 5′ end of a decayed mRNA. In TI-seq and Ribo-seq, the position of the ribosome footprint is often presented by the first nucleotide of the peptidyl site (P-site) within the ribosome—often the 13th nucleotide in a 28-nt ribosome footprint across many species (Ingolia et al. 2011; Hsu et al. 2016; Wu et al. 2019). TI-seq reads enrich for ribosomes engaged in translation initiation, while Ribo-seq reads capture ribosomes involved in both initiation and elongation (Schneider-Poetsch et al. 2010; Ingolia et al. 2011; Lee et al. 2012; Li and Liu 2020). m6A-SAC-seq and ac4C-seq identify the nucleotide positions of their corresponding

epitranscriptomic modifications, specifically N^6-methyladenosine and N^4-acetylcytidine, respectively. Integrating and comparing Ribo-seq data with other single-nucleotide resolution sequencing data can deepen our understanding of gene regulation at multiple levels and generate new hypotheses for further research.

RESULTS:

Design of *ggRibo*

To facilitate the visual interpretation of translation events from periodic Ribo-seq reads, we developed *ggRibo*, a single-gene viewer that displays Ribo-seq periodicity alongside RNA-seq coverage within the gene structure context, including 5′ and 3′ UTRs and introns (**Figure 2**). *ggRibo* supports all GTF and GFF annotation files compatible with the *GenomicRanges* package (Lawrence et al. 2013). Built on the *ggplot2* framework, *ggRibo* generates publication-quality plots. Notably, *ggRibo* visualizes both annotated ORFs and extra ORFs (called “eORFs” in the *ggRibo* package) (**Figure 3A**). It can also extend the main ORF reading frame into the 5′ UTR and 3′ UTR, allowing the visualization of alternative start sites (either AUG or non-AUG) and stop codon readthrough, respectively (**Figure 3A**). Additionally, *ggRibo* can plot Ribo-seq reads on annotated noncoding RNAs to evaluate whether a potential small ORF is translated with robust 3-nt periodicity (**Figure 3B**). A summary of *ggRibo*’s improvements over its prototype, *RiboPlotR*, is provided in **Figure S2**.

For each Ribo-seq read, *ggRibo* plots the position of the first nucleotide of the P-site (**Figure 1**). The P-site positions can be determined using ORF discovery tools such as *RiboTaper* or *Ribo-seQC* (Calviello et al. 2016, 2019). With these P-site information files, *ggRibo* highlights unannotated ORFs and noncanonical translation features, thereby enhancing the detection of complex translational activities within genes of interest. By default, *ggRibo* displays genes from left to right, regardless of whether they are on the Watson or Crick strand. While this behavior differs from conventional genome browsers, it simplifies visualization and improves user interpretation.

In addition to the primary *ggRibo* function, we provide *ggRibo_decom* and *ggRNA* functions. *ggRibo_decom* is designed for reading frame decomposition, plotting Ribo-seq reads assigned to each reading frame separately (**Figure S3A**). It processes one dataset at a time, allowing a more detailed view of frame-specific distributions. *ggRNA*, in contrast, plots RNA-seq data alone when only RNA-seq coverage is of interest (**Figure S3B**). Similar to *ggRibo* main function, *ggRNA* supports multiple samples and both paired-end and single-end RNA-seq data. Below, we present examples illustrating how *ggRibo* can be applied to explore unannotated

ORFs and noncanonical translation features within the translatome and to integrate with other omics datasets to advance our comprehensive understanding of gene regulation. The setup, requirements, and functions of the *ggRibo* package are detailed in the 'Methods and Utility' section.

Visualizing Translated ORFs with Human Data Using *ggRibo*

ggRibo can illustrate both simple and complex translation events. To demonstrate its basic usage, we first provide an example using the human *MRPL11* gene (**Figure 4**), which contains a uORF in the 5′ UTR that represses downstream main ORF translation (Calvo et al. 2009). **Figure 4A** presents the Ribo-seq and RNA-seq data of *MRPL11* using human embryonic stem cell data (Chothani et al. 2022), displaying all annotated RNA isoforms, and the uORF is highlighted in yellow in the gene models. **Figure 4B** adds a DNA sequence track for additional context, and **Figure 4C** zooms in on the *MRPL11* uORF, displaying both DNA nucleotide and amino acid sequence tracks.

Next, we show a more complex example from the human immune-related gene *IFITM3* (*Interferon-Induced Transmembrane Protein 3*), which contains isoforms both with and without an annotated CDS, using the human brain data (Chothani et al. 2022). The *ggRibo* plot reveals that one *IFITM3* isoform, ENST00000399808, is the major expressed and translated isoform in the human brain (**Figure 5A**: a gene view displaying all RNA isoforms, **5B**: zoom in to view a subset of isoforms, offering a closer inspection of RNA-seq and Ribo-seq profiles). This example demonstrates how visualizing RNA-seq and Ribo-seq reads in the gene structure context helps identify expressed and translated isoforms.

Visualizing Translated ORFs with Arabidopsis Data Using *ggRibo*

We next present 4 examples using a high-coverage, super-resolution Ribo-seq dataset from the model plant *Arabidopsis thaliana* (Wu et al. 2024a) . Because ~90% of Ribo-seq reads from this dataset map to the annotated reading frame, its optimal depth and resolution have enabled the discovery of numerous unannotated translation events (Wu et al. 2024a).

PAE7 encodes a pectin acetylesterase and has two RNA isoforms. RNA-seq coverage indicates that both isoforms are expressed, and the Ribo-seq periodicity supports that both isoforms are translated (**Figure 6A**). Additionally, cross-referencing the RNA-seq profile with gene models reveals that the expressed exon 14 of isoform 2 is longer than the annotated range (**Figure 6A**, blue dashed box). This finding further demonstrates how gene-structure-based visualization aids in interpreting isoform translation and refining gene annotations.

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, we plotted the entire *SnRK2.2* gene (**Figure 6B**) and a zoomed-in view with the DNA nucleotide and amino acid sequence tracks (**Figure 6C**). 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). This read distribution pattern reflects the altered ribosome footprint length at the codon preceding the stop codon and is uniquely associated with minimal uORFs (Wu et al. 2024a).

TAS3 is an annotated non-coding RNA that encodes the primary transcript of trans-acting short-interfering RNA (*tasiRNA*). The *TAS3* primary transcript is known to translate a small ORF, 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). Using *ggRibo*, we can clearly observe that the region corresponding to this small ORF is highly enriched with Ribo-seq reads with strong 3-nt periodicity (**Figures 6D** and **S4**: zoom in to view the sequences near the start and stop codons). Further analysis of *TAS3* expression profiles with *ggRibo_decom*, which plots 3-frame reads separately, confirmed that the Ribo-seq reads are predominantly enriched in frame 0 (**Figure S5**). The *ggRibo_decom* function is particularly useful for highlighting such frame enrichment in non-coding RNAs.

Notably, for non-coding RNAs, *ggRibo* assigns the reading frame from the first nucleotide of the annotated RNA sequence, rather than from the start of a CDS/ORF, as it does for coding RNAs. As a result, a translated ORF in a non-coding RNA may use one of the reading frames in red, blue, or green. **Figure S6** shows an example of a small ORF enriched in the 'blue' reading frame from an annotated non-coding RNA.

Comparing Differential Translated Isoforms Across Multiple Datasets with *ggRibo*

ggRibo can visualize multiple Ribo-seq/RNA-seq datasets to compare translated isoforms and/or translation efficiency of genes of interest. Here, we show an example of tissue-specific isoform translation of *BCA4* (*β-Carbonic Anhydrase 4*) comparing Arabidopsis 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 3 annotated isoforms; from the *ggRibo* plot, we observe isoforms 2 and 3 expressed in the tissues examined (**Figure 7**: comparing the gene models at the bottom with the expression profiles). Interestingly, while isoform 2 is highly expressed and translated in the shoot and whole seedling, it is nearly undetectable in the root

(**Figure 7**). Instead, the root preferentially expresses isoform 3. Recent work has shown that the isoform 2 (the longer isoform) protein localizes to the plasma membrane while the isoform 3 (the shorter isoform) protein is localized in the cytosol (Weerasooriya et al. 2024). This example highlights *ggRibo*'s capacity to compare multiple datasets, enabling investigations into differential translation or isoform-specific translation across tissues and conditions and providing valuable insights for future functional studies.

Multi-omic Data Comparisons Using *ggRibo*

Besides visualizing Ribo-seq/RNA-seq profiles, *ggRibo* can integrate other omics datasets presented with single nucleotide resolution. To illustrate its utility in multi-omic data comparison, we provide 3 examples below. Note that other omics data can be displayed either with a single color (examples in **Figures 8** and **9**) or with 3 colors to view 3-nt periodicity (example in **Figure 10**) by setting the 'sample_color' parameter in the *ggRibo* function.

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

Little Zippers (ZPR1 to 4) 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 multi-domain homologs and act as dominant negative regulators (Eguen et al. 2015; Bhati et al. 2021) Originally, ZPR1 was predicted to be a 93 a.a. peptide (Wenkel et al. 2007). However, it is annotated as 136 a.a. in both 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* begin downstream of the annotated AUG start (**Figure 8: Ribo-seq panel**). In fact, the Ribo-seq reads mostly start at the next in-frame AUG, i.e., from the ORF in the AT2G45450.x isoform we defined (**Figure 8:** black arrowheads in Ribo-seq panel and the start/stop codon track), supporting that *ZPR1* is predominantly translated into a 93 a.a. peptide.

Why does *ZPR1* primarily use the second AUG? Our recent study revealed that transcription start sites in Arabidopsis can be highly variable, affecting ORF translation initiation sites (Wu and Hsu 2024). More specifically, some transcripts may begin transcription downstream of 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 transcription start sites (TSSs) (Nielsen et al. 2019), and compared them with the Ribo-seq data and the gene model (**Figure 8**). Consistent with our hypothesis, the TSS-seq signals for *ZPR1* occur primarily downstream of the annotated start codon, explaining why *ZPR1* translation preferentially initiates at the second AUG (**Figure 8: TSS-seq panel and gene model**).

Second Example: a downstream ORF (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, identified using an ORF discovery tool, *RiboTaper* (Calviello et al. 2016; Wu et al. 2024a). *RiboTaper* classified this dORF as an overlapping dORF in *AT1G49980.1* isoform, overlapping with the stop codon of the main ORF (see gene models in **Figure 8A-B**). Although plotting the annotated gene range for *AT1G49980* confirmed a highly translated ORF within the annotated 3′ UTR (**Figure 8A, Ribo-seq panel**: green reading frame), the higher RNA coverage in the dORF region compared to the main ORF region suggests that this dORF may arise from a separate unannotated gene/transcript. To test this possibility, we examined transcription start sites reported in TSS-seq data (Nielsen et al. 2019). We found two distinct TSS clusters: one near the annotated transcription start site and the other upstream of the dORF (**Figure 8A-B, TSS-seq panel**: purple lines). This pattern suggests the presence of an unannotated gene/transcript downstream of *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 2 potential ATG start codons: one overlapping with the main ORF's stop codon and the other located 2 codons downstream (**Figure 8B**, 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 specialized in enriching initiating ribosomes to identify translation initiation sites. The TI-seq data show 2 major peaks: the first one near the annotated start codon of *AT1G49980* and the second one at the dORF region (arrows in **Figure 8A, TI-seq panel**). Zooming in on the 5′ region of the dORF revealed that this dORF exclusively uses the second ATG (**Figure 8B, TI-seq panel**), indicating that this dORF does not overlap with the *AT1G49980* main ORF.

Third Example: co-translation mRNA decay observed in a conserved peptide uORF

Co-translational 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, where 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 (**Figure 10**). As noted in previous studies (Hou et al. 2016; Yu et al. 2016; Wu et al. 2024a), Ribo-seq and PARE-seq suggested

stalled ribosomes at approximately 30-nucleotide intervals (the estimated length of ribosome footprint) from the end of the uORF (**Figure 10A, B**). Meanwhile, PARE-seq data display corresponding peaks 13 nucleotides upstream of the Ribo-seq peaks (**Figure 10B**). This pattern aligns with the expectations for co-translation decay, where ribosomes act as a protective barrier against exonucleolytic degradation, causing mRNA decay to stop at the 5′ boundary of the ribosome (Hou et al. 2016; Yu et al. 2016). Remarkably, the shift between red and blue reading frames in Ribo-seq data prior to the stop codon (Wu et al. 2024a) is also reflected in PARE-seq data (blue/green) (**Figure 10B**).

In summary, these examples demonstrate the versatility of *ggRibo* in integrating Ribo-seq with other genome-wide sequencing data, allowing researchers to gain deeper insights into the complexity of gene expression regulation.

DISCUSSION:

Visualization is emphasized in cell biology research to generate hypotheses and results. However, in omics studies, the visualization of sequencing data remains underutilized due 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 (e.g., across different organs, under stress, or with chemical treatments). By integrating Ribo-seq data with other sequencing datasets, *ggRibo* facilitates multi-omics comparisons, providing deeper insights into translational regulation and different levels of gene regulation.

One example of using visualization tools to study translational regulation comes from an earlier application of *RiboPlotR* (the prototype of *ggRibo*) (Wu and Hsu 2021) to investigate how uORF repressiveness is determined. We observed that weakly translated uORFs are often located near the 5′ end of RNA-seq coverage. Cross-referencing transcription start sites from CAGE-seq and TSS-seq datasets with Ribo-seq and RNA-seq data revealed that only a small fraction of those uORF mRNAs actually include the uORFs (Wenkel et al. 2007). In other words, weak uORFs often map upstream of the main transcription start sites, and, therefore, they are not present in the majority of the transcripts. While transcription start sites can be dispersed among many organisms (Valen et al. 2009; Morton et al. 2014; Cheng et al. 2018), the impact of this dispersion on uORF repressiveness remained largely unexplored until our visualization-based study (Wu and Hsu 2024).

The above example highlights the importance of visualizing sequencing data within genes of interest to better understand biological regulation and develop new hypotheses for

studying its molecular mechanisms. With its flexibility and simple workflow, *ggRibo* is a powerful tool for researchers to examine mRNA translation at an unparalleled resolution. Finally, *ggRibo* can also visualize other single-nucleotide resolution sequencing data, independent of Ribo-seq. Thus, the broad applicability of *ggRibo* will help researchers efficiently visualize individual gene expression from large-scale datasets, produce publication-quality figures, and validate or generate hypotheses to advance the study of gene expression regulation.

METHODS AND UTILITY:

ggRibo package includes multiple functions for processing and visualizing Ribo-seq and RNA-seq data, with a focus on ORF analysis and mRNA translation within the gene structure context. All data files are imported and saved as R6 objects. The steps and functions for *ggRibo* visualization are outlined below:

First, annotation files are imported using gtf_import() and eORF_import() functions (eORF refers to extra ORFs, i.e., unannotated ORFs).

- gtf_import() function imports GTF/GFF annotation files, creating a Range_info object that extracts genomic features such as exons, transcripts, coding sequences (CDS), and untranslated regions (UTRs). This object serves as a foundational dataset for subsequent analyses.
- eORF_import() function imports user-defined extra ORF (eORF) annotations, creating an eORF_Range_info object to accommodate non-standard ORFs that may not be included in standard annotations.

Secondly, the Ribo_data() function reads Ribo-seq data files (or other single-nucleotide resolution data) and organizes them into a list of data frames. Each data frame contains read counts, chromosome numbers, chromosome positions, and strand information. This structured format enables efficient manipulation and analysis of Ribo-seq data.

The code provides several frame assignment functions that are essential for assigning translated reading frames:

- assign_frames() assigns reading frames to Ribo-seq reads based on provided CDS ranges, allowing for precise determination of which codons are being translated.
- assign_frames_extended() extends annotated frame assignments into UTRs, accommodating overlapping eORFs and ensuring accurate frame calculations even when non-standard ORFs are present.

- assign_frames_with_extension() enables custom extensions into the 5' and 3' UTRs (i.e., takes fExtend or tExtend), offering flexibility in analyzing alternative AUG or non-AUG starts in the 5' UTR and stop codon readthrough in the 3' UTR.
- exclude_eORF_reads() filters out Ribo-seq reads that overlap with eORF regions from the main dataset, preventing redundant plotting.

The main plotting function, ggRibo(), serves as the core tool for Ribo-seq and RNA-seq data visualization. It plots Ribo-seq periodicity alongside RNA-seq coverage within the gene structure context, integrating transcript isoform and genome annotations to highlight unannotated ORFs and unconventional translation events. This function supports customization of plot aesthetics, scaling, and the inclusion of additional single-nucleotide resolution sequencing methods, such as TI-seq, degradome sequencing, and CAGE-seq. By accommodating various data types and providing detailed visualizations, *ggRibo* facilitates the identification of factors influencing translation and aids in generating hypotheses about gene expression and translational regulation.

For visualizing DNA sequences and gene models, the code includes the functions plotDNAandAA() and plotGeneTxModel(), respectively. When using the plot_range parameter in ggRibo() for a genomic region ≤ 200 nucleotides, plotDNAandAA() generates plots displaying both DNA nucleotides and their corresponding amino acids, enabling visualization of codon usage and reading frames. For genomic regions > 200 nucleotides, plotDNAandAA() highlights start codons (green) and stop codons (red) in each frame. The plotGeneTxModel() function creates gene models illustrating exons, UTRs, CDS, and eORFs for specified genes and their isoforms. It accounts for overlapping features and adjusts visual elements such as colors and labels for clarity, providing a comprehensive view of gene structure and potential translation events.

In conclusion, the *ggRibo* package leverages the *ggplot2* framework for high-quality plotting and follows object-oriented principles using custom classes. It efficiently handles complex genomic data structures and ensures accurate frame assignments, which are critical for Ribo-seq and related genomic data analysis.

The coding and part of the methods section were assisted by the ChatGPT model o1. The initial code was a heavily modified RiboPlotR that still used base R plots but was gradually transformed into ggplot2 by ChatGPT and Larry Wu. ChatGPT also facilitated the addition of new functionalities, such as the DNA/AA plot and Ribo-seq plot for ncRNAs with reading frame coloring. All tasks done by ChatGPT have detailed instructions from Larry Wu.

DATA ACCESS

ggRibo is available on GitHub (https://github.com/hsinyenwu/ggRibo). The example data are stored on Mendeley Data: Arabidopsis (https://data.mendeley.com/datasets/wm6cs4zbtw/1) and human (https://data.mendeley.com/datasets/m3t293k4wr/1)

ACKNOWLEDGEMENTS

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 to IDK, and research grants from the National Science Foundation under Award Numbers 2051885 and 2425390 and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM155375 to PYH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or NIH.

AUTHOR CONTRIBUTIONS

HLW wrote the *ggRibo* package and tested it with Arabidopsis data. IDK tested *ggRibo* with human data and fixed few bugs. HLW and IDK prepared the GitHub tutorial. HLW and PYH wrote the paper with input from IDK.

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

REFERENCES:

Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, and Bailey-Serres J. Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. Proceedings of the National Academy of Sciences of the United States of America. 2017:**114**(46):E10018–E10027. https://doi.org/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. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO Journal. 2014:**33**(9):981–993. https://doi.org/10.1002/embj.201488411

Bhati KK, Dolde U, and Wenkel S. MicroProteins: Expanding functions and novel modes of regulation. Mol Plant. 2021:**14**(5):705–707. https://doi.org/10.1016/j.molp.2021.01.006

Brar GA and Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. Nature Reviews Molecular Cell Biology. 2015:**16**(11):651–664. https://doi.org/10.1038/nrm4069

Calviello L, Hirsekorn A, and Ohler U. Quantification of translation uncovers the functions of the alternative transcriptome. Nature Structural and Molecular Biology. 2020:**27**(8):717–725. https://doi.org/10.1038/s41594-020-0450-4

Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, and Ohler U. Detecting actively translated open reading frames in ribosome profiling data. Nature Methods. 2016:**13**(2):165–170. https://doi.org/10.1038/nmeth.3688

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

Calvo SE, Pagliarini DJ, and Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proceedings of the National Academy of Sciences of the United States of America. 2009:**106**(18):7507–7512. https://doi.org/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. Pervasive functional translation of noncanonical human open reading frames. Science. 2020:**367**(6482):1140–1146. https://doi.org/10.1126/science.aay0262

Cheng Z, Otto GM, Powers EN, Keskin A, Mertins P, Carr SA, Jovanovic M, and Brar GA. Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. Cell. 2018:**172**(5):910-923.e16. https://doi.org/10.1016/j.cell.2018.01.035

Chothani S, Ho L, Schafer S, and Rackham O. Discovering microproteins: making the most of ribosome profiling data. RNA Biol. 2023:**20**(1):943–954. https://doi.org/10.1080/15476286.2023.2279845

Chothani SP, Adami E, Widjaja AA, Langley SR, Viswanathan S, Pua CJ, Zhihao NT, Harmston N, D'Agostino G, Whiffin N, et al. A high-resolution map of human RNA translation. Mol Cell. 2022:**82**(15):2885-2899.e8. https://doi.org/10.1016/j.molcel.2022.06.023

Choudhary S, Li W, and D Smith A. Accurate detection of short and long active ORFs using Ribo-seq data. Bioinformatics. 2020:**36**(7):2053–2059. https://doi.org/10.1093/bioinformatics/btz878

Eguen T, Straub D, Graeff M, and Wenkel S. MicroProteins: small size-big impact. Trends Plant Sci. 2015:**20**(8):477–482. https://doi.org/10.1016/j.tplants.2015.05.011

German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, et al. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol. 2008:**26**(8):941–946. https://doi.org/10.1038/nbt1417

Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, and Ecker JR. A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. Developmental Cell. 2008:**14**(6):854–866. https://doi.org/10.1016/j.devcel.2008.04.005

Guo R, Yu X, and Gregory BD. The identification of conserved sequence features of co-translationally decayed mRNAs and upstream open reading frames in angiosperm transcriptomes. Plant Direct. 2023:**7**(1):e479. https://doi.org/10.1002/pld3.479

Harnett D, Meerdink E, Calviello L, Sydow D, and Ohler U. Genome-Wide Analysis of Actively Translated Open Reading Frames Using RiboTaper/ORFquant. Methods in Molecular Biology. 2021:**2252**:331–346. https://doi.org/10.1007/978-1-0716-1150-0_16

Hou CY, Lee WC, Chou HC, Chen AP, Chou SJ, and Chen HM. Global analysis of truncated RNA ends reveals new insights into Ribosome Stalling in plants. Plant Cell. 2016:**28**(10):2398–2416. https://doi.org/10.1105/tpc.16.00295

Hsu PY and Benfey PN. Small but Mighty: Functional Peptides Encoded by Small ORFs in Plants. Proteomics. 2018:**18**(10):1700038. https://doi.org/10.1002/pmic.201700038

Hsu PY, Calviello L, Wu HYL, Li FW, Rothfels CJ, Ohler U, and Benfey PN. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America. 2016:**113**(45):E7126–E7135. https://doi.org/10.1073/pnas.1614788113

Hu H, Boisson-Dernier A, Israelsson-Nordström M, Böhmer M, Xue S, Ries A, Godoski J, Kuhn JM, and Schroeder JI. Carbonic anhydrases are upstream regulators of CO2-controlled stomatal movements in guard cells. Nat Cell Biol. 2010:**12**(1):87–93; sup pp 1-18. https://doi.org/10.1038/ncb2009

Hu H, Rappel W-J, Occhipinti R, Ries A, Böhmer M, You L, Xiao C, Engineer CB, Boron WF, and Schroeder JI. Distinct Cellular Locations of Carbonic Anhydrases Mediate Carbon Dioxide Control of Stomatal Movements. Plant Physiol. 2015:**169**(2):1168–1178. https://doi.org/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. m6A RNA modifications are measured at single-base resolution across the mammalian transcriptome. Nat Biotechnol. 2022:**40**(8):1210–1219. https://doi.org/10.1038/s41587-022-01243-z

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, and Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature Protocols. 2012:**7**(8):1534–1550. https://doi.org/10.1038/nprot.2012.086

Ingolia NT, Ghaemmaghami S, Newman JRS, and Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009:**324**(5924):218–223. https://doi.org/10.1126/science.1168978

Ingolia NT, Lareau LF, and Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011:**147**(4):789–802. https://doi.org/10.1016/j.cell.2011.10.002

Iwakawa H oki, Lam AYW, Mine A, Fujita T, Kiyokawa K, Yoshikawa M, Takeda A, Iwasaki S, and Tomari Y. Ribosome stalling caused by the Argonaute-microRNA-SGS3 complex regulates the production of secondary siRNAs in plants. Cell Reports. 2021:**35**(13):109300. https://doi.org/10.1016/j.celrep.2021.109300

Ji Z, Song R, Regev A, and Struhl K. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. eLife. 2015:**4**(DECEMBER2015):e08890. https://doi.org/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. HD-ZIP III activity is modulated by competitive inhibitors via a feedback loop in Arabidopsis shoot apical meristem development. Plant Cell. 2008:**20**(4):920–933. https://doi.org/10.1105/tpc.107.057448

Kiniry SJ, O'Connor PBF, Michel AM, and Baranov PV. Trips-Viz: A transcriptome browser for exploring Ribo-Seq data. Nucleic Acids Research. 2019:**47**(D1):D847–D852. https://doi.org/10.1093/nar/gky842

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, and Carey VJ. Software for Computing and Annotating Genomic Ranges. PLoS Computational Biology. 2013:**9**(8):e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Lee S, Liu B, Lee S, Huang SX, Shen B, and Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proceedings of the National Academy of Sciences of the United States of America. 2012:**109**(37):E2424-32. https://doi.org/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. Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. eLife. 2016:**5**:e22750. https://doi.org/10.7554/elife.22750

Li YR and Liu MJ. Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants. Genome Research. 2020:**30**(10):1418–1433. https://doi.org/10.1101/GR.261834.120

Michel AM, Kiniry SJ, O'Connor PBF, Mullan JP, and Baranov PV. GWIPS-viz: 2018 update. Nucleic Acids Research. 2018:**46**(D1):D823–D830. https://doi.org/10.1093/nar/gkx790

Morton T, Petricka J, Corcoran DL, Li S, Winter CM, Carda A, Benfey PN, Ohler U, and Megraw M. Paired-end analysis of transcription start sites in Arabidopsis reveals plant-specific promoter signatures. Plant Cell. 2014:**26**(7):2746–2760. https://doi.org/10.1105/tpc.114.125617

Nagarajan VK, Kukulich PM, Von Hagel B, and Green PJ. RNA degradomes reveal substrates and importance for dark and nitrogen stress responses of Arabidopsis XRN4. Nucleic Acids Research. 2019:**47**(17):9216–9230. https://doi.org/10.1093/nar/gkz712

Nielsen M, Ard R, Leng X, Ivanov M, Kindgren P, Pelechano V, and Marquardt S. Transcription-driven chromatin repression of Intragenic transcription start sites. PLoS Genetics. 2019:**15**(2). https://doi.org/10.1371/journal.pgen.1007969

Pelechano V, Wei W, and Steinmetz LM. Widespread co-translational RNA decay reveals ribosome dynamics. Cell. 2015:**161**(6):1400–1412. https://doi.org/10.1016/j.cell.2015.05.008

Sas-Chen A, Thomas JM, Matzov D, Taoka M, Nance KD, Nir R, Bryson KM, Shachar R, Liman GLS, Burkhart BW, et al. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature. 2020:**583**(7817):638–643. https://doi.org/10.1038/s41586-020-2418-2

Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, Merrick WC, Green R, Shen B, and Liu JO. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nature Chemical Biology. 2010:**6**(3):209–217. https://doi.org/10.1038/nchembio.304

Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, and Sandelin A. Characterization of arabidopsis thaliana promoter bidirectionality and antisense RNAs by inactivation of nuclear RNA decay pathways. Plant Cell. 2020:**32**(6):1845–1867. https://doi.org/10.1105/tpc.19.00815

Valen E, Pascarella G, Chalk A, Maeda N, Kojima M, Kawazu C, Murata M, Nishiyori H, Lazarevic D, Motti D, et al. Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. Genome Res. 2009:**19**(2):255–265. https://doi.org/10.1101/gr.084541.108

Wakaguri H, Yamashita R, Suzuki Y, Sugano S, and Nakai K. DBTSS: database of transcription start sites, progress report 2008. Nucleic Acids Res. 2008:**36**(Database issue):D97-101. https://doi.org/10.1093/nar/gkm901

Weerasooriya HN, Longstreth DJ, DiMario RJ, Rosati VC, Cassel BA, and Moroney JV. Carbonic anhydrases in the cell wall and plasma membrane of Arabidopsis thaliana are required for optimal plant growth on low CO2. Front Mol Biosci. 2024:**11**:1267046. https://doi.org/10.3389/fmolb.2024.1267046

Wenkel S, Emery J, Hou B-H, Evans MMS, and Barton MK. A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes. Plant Cell. 2007:**19**(11):3379–3390. https://doi.org/10.1105/tpc.107.055772

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer. 2016.

Willems P, Ndah E, Jonckheere V, Stael S, Sticker A, Martens L, Van Breusegem F, Gevaert K, and Van Damme P. N-terminal proteomics assisted profiling of the unexplored translation initiation landscape in Arabidopsis thaliana. Molecular and Cellular Proteomics. 2017:**16**(6):1064–1080. https://doi.org/10.1074/mcp.M116.066662

Wu H-YL, Ai Q, Teixeira RT, Nguyen PHT, Song G, Montes C, Elmore JM, Walley JW, and Hsu PY. Improved super-resolution ribosome profiling reveals prevalent translation of upstream ORFs and small ORFs in Arabidopsis. Plant Cell. 2024a:**36**(3):510–539. https://doi.org/10.1093/plcell/koad290

Wu HYL and Hsu PY. RiboPlotR: a visualization tool for periodic Ribo-seq reads. Plant Methods. 2021:**17**(1):1–7. https://doi.org/10.1186/s13007-021-00824-4

Wu H-YL and Hsu PY. Dispersed Transcription Start Sites Modulate uORF-Mediated Regulation in Arabidopsis. 2024:2024.04.25.591216. https://doi.org/10.1101/2024.04.25.591216

Wu H-YL, Jen J, and Hsu PY. What, where, and how: Regulation of translation and the translational landscape in plants. The Plant Cell. 2024b:**36**(5):koad197. https://doi.org/10.1093/plcell/koad197

Wu HYL, Song G, Walley JW, and Hsu PY. The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. Plant Physiology. 2019:**181**(1):367–380. https://doi.org/10.1104/pp.19.00541

Xiao Z, Huang R, Xing X, Chen Y, Deng H, and Yang X. De novo annotation and characterization of the translatome with ribosome profiling data. Nucleic Acids Res. 2018:**46**(10):e61. https://doi.org/10.1093/nar/gky179

Yu X, Willmann MR, Anderson SJ, and Gregory BD. Genome-wide mapping of uncapped and cleaved transcripts reveals a role for the nuclear mrna cap-binding

complex in cotranslational RNA decay in Arabidopsis. Plant Cell. 2016:**28**(10):2385–2397. https://doi.org/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. Genome-wide identification and differential analysis of translational initiation. Nat Commun. 2017:**8**(1):1749. https://doi.org/10.1038/s41467-017-01981-8

Figures

Figure 1

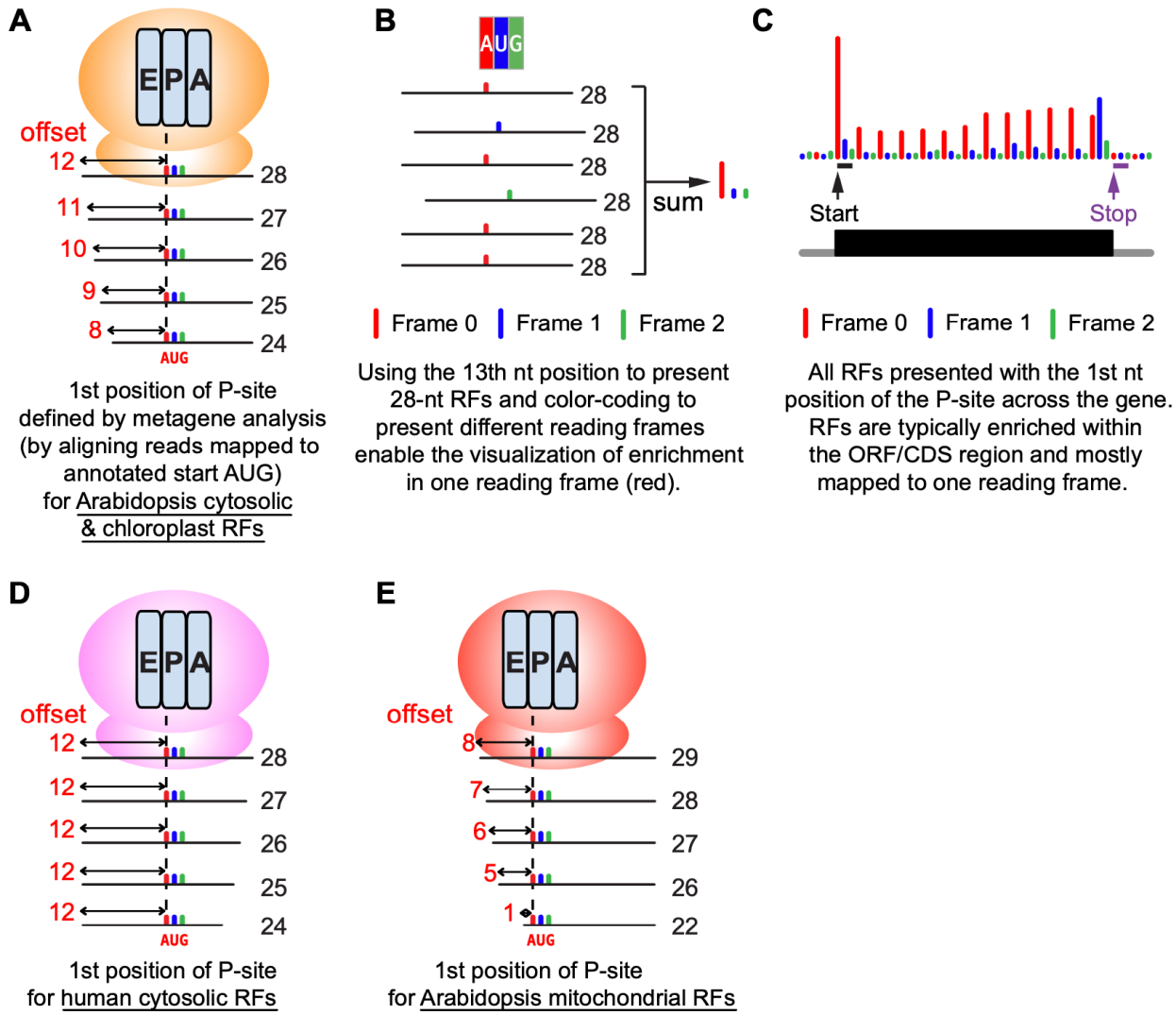

Figure 1. An illustration of how ribosome footprints (RFs) can be represented by the first nucleotide (nt) of the P-site within ribosomes and color-coded to indicate reading frame enrichment and highlight 3-nt periodicity

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 nt (offset by 12 nt from 5′) is typically the 1st nt position of the P-site for a 28-nt RF. Similarly, the 12th nt (offset by 11 nt from 5′) is the 1st nt position of the P-site for a 27-nt RF. Note the P-site offsets could vary in different organisms and organelles (see panels D-E) (Ingolia et al. 2011; Hsu et al. 2016; Wu et al. 2024).

B) An illustration showing how 28-nt RFs are mapped to a codon. The 28-nt RFs are presented by their 13th nt 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 datasets.

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 translation termination causes RFs to have a different length and results in an enrichment of frame 1 (blue) (Wu et al. 2024).
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 to the 5′ of the RFs in Arabidopsis (see panel 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 panel A).

Figure 2

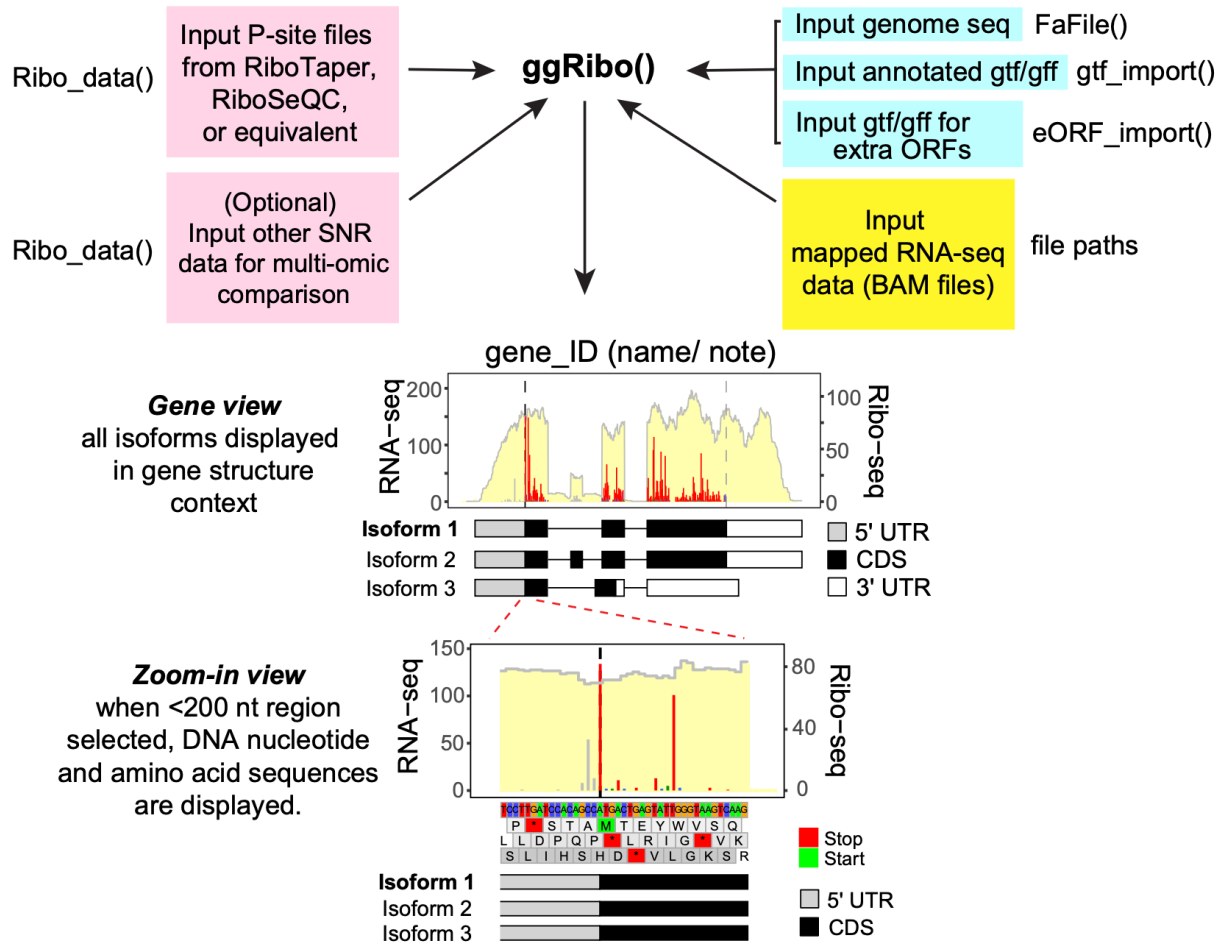

Figure 2. Input files required for running *ggRibo* and typical output plots in gene view and zoom-in view. *ggRibo* requires Ribo-seq P-site information, genome annotation for annotated ORFs and extra ORFs, and mapped RNA-seq BAM file(s), Other omics data presented with single-nucleotide position can be input and visualized in parallel. 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 zoom-in view.

Figure 3

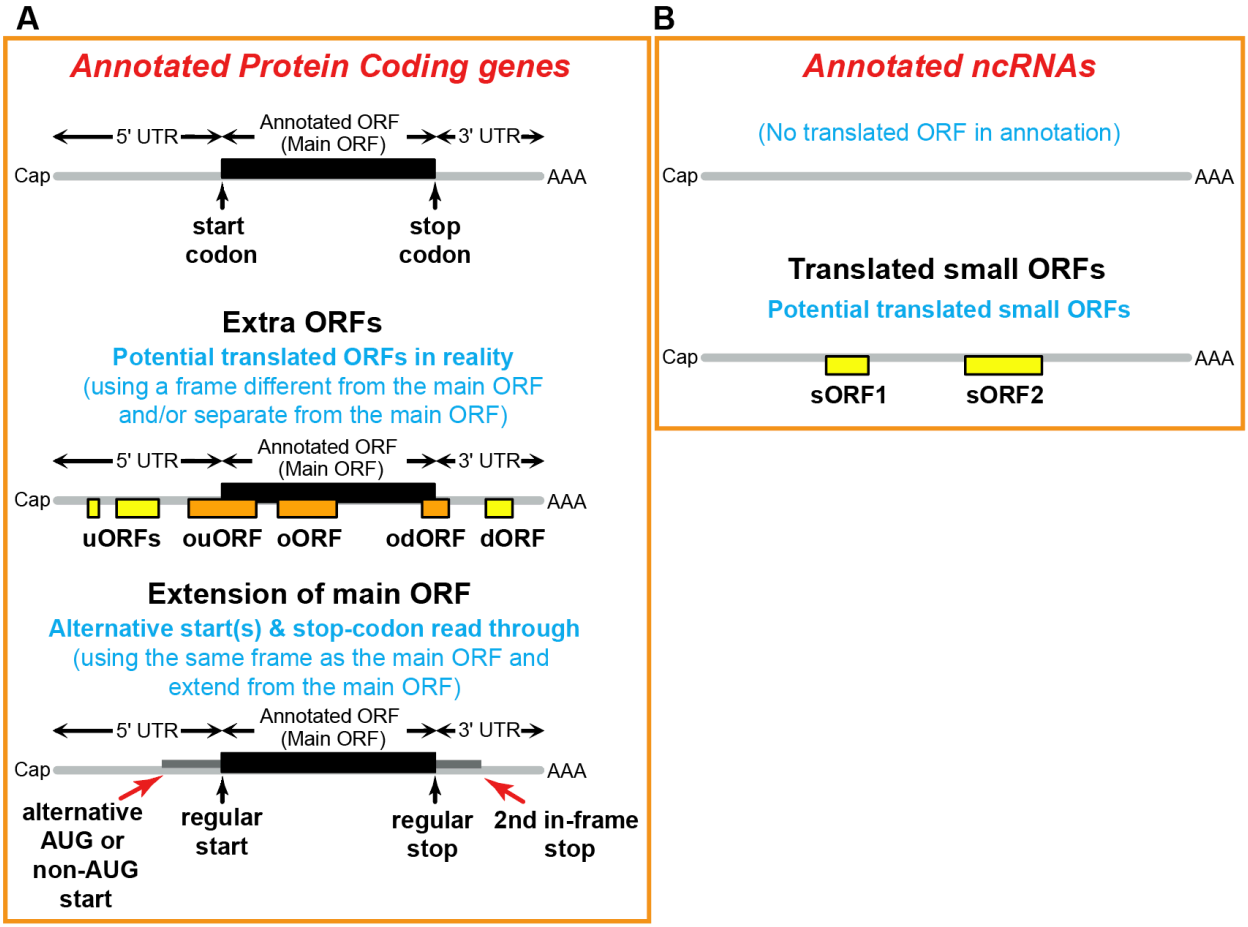

Figure 3. *ggRibo* visually presents various types of ORFs

Annotated genes are typically grouped into either coding (A) or noncoding (B), and *ggRibo* can plot ORFs in both gene types. For annotated protein-coding genes, typically only one ORF (main ORF) is annotated, while 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 5′ UTR and downstream ORFs (dORFs) within 3′ UTR. Some of the uORFs and dORFs may partially overlap with the main ORF (ouORFs and odORFs). Some ORFs may also completely overlap with the main ORF (oORF). 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 still be translated.

Figure 4

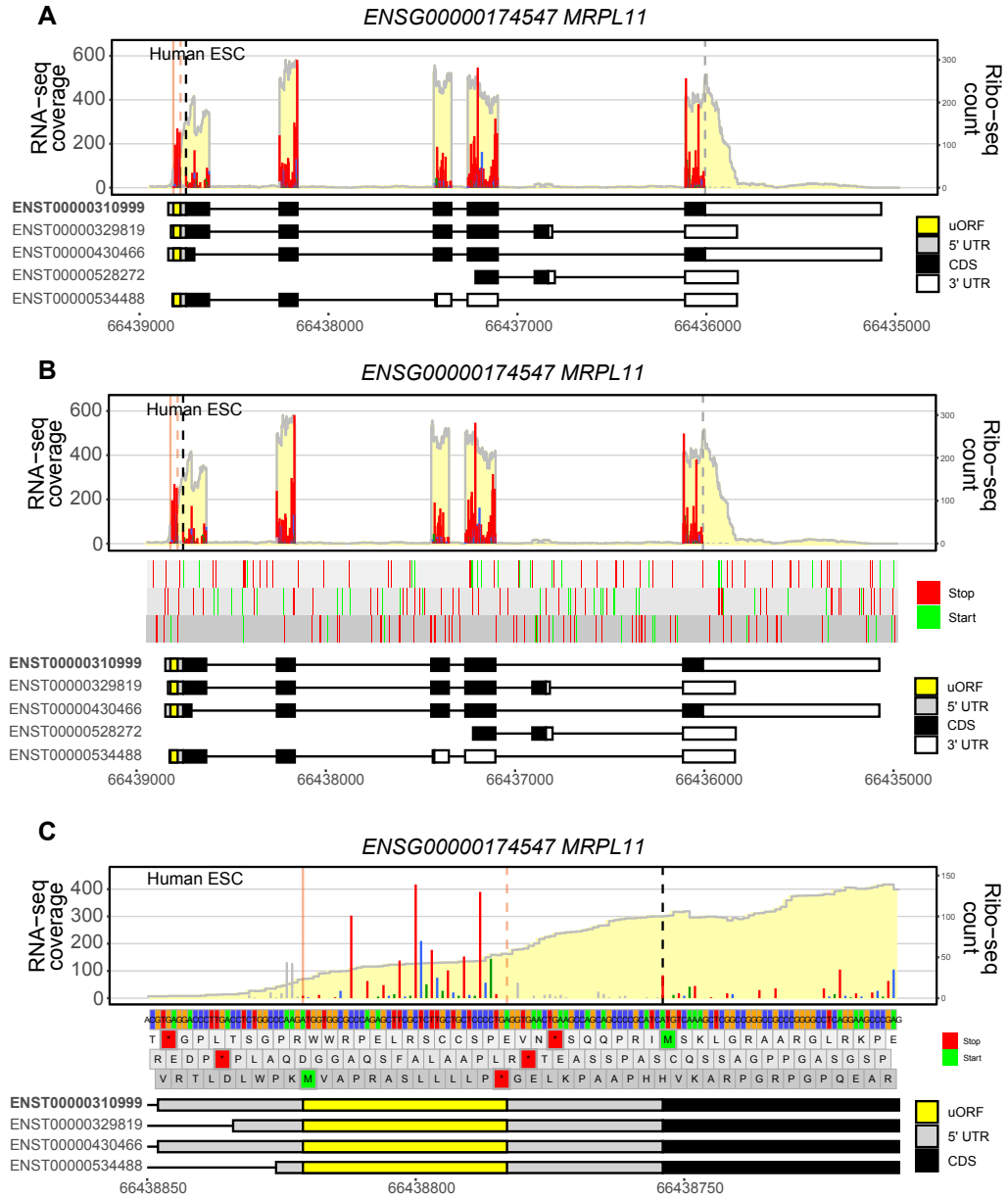

Figure 4. Examples of *ggRibo* plots for the human *MRPL11* gene

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).

(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 due 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 area <201 nt).

RNA-seq coverage is shown with a light-yellow background. Ribo-seq reads are presented with their first nt of the P-site, and they are color-coded in red, blue, and green to indicate they map

to reading frame 0 (expected), 1, and 2, respectively. *MRPL11* possesses 5 annotated transcript isoforms, and both a uORF and the main ORF are translated. The user can specify which transcript isoform to focus on and 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 for the annotated mORF, respectively. Orange solid and dashed vertical lines represent start and stop codons for the uORF, respectively. Reads outside of the ORF range are shown in grey. Genomic coordinates on the chromosomes are indicated below the gene models.

Figure 5

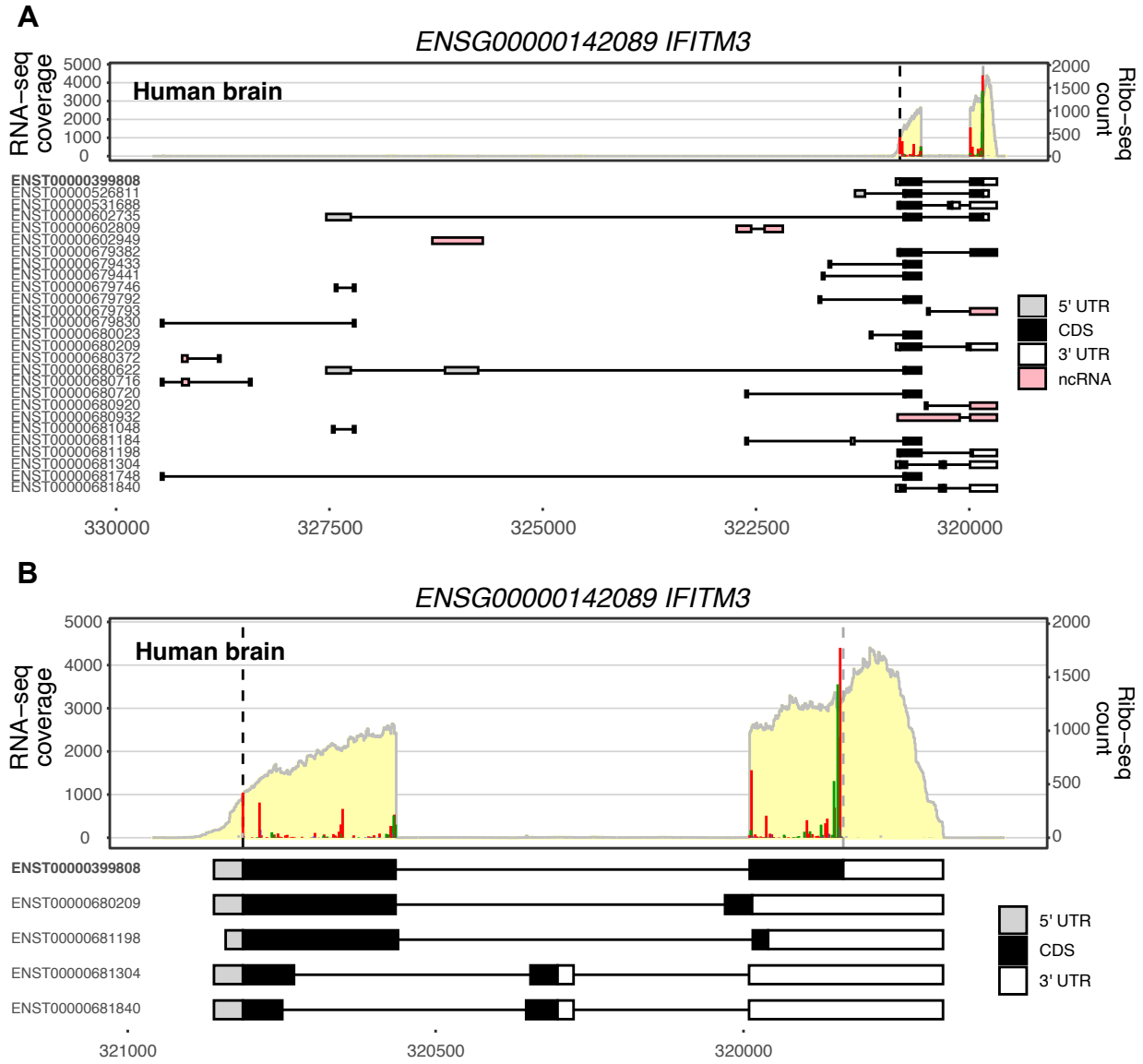

Figure 5. Examples of *ggRibo* plots for the human *IFITM3* gene

Plotting RNA-seq and Ribo-seq profiles of the human *IFITM3* gene using *ggRibo*. The human brain Ribo-seq and RNA-seq data are from (Chothani et al. 2022). Data presentations by *ggRibo* were described in Figure 4 legend.

(A) A gene view of the *IFITM3* gene. Although multiple isoforms are annotated (including some without an annotated CDS), it appears that only the ENST00000399808 isoform (bolded) is expressed and translated in this particular tissue.

(B) A closer view of ENST00000399808 and a subset of other isoforms in the same region confirms that ENST00000399808 is the major expressed and translated isoform.

Figure 6

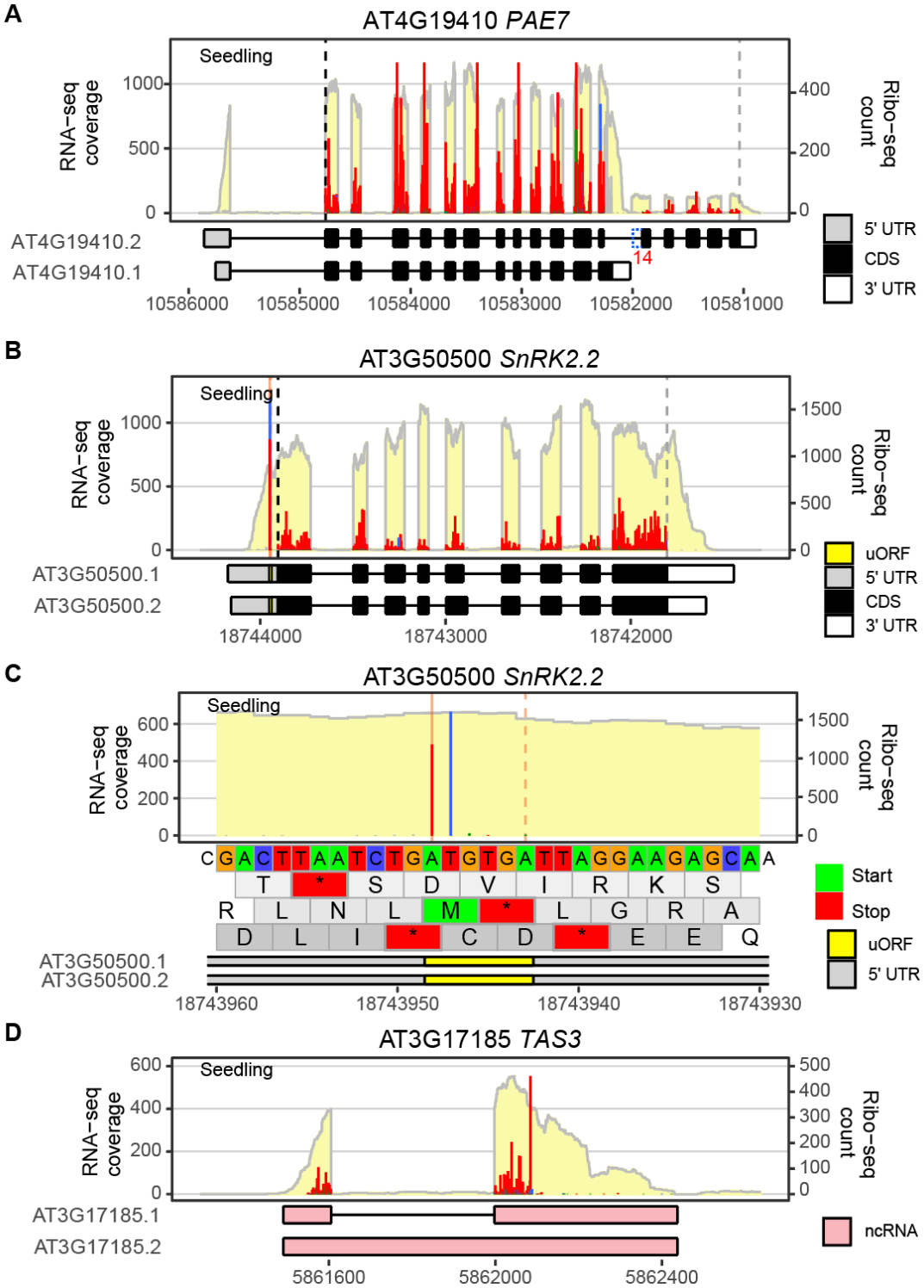

Figure 6. Examples of *ggRibo* plots with Arabidopsis genes

Plotting RNA-seq and Ribo-seq profiles of Arabidopsis genes using *ggRibo*. The Arabidopsis seedling Ribo-seq and RNA-seq data are from (Wu et al. 2024). Data presentations by *ggRibo* were described in Figure 4 legend.

(A) Gene view *of PAE7* shows that both isoforms are translated. Note 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.
(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.

Figure 7

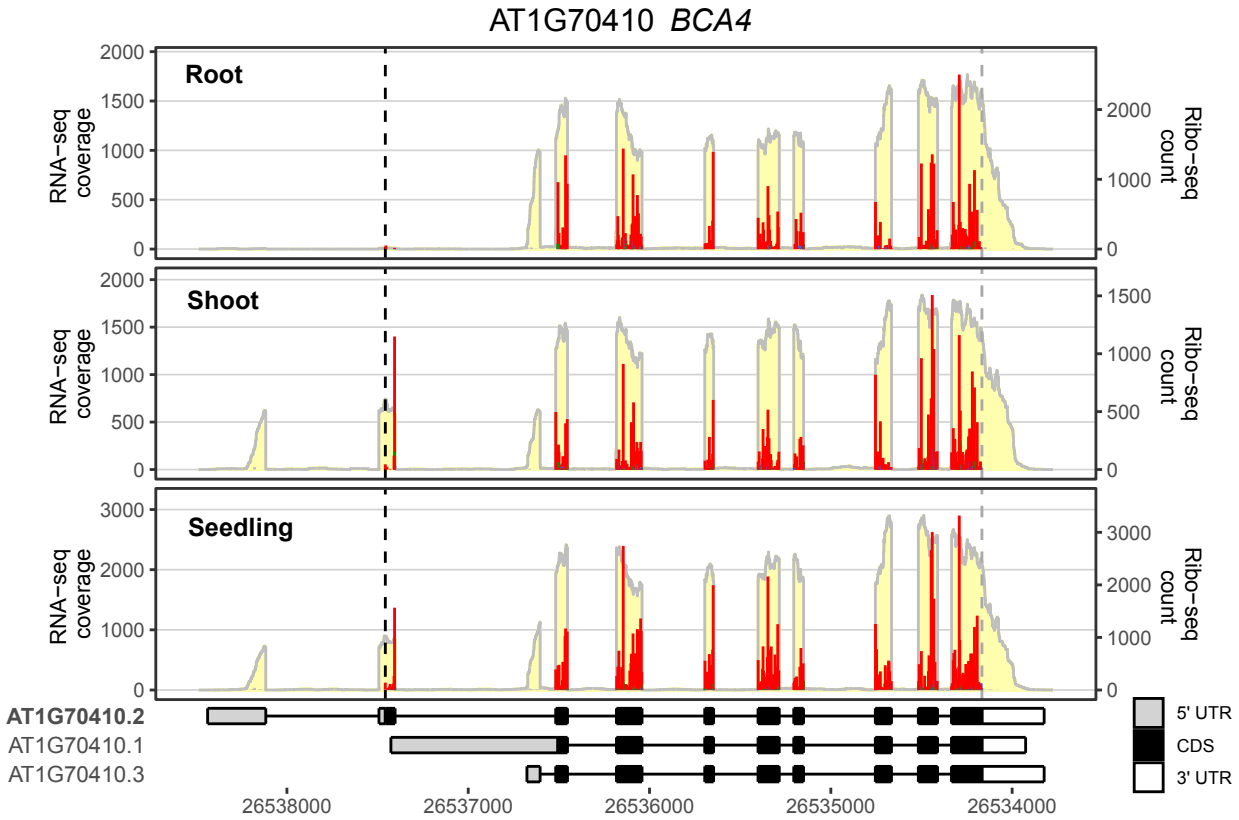

Figure 7. Comparing multiple Ribo-seq and RNA-seq datasets reveals differential isoform translation in Arabidopsis tissues

Ribo-seq/RNA-seq expression profiles of *BCA4* in Arabidopsis root, shoot, and whole seedling are compared. The root and shoot data were reported in (Hsu et al. 2016), and the seedling data were reported in (Wu et al. 2024). Note that isoform 2 is nearly undetectable in the root, while it is highly expressed and translated in the shoot and whole seedling. Since the 5' UTR for the isoform 3 was overestimated in the Araport11 annotation, the exon 1 was manually shortened to match the expressed exon range in Adobe Illustrator.

Figure 8.

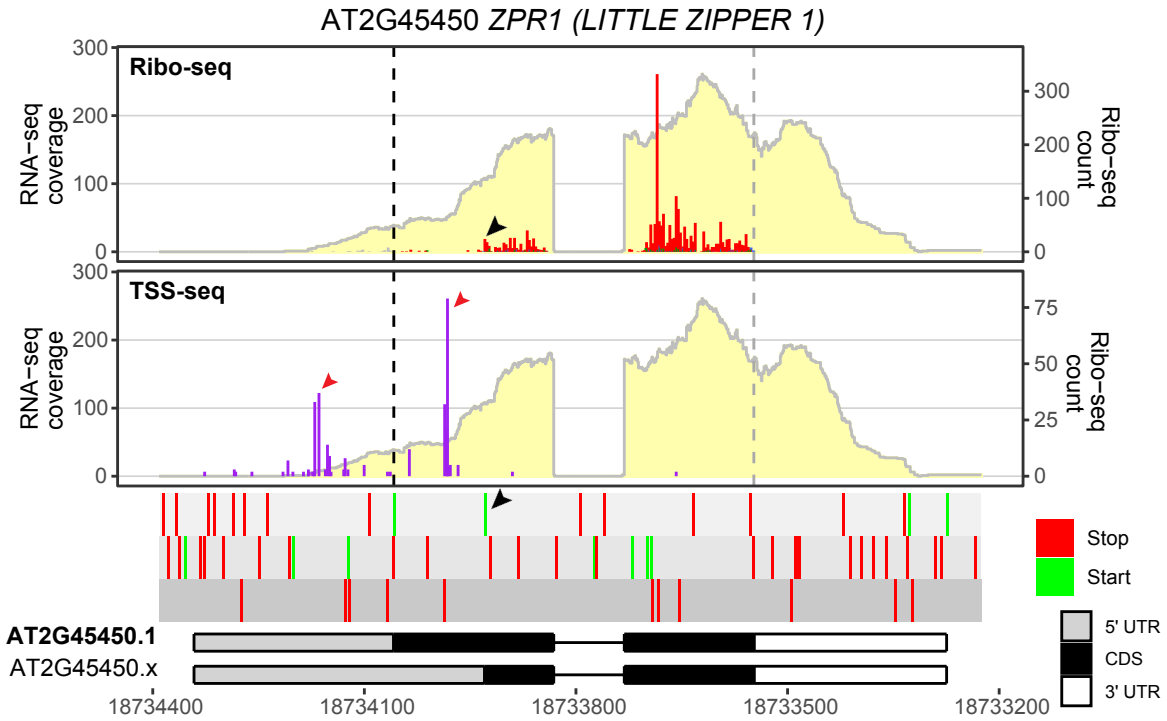

Figure 8. A multi-omic comparison reveals two transcription start site clusters in *LITTLE ZIPPER* 1

A gene view showing Ribo-seq/RNA-seq (Wu et al. 2024) 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 and one downstream of 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.

Figure 9

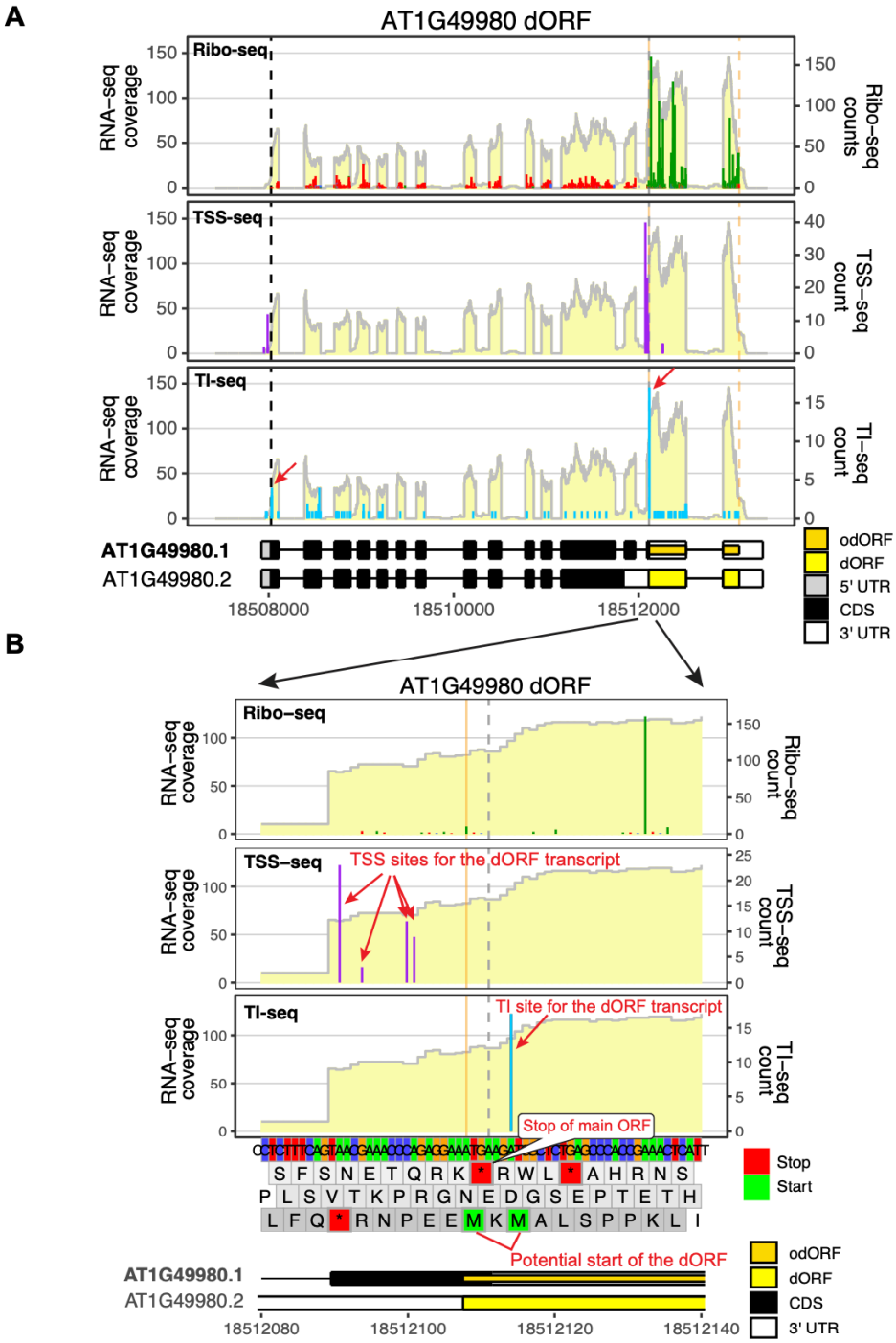

Figure 9. Multi-omic data comparison reveals the nature of a translated "dORF"
Ribo-seq and RNA seq data are reported in (Wu et al. 2024), TSS-seq data are reported in (Nielsen et al. 2019), and TI-seq data are reported in (Willems et al. 2017).

(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. 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 2 major TSS clusters (purple lines);

one corresponds to the annotated transcription start site for *AT1G49980*, and the other upstream of this dORF. The existence of a transcription start site upstream of the dORF supports this dORF arising from an unannotated gene/transcript, which has higher RNA levels than *AT1G49980*.

(B) A zoom-in view focusing on the 5′ of the dORF using the same data above. The grey 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 reveals 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), while 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.

Figure 10

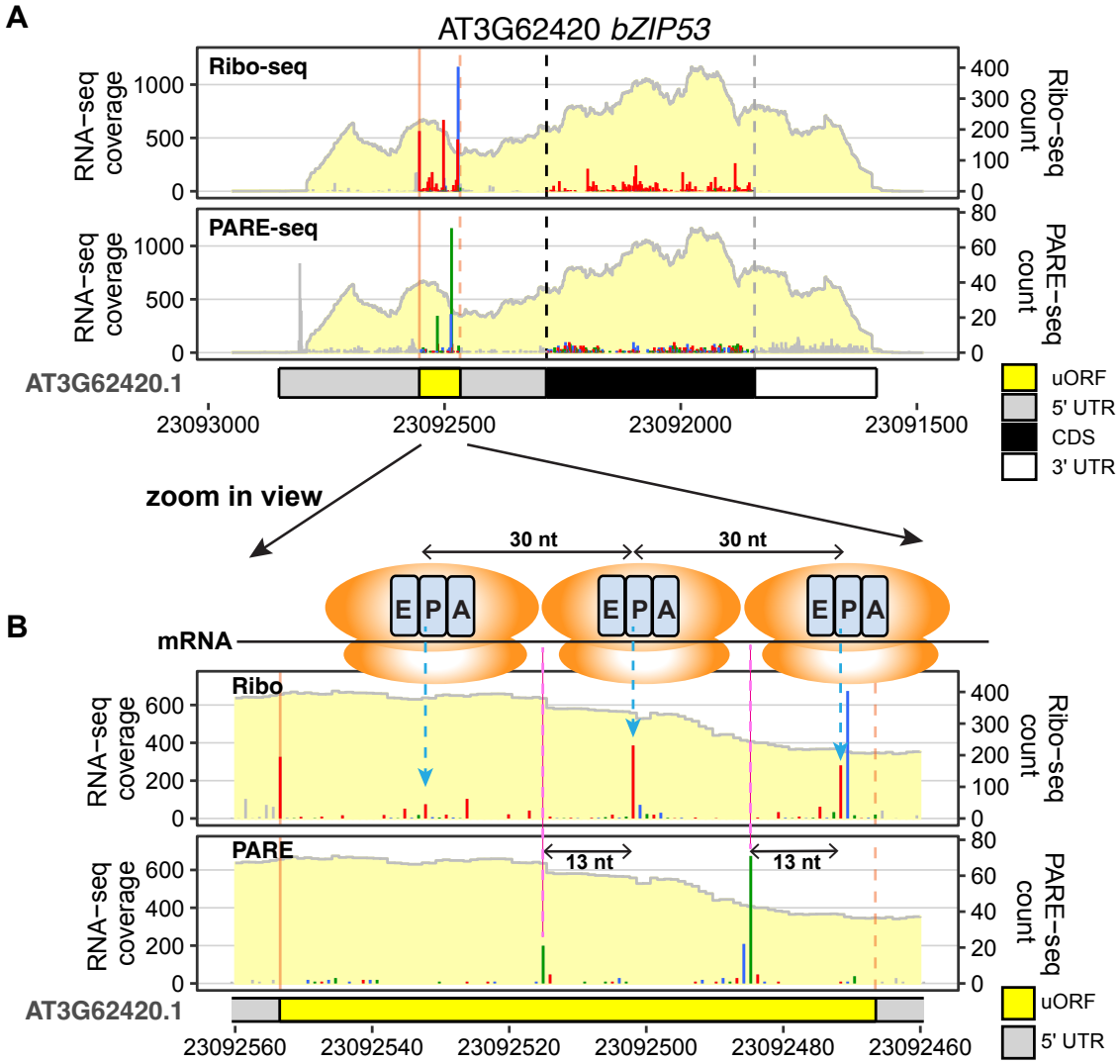

Figure 10. Visualization of co-translational mRNA decay in *bZIP53* conserved peptide uORF. Ribo-seq and RNA seq data are reported in (Wu et al. 2024), and PARE-seq data are reported in (Nagarajan et al. 2019).

(A) A gene view showing Ribo-seq and PARE-seq in parallel for *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.

(B) A zoom-in view of the conserved peptide uORF region. Accumulation of Ribo-seq reads at 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 co-translational 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 S1.

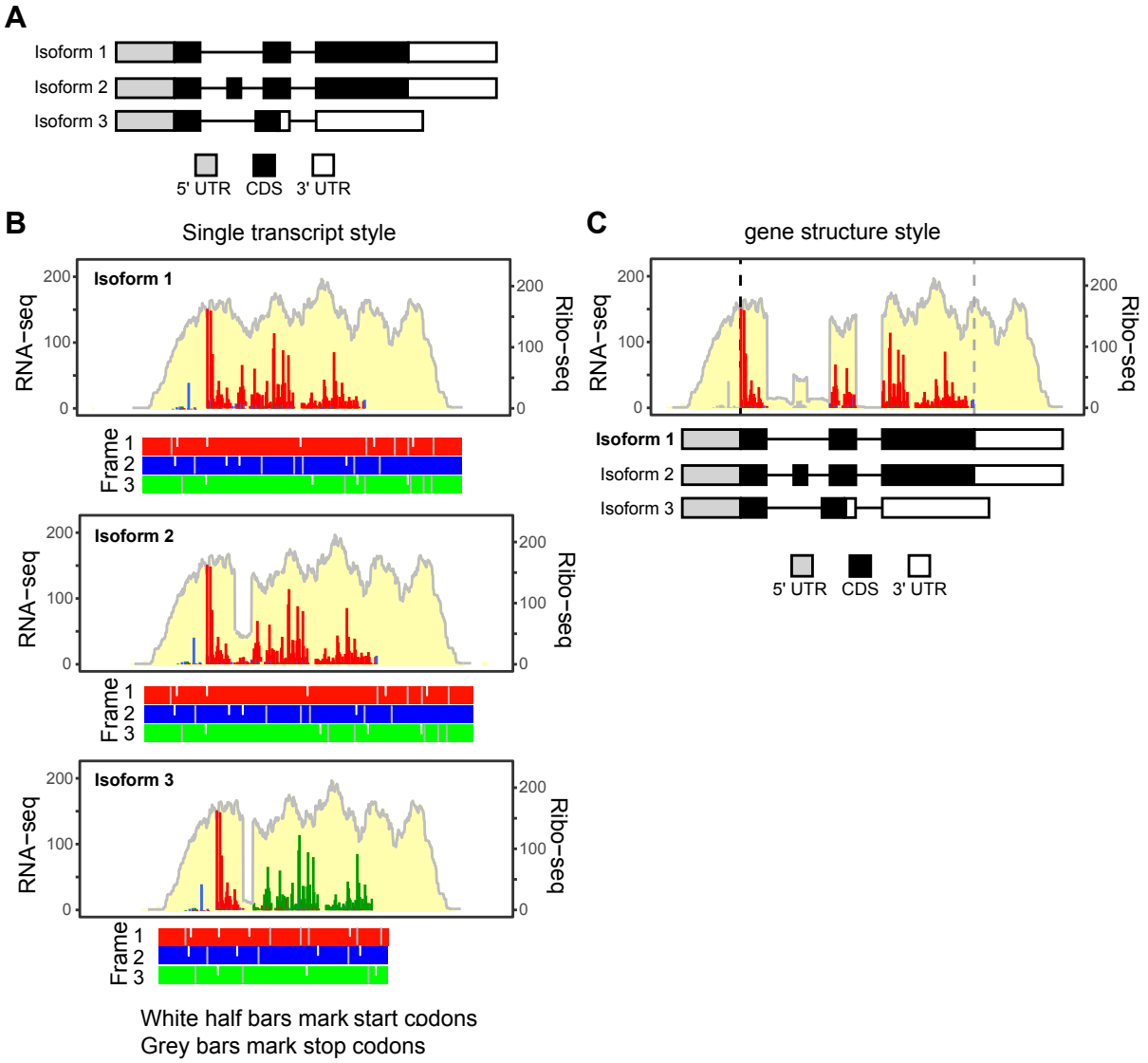

Figure S1. Single transcript style vs. gene structure style for Ribo-seq visualization

A) Gene structure of a hypothetical gene encoding three RNA isoforms.

B) In the single transcript style, the mature mRNA of each isoform is presented separately. Ribo-seq reads in different reading frames are color-coded to show the periodicity, and RNA-seq coverage is plotted as the light-yellow background for each panel. The start codon and stop codon are shown with white half bars and grey bars, respectively. Visually, it is challenging to identify which RNA isoform is expressed or translated.

C) In the gene-structure style, both UTRs and introns are included, and gene models of each isoform are presented in parallel. Ribo-seq reads in different reading frames are color-coded to show the periodicity, and RNA-seq coverage is plotted as the light-yellow background. In this example, isoform 1 is selected (bold), and therefore reading frames are assigned according to the annotation of this isoform. Reads outside of its ORF range are shown in grey. The start codon and stop codon are shown with black and grey dashed lines, respectively. Even though the color coding is based on isoform 1, we can tell that isoform 2

is also expressed based on RNA-seq coverage in its 2nd exon. Isoform 2 is likely also translated based on Ribo-seq reads (grey) in the 2nd exon. In contrast, isoform 3 is likely not expressed or translated as there is no RNA-seq or Ribo-seq support for its 2nd exon. Additionally, we can observe significant intron retention between 1st and 2nd exons for isoform 1.

Figure S2. A summary of *ggRibo* improvement compared to *RiboPlotR.*

	RiboPlotR	***ggRibo***
Plot system	Base R	ggplot2
Object oriented programming	No	Yes (R6)
Annotation style	Transcript id = gene id + . + number	Any gtf/gff3 accepted by GenomicFeatures
Number of samples	2~3	Any as RAM allows
ORF types visualized	uORF, dORF, mORF	uORF, dORF, mORF ouORF, odORF, oORF, ncORF
Visualize Ribo-seq reads on ncRNA	No (need a defined ORF)	Yes
Zoom in on a specific region of a gene	No	Yes (use plot_range)
Adjust the y-axis of Ribo-seq to see lowly translated events	No	Yes (use Ribo_fix_height)
Extend main ORF frame to 5′ UTR or 3′ UTR (check non-AUG start or stop codon readthrought)	No	Yes (use fExtend for 5′ UTR and tExtend for 3′ UTR)
Change frame colors from red, blue, green	No	Yes (use frame_colors)
Plot DNA sequence	No	Yes (set show_seq = TRUE, and input FASTA file)
Plot direction	As genome browser	Always left to right (always 5′ -> 3′)

Figure S3

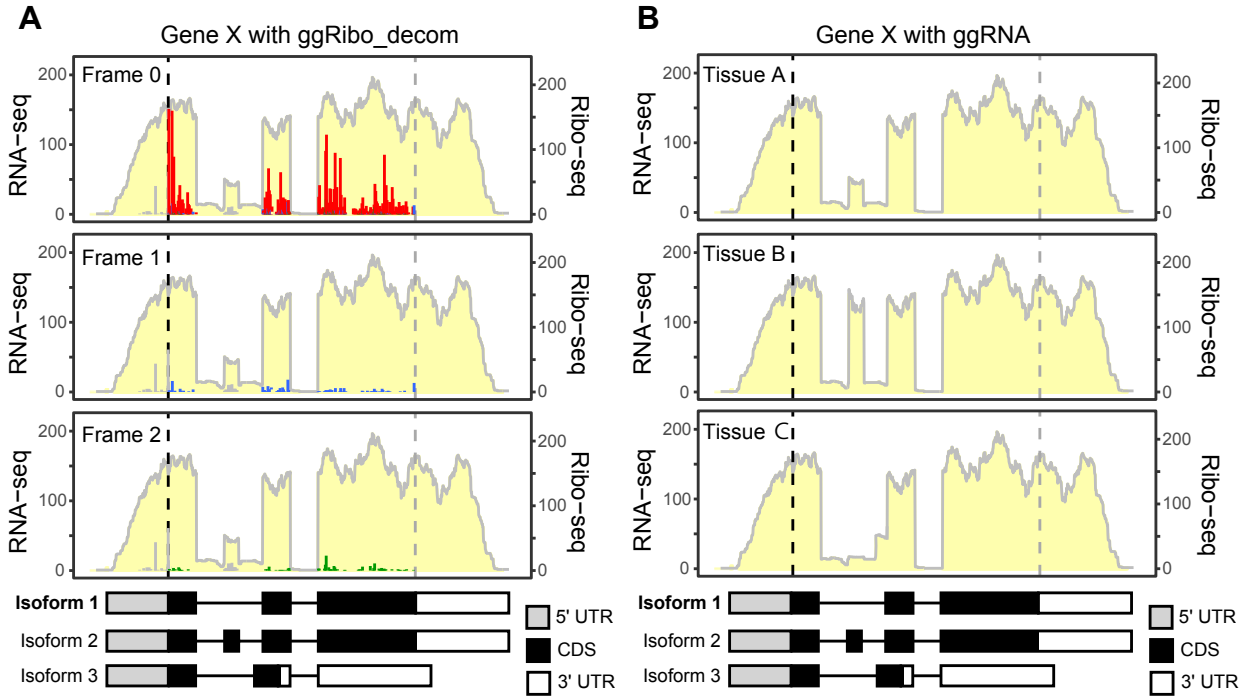

Figure S3. *ggRibo_decom* plots Ribo-seq reads in each frame separately, and *ggRNA* plots RNA-seq coverage

A) *ggRibo_decom* takes one dataset (i.e., one Ribo-seq and one RNA-seq dataset) and separates Ribo-seq reads in different frames. The selected isoform can be either coding or noncoding, and reads that fall outside of the corresponding CDS are colored in gray. Users can use *ggRibo_decom* to visualize whether Ribo-seq reads are enriched in one or more frames.

B) *ggRNA* can handle multiple single-end or paired-end RNA-seq datasets. Users can leverage *ggRNA* to plot coverage (using BAM files) for RNA-seq, Ribo-seq, or other sequencing data, providing an effective initial quality check for new datasets. Ribo-seq reads should be enriched in the coding sequences or known translated uORFs.

Figure S4

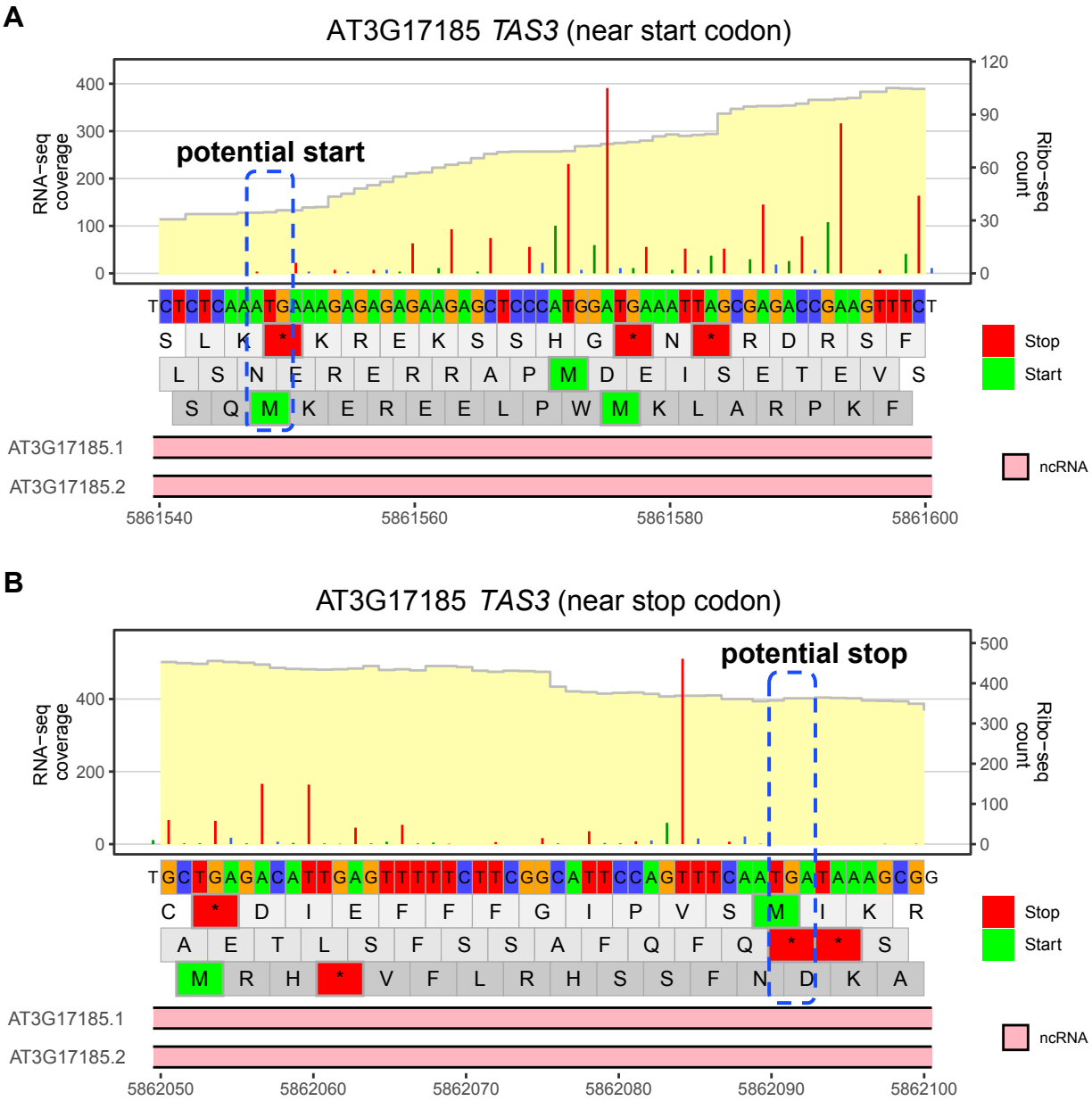

Figure S4. Zoom in plots for Arabidopsis *TAS3* to view the regions near the start and stop codons

Regions near the start codon (A) and stop codon (B) of *TAS3*. The potential start and stop are highlighted and annotated in Adobe Illustrator.

Figure S5

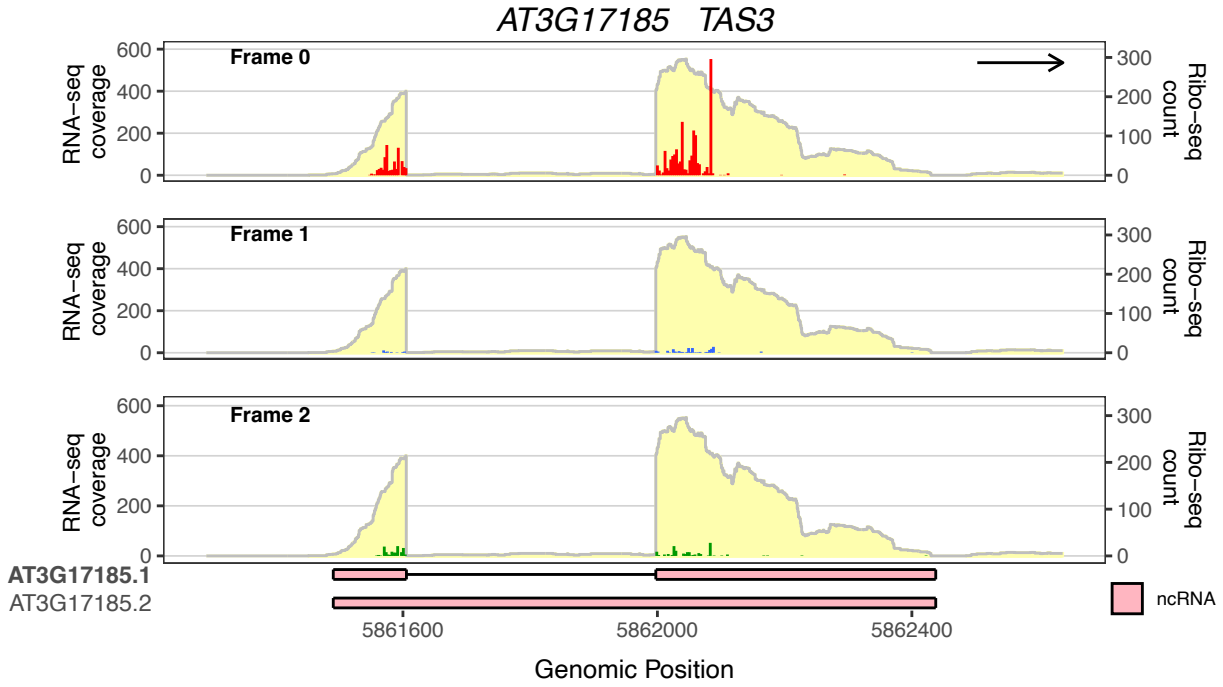

Figure S5. Plotting Arabidopsis *TAS3* gene using *ggRibo_decom* function
ggRibo_decom function is used for the decomposition of the Ribo-seq reads in different reading frames for *TAS3*. The arrow on the top right, which points to the right, indicates the gene direction is consistent with the direction shown on the genome browser. Ribo-seq and RNA-seq data are reported in (Wu et al. 2024).

Figure S6

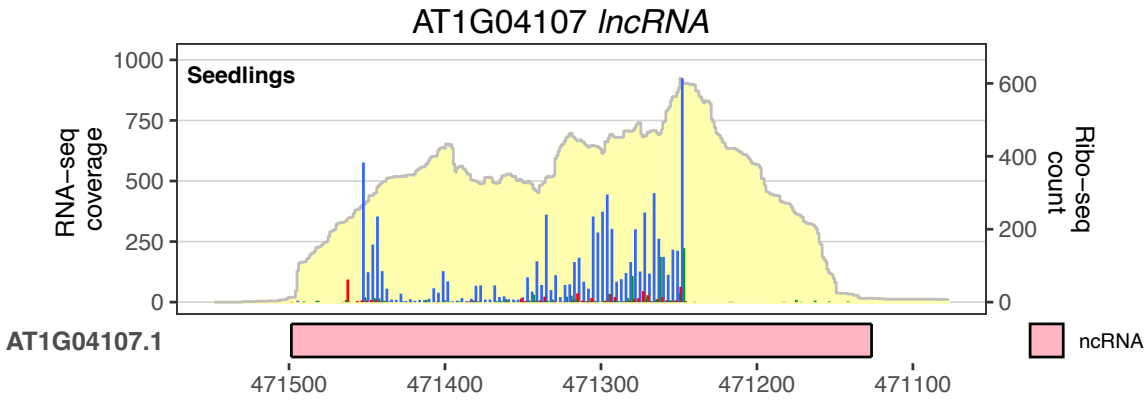

Figure S6. Example of a novel sORF identified within a lncRNA using the 'blue' reading frame. Ribo-seq and RNA-seq expression profiles of a novel sORF identified in an annotated lncRNA in Arabidopsis seedlings (Wu et al. 2024) are shown. The strong 3-nt periodicity of Ribo-seq reads in the 'blue' reading frame suggests the presence of a translated sORF.