Supplemental Material

Supplemental Figures

Supplemental Figure S1. Profiling of initiating ribosomes in tomato.

Supplemental Figure S2. Similarity of gene expression between biological replicates of the LTM, CHX and mRNA samples.

Supplemental Figure S3. LTM, CHX and mRNA read densities for tomato and *Arabidopsis* genes.

Supplemental Figure S4. The types of ORFs initiated from alternative TISs located in 5'UTRs and CDSs in tomato.

Supplemental Figure S5. Codon composition of CDSs for tomato, *Arabidopsis* and human genes.

Supplemental Figure S6. The correlation between the LTM read density in a given TIS and the CHX/mRNA read densities in the corresponding ORF for upstream non-AUG TISs.

Supplemental Figure S7. The correlation between alternative AUG/near-cognate TISs and translation efficiency/transcript abundance in *Arabidopsis*.

Supplemental Figure S8. The impact of alternative in-frame TISs on protein subcellular localization.

Supplemental Figure S9. The conservation of alternative TIS peaks and corresponding ORF types between tomato and *Arabidopsis*.

Supplemental Tables

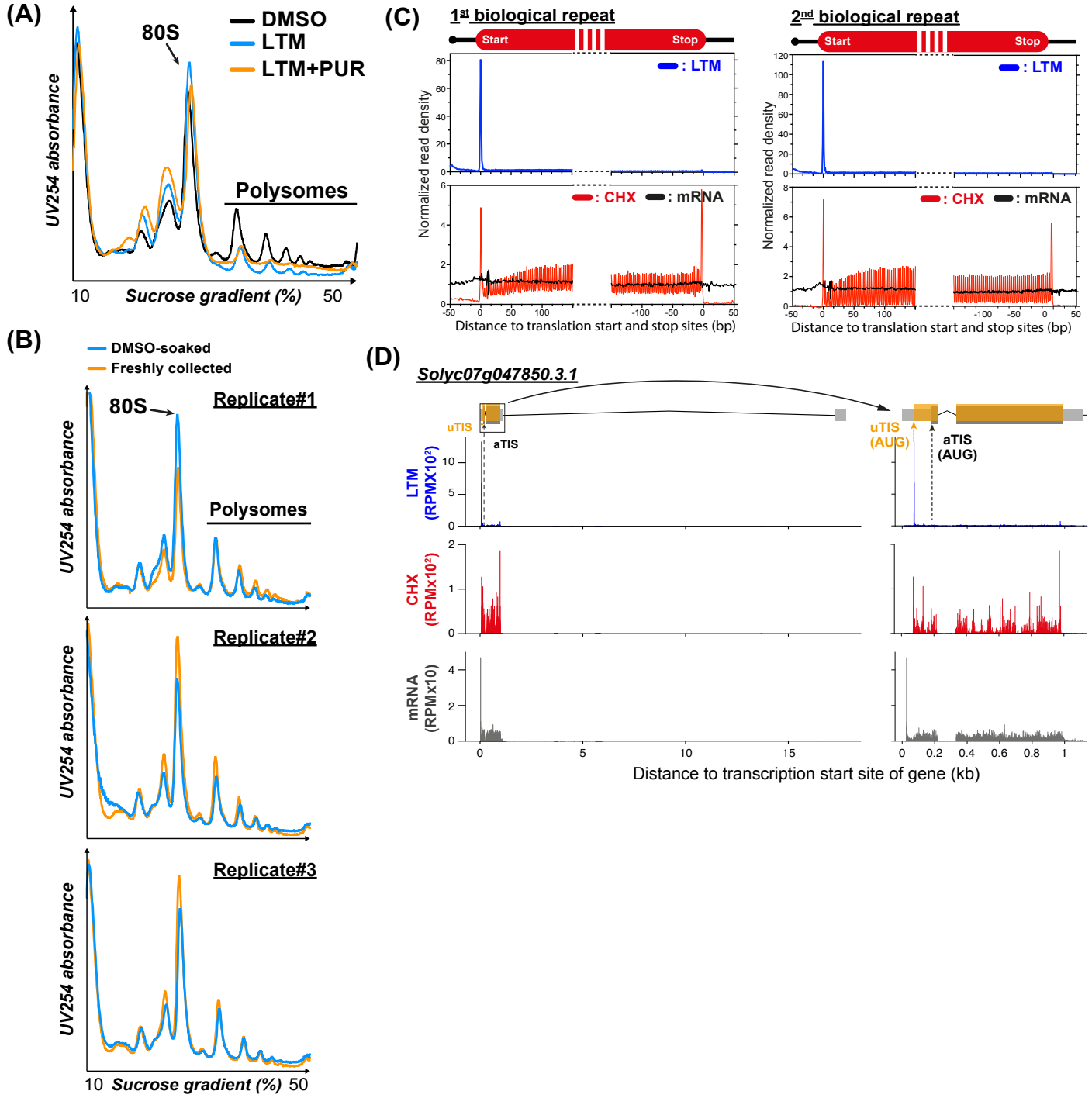
Supplemental Table S1. List of *in vivo* translation initiation sites identified in the study.

Supplemental Table S2. Summary of 5'-extended/truncated ORFs and their predicted localization.

Supplemental Table S3. RNA-seq mapping statistics for the LTM, CHX and mRNA samples.

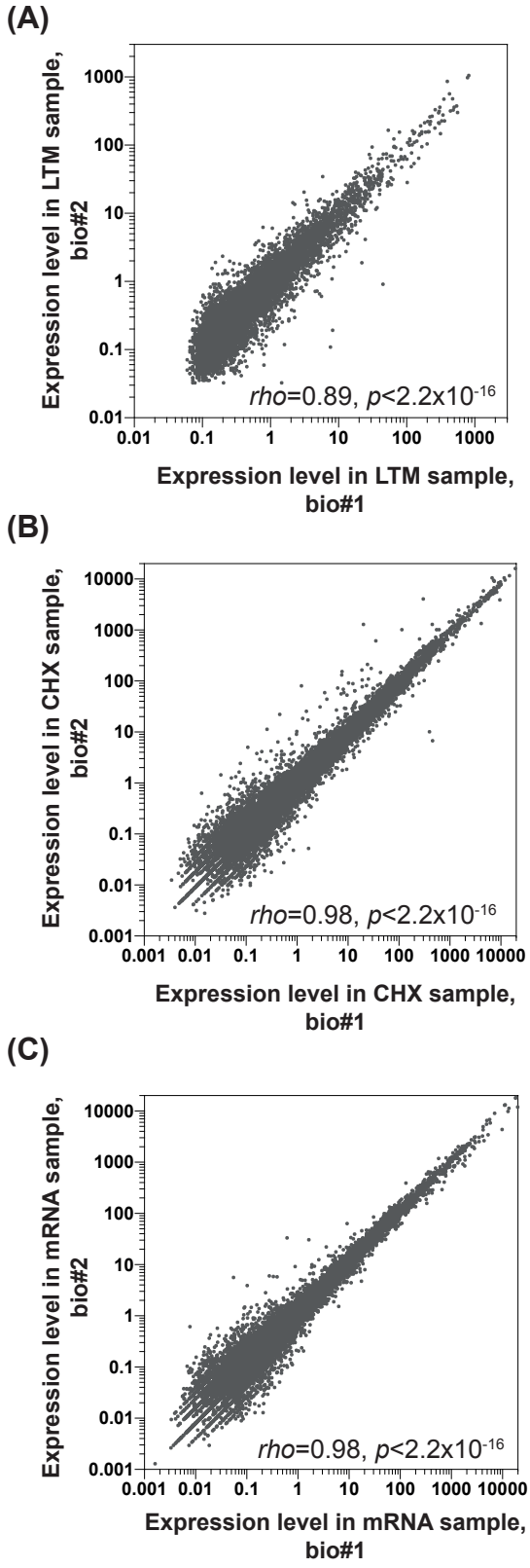
Supplemental Table S4. List of primers used in the study.

Supplemental Figure S1



Supplemental Figure S1. Profiling of initiating ribosomes in tomato.
(A) Polysome profiling analyses of the samples treated with LTM (lactimidomycin) (blue) and LTM plus puromycin (PUR) (orange). Monosomes (80S) and polysomes are marked. DMSO: control treatment (black). (B) Polysome profiling analyses of a freshly collected leaf sample (i.e., the excised leaves immediately frozen in liquid nitrogen; orange) and a DMSO-soaked leaf sample (i.e., the excised leaves soaked in DMSO solution for 30 min with gentle shaking; blue) to reveal the effects of hypoxia on global translational status in three biological replicates. (C) Metagene plots for mean read densities in regions around the annotated translation start and stop sites of genes calculated for the LTM, CHX and mRNA samples. Only genes with a 5'UTR and 3'UTR ≥ 20 nt were included. Sequencing reads from two independent biological repeats were analyzed and are plotted separately. (D) As indicated in (C), but for a single gene, *Solyc07g047850*, with alternative in-frame TISs located in the 5'UTR. Within the gene models (top), light gray boxes indicate UTRs, dark gray boxes indicate annotated CDSs, thin lines indicate introns, orange boxes indicate alternative TIS-initiated CDSs and black and orange arrows indicate the annotated and upstream alternative TISs (aTIS and uTIS), respectively. RPM: reads per million mapped reads. The full-length gene model (left) and a partial model with aTIS and uTIS (right) are shown.

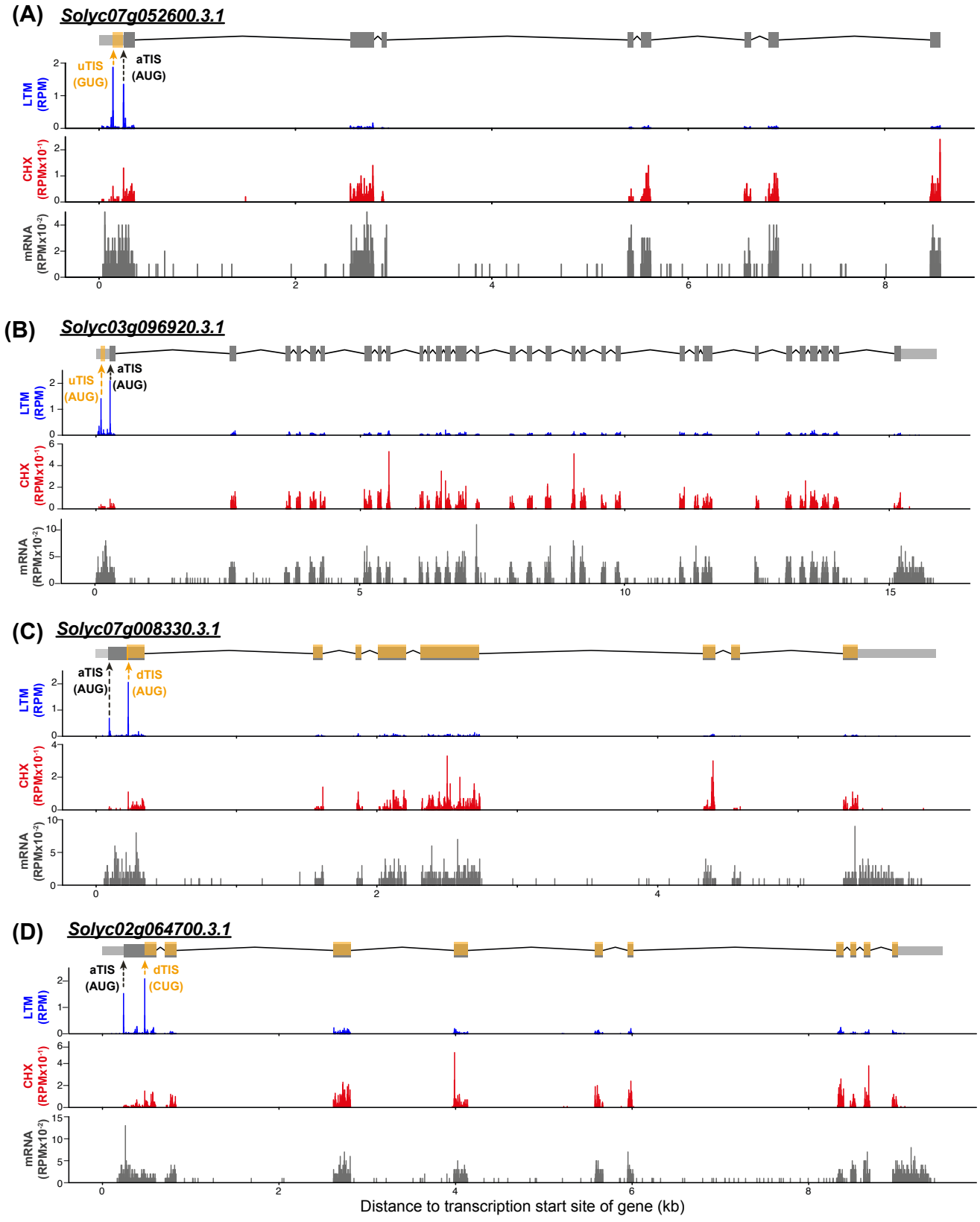
Supplemental Figure S2

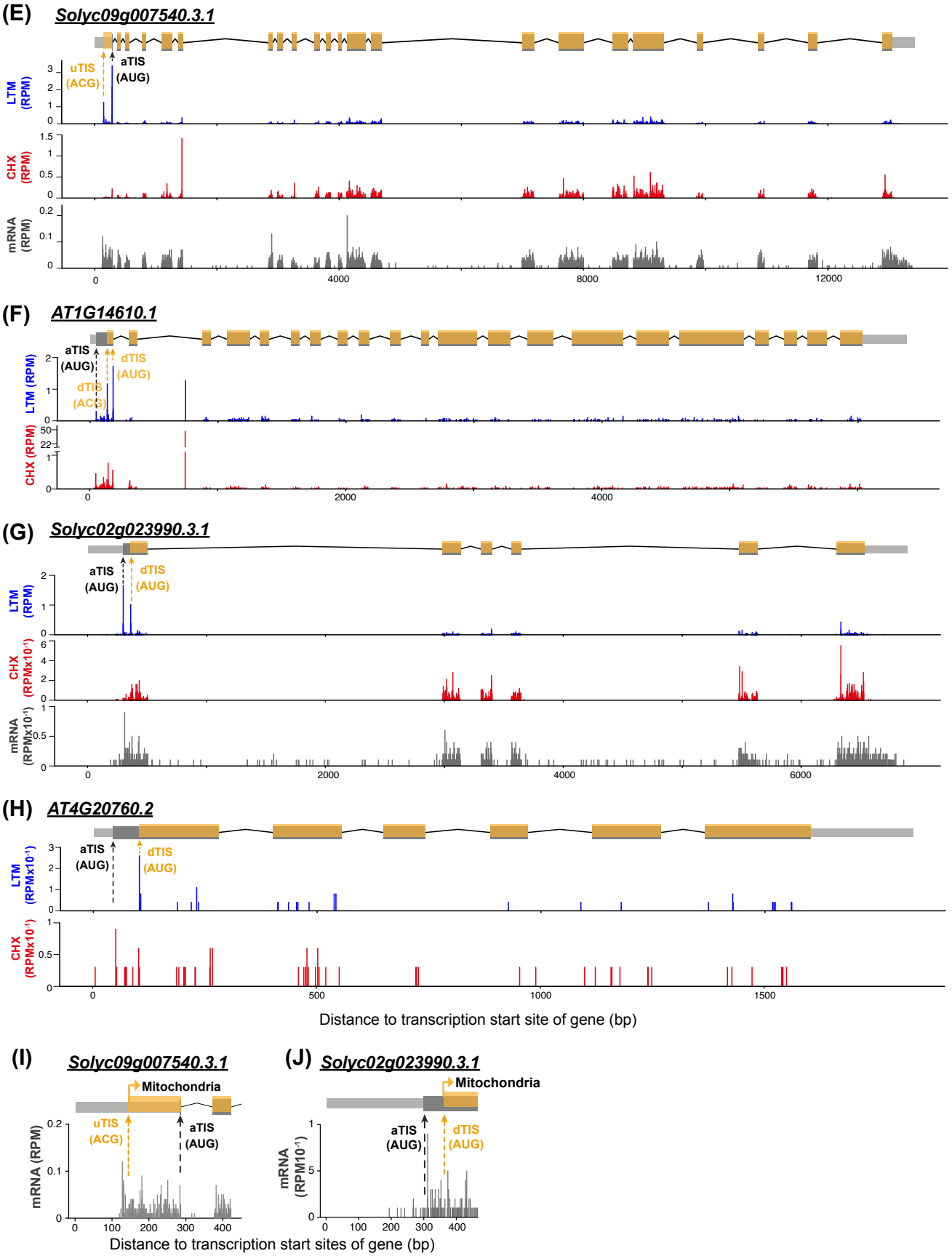


Supplemental Figure S2. Similarity of gene expression between biological replicates of the LTM, CHX and mRNA samples.

(A) Scatter plots showing the correlation of normalized LTM read intensities for the identified translation initiation sites (TISs) between biological replicates. The TISs identified in both replicates were included. The normalized LTM read intensity was determined by normalizing the read counts at a given TIS to the upper quintile of the read counts at all TISs, as described in a previous study (Gao et al., 2015). (B, C) As described in (A), but for CHX read densities determined from reads mapped to CDSs (B) and for mRNA read density determined from reads mapped to transcript regions (C) of a given gene. RPKM: reads per million mapped reads per kilobase. *Rho*: Spearman's rank correlation coefficient.

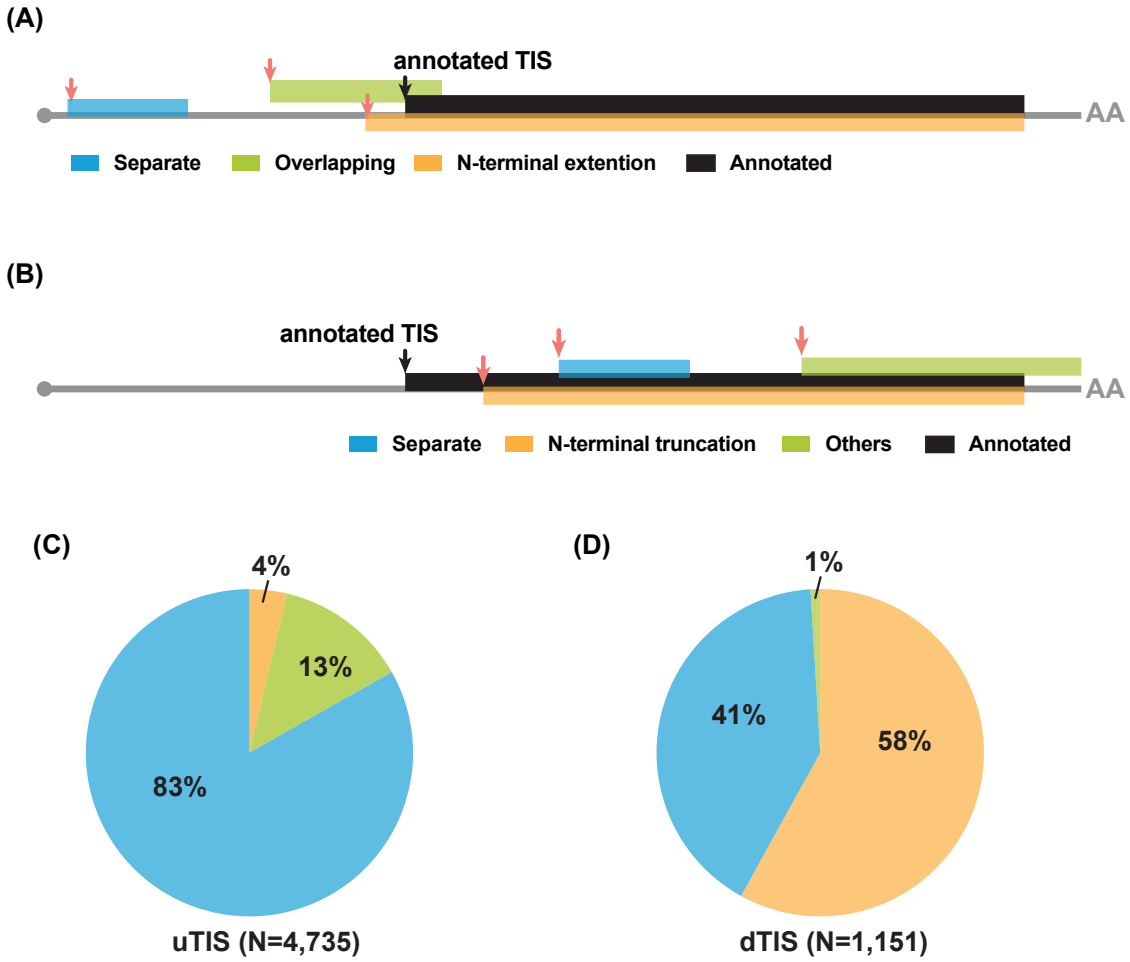
Supplemental Figure S3





Supplemental Figure S3. LTM, CHX and mRNA read densities for tomato and *Arabidopsis* genes.
(A-H) Plots for read densities in LTM, CHX and mRNA samples within the full-length gene models for genes with alternative TISs. Within the gene models (top), light and dark gray boxes indicate UTRs and annotated CDSs, thin lines indicate introns, orange boxes indicate alternative TIS-initiated ORFs and black and orange arrows indicate the annotated and upstream/downstream alternative TISs (aTIS and uTIS/dTIS). RPM: reads per million mapped reads. (I,J) As indicated in (A-H), but shown for the read densities in the mRNA sample within the partial gene models with alternative and annotated TISs (right).

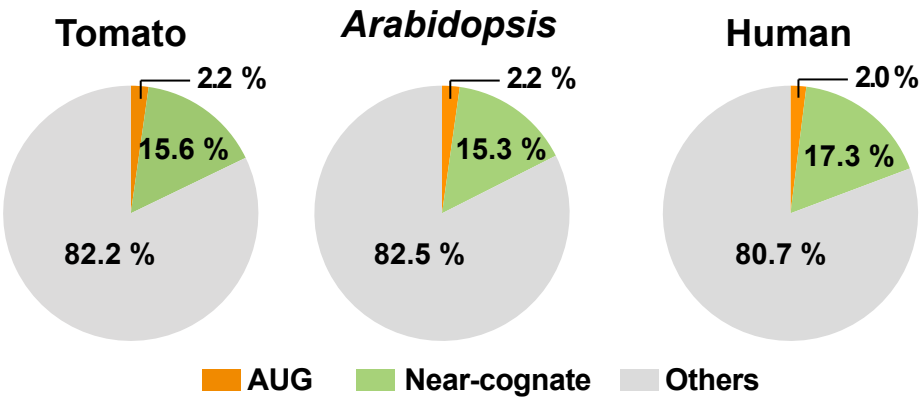
Supplemental Figure S4



Supplemental Figure S4. The types of ORFs initiated from alternative TISs located in 5'UTRs and CDSs in tomato.

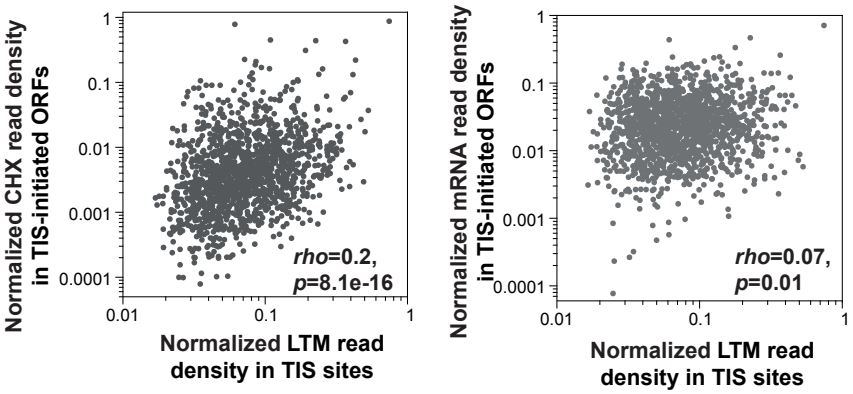
(A) Illustration of ORFs initiated from upstream TISs (uTISs, located in 5'UTR, pink arrows), which are separate from (blue box; separate), overlap with (green box; overlapped) and have an N-terminal extension (orange box; N-terminal extension) compared with the annotated ORF (black box). Gray line: mRNA; black arrow: annotated TIS. (B) As described in (A), but for downstream TIS (dTIS, located in CDS, pink arrows)-initiated ORFs, which are out-of-frame and different from the annotated ORF (blue box; separate), have an N-terminal truncation (orange box; N-terminal truncation) compared with the annotated ORF (black box), or lack a stop site in the downstream transcribed region (green box; others). (C,D) Pie charts showing the proportion (%) of the types of ORFs initiated from uTISs (C) and dTISs (D) in tomato. The types of ORFs in (C) and (D) are illustrated in (A) and (B), respectively.

Supplemental Figure S5



Supplemental Figure S5. Codon composition of CDSs for tomato, *Arabidopsis* and human genes.
Pie charts showing the proportion (%) of annotated CDSs of all protein-coding genes in tomato, *Arabidopsis* and humans with AUG (orange), near-cognate codons (green) and all other codons (gray).

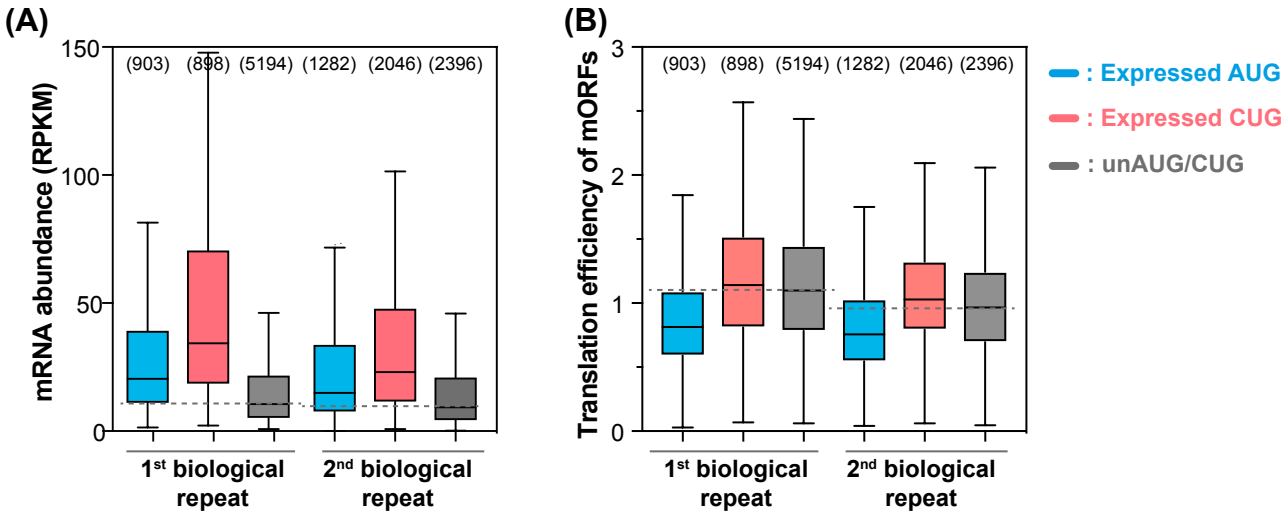
Supplemental Figure S6



Supplemental Figure S6. The correlation between the LTM read density in a given TIS and the CHX/mRNA read densities in the corresponding ORF for upstream non-AUG TISs.

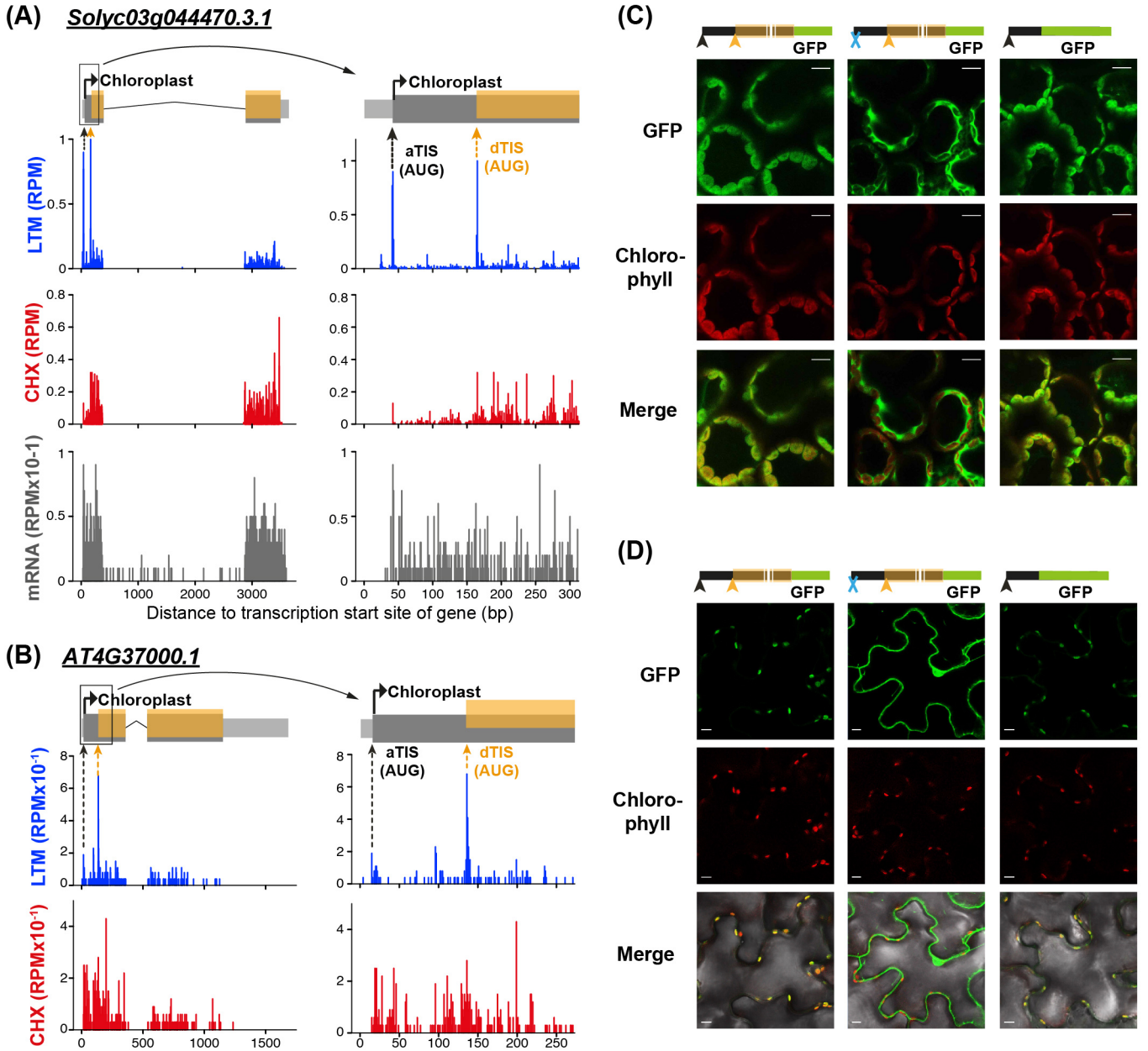
The non-AUG TISs (n=1,571) that are located in 5'UTRs and lead to separate upstream ORFs were included. The normalized LTM read densities in a given TIS (x-axis) were determined by normalizing the LTM read counts of the TIS (a 5-nt window flanking the TIS) to the LTM read counts in the transcribed regions of a gene with the TIS in question. The normalized CHX (left panel) and mRNA (right panel) read intensities in a given TIS-initiated ORF (y-axis) were determined by normalizing the CHX/mRNA read counts in the ORF to the CHX and mRNA read counts in the transcribed regions of a gene with the TIS in question, respectively.

Supplemental Figure S7



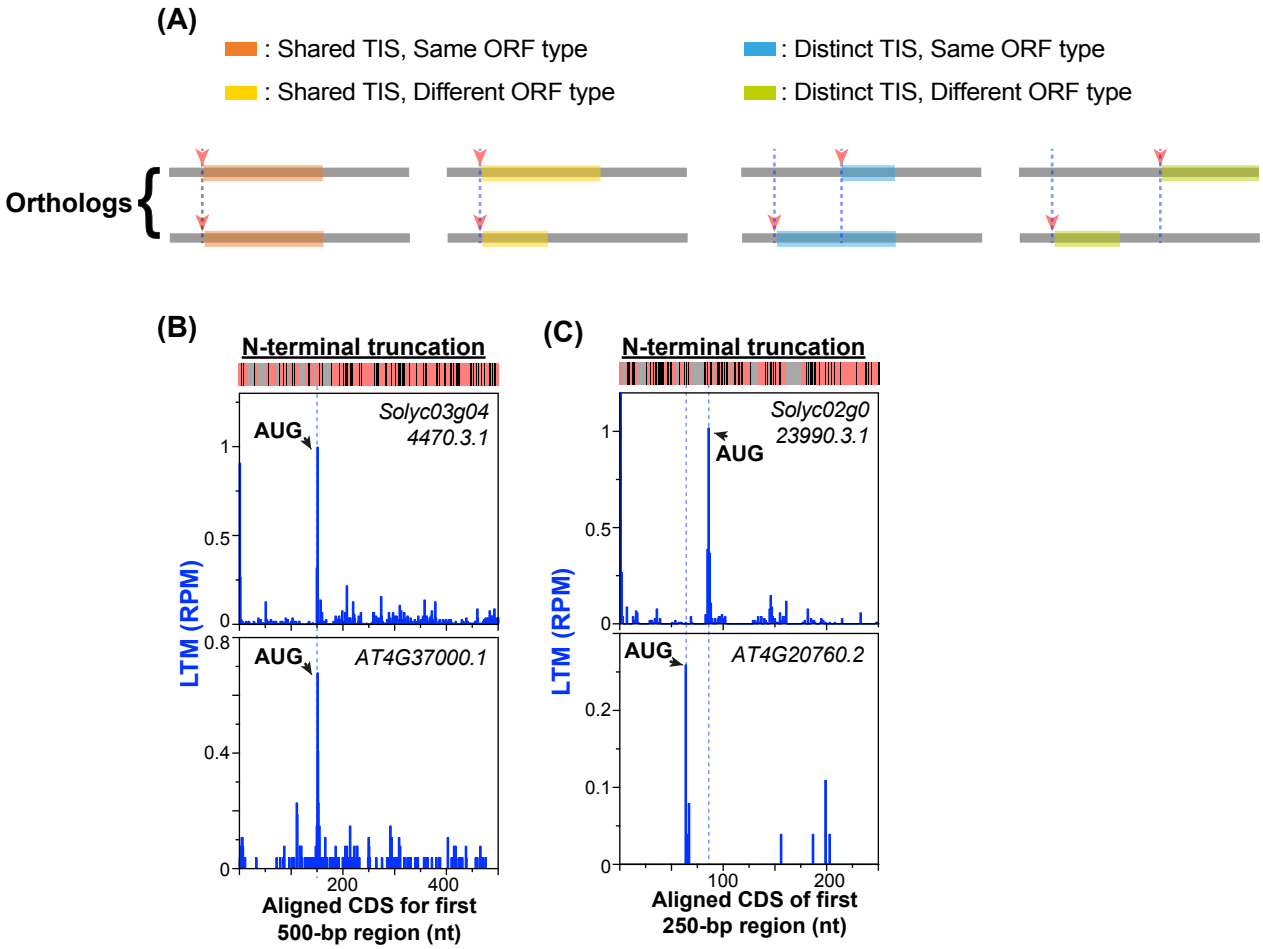
Supplemental Figure S7. The correlation between alternative AUG/near-cognate TISs and translation efficiency/transcript abundance in *Arabidopsis*.
(A) The steady-state mRNA abundances for *Arabidopsis* genes with 5'UTRs containing exclusively AUG-initiated (blue) and exclusively CUG-initiated (pink) TISs and genes without AUG/CUG-initiated TISs in the 5' UTR (gray). Steady-state mRNA abundances for two independent biological replicates are shown. Only genes with an AUG or CUG site in the 5'UTR and with RPKM values ≥ 1 in the CDS in both the CHX and mRNA RNA-seq data sets were included. (B) As described in (A), but for translational efficiencies of main open reading frames (mORFs). The number of genes within a given group is shown in parentheses.

Supplemental Figure S8



Supplemental Figure S8. The impact of alternative in-frame TISs on protein subcellular localization. (A,B) As described in **Fig. S1D**, but for the orthologous gene pair *Solyc03g044470* (A) and *AT4G37000* (B), which have downstream in-frame AUG TISs (orange arrows) and encode N-terminally truncated protein isoforms. The TIS giving rise to the protein isoform with a predicted chloroplast targeting signal is indicated. The full-length gene model (left) and a partial model with aTIS and dTIS (right) are shown. (C,D) Confocal images showing the localization of Solyc03g044470-GFP proteins in tobacco mesophyll (C) and epidermal cells (D). Solyc03g044470-GFP proteins were encoded by the wild-type full-length CDS (left panels), annotated TIS-mutated full-length CDS (middle panels), and the CDS region between the annotated TIS (aTIS) and downstream TIS (dTIS) (right panels) and transiently expressed in tobacco leaves. Scale bar = 10 μm. The aTIS, dTIS and mutated TIS are shown as a black arrow, orange arrow, and a blue cross, respectively.

Supplemental Figure S9



Supplemental Figure S9. The conservation of alternative TIS peaks and corresponding ORF types between tomato and *Arabidopsis*.
(A) The alternative TISs (pink arrows) identified in a tomato-*Arabidopsis* orthologous gene pair were classified depending on (1) whether the TIS position/codon between orthologs is identical (i.e., "Shared") or different (i.e., "Distinct"); (2) whether the corresponding ORF type between orthologs is the same (i.e. "Same") or different (i.e. "Different") (see **Methods**). (B,C) Plots of LTM read density along the aligned CDS for an orthologous gene pair with alternative TISs (arrow) at the same position (B) and at different positions (C). The TIS codon and ORF type for each ortholog are shown next to the TIS peak and on the top left, respectively. The heatmap (top) shows the matched (pink) and mismatched (black) sequences and gapped regions (gray).

Supplemental Table S2. Summary of 5'-extended/truncated ORFs and their predicted localization.

	# alternative ORFs	Chloroplast		Mitochondrial	
		+	–	+	–
N-terminally extended ORFs	175	16	21	19	10
N-terminally truncated ORFs	668	101	93	35	88

+: Targeting signal in the N-terminally extended/truncated ORF but not the annotated ORF.

-: Targeting signal in the annotated ORF but not the N-terminally extended/truncated ORF.

Supplemental Table S3. RNA-seq mapping statistics for the LTM, CHX and mRNA samples

Read number\ Sample	LTM, rep1	LTM, rep2	CHX, rep1	CHX, rep2	Total RNAs, rep1	Total RNAs, rep2
Raw	95,608,270	115,710,376	89,588,181	106,066,244	106,903,465	123,125,844
QC-passed	86,973,843	103,785,394	80,620,935	97,516,630	96,427,418	115,136,413
Not mapped to tRNAs, rRNAs, snRNAs and snoRNAs	35,570,939	73,584,367	65,854,095	76,227,750	65,593,674	78,718,074
Mapped to genomes	30,115,112	65,203,599	57,591,995	63,932,617	44,658,947	55,843,819

Supplemental Table S4. List of primers used in the study.

Primer	5' to 3' sequence
Solyc07g052600_F	AAAGGAAAATAAACGAATAAAAAGTTAAAATGAGAATATGC
Solyc07g052600_R	GCAAATTTGGCTCCTCCACC
Solyc03g096920_F	CTCCTATTTTTCCTCTATATACTCTTTCTGCGTC
Solyc03g096920_R	AGCAGAGAAAGAAAAAGTGAGCG
Solyc02g064700_F	TCATTTGGAGAGAACACGGGGGACTATGGCCCTACGGCCTTCT
Solyc02g064700_R	ATGACGTCCTCGGAGGAGGCCATCTGGTGAGCGCGGGTTCTACT
Solyc07g008330-F	TCATTTGGAGAGAACACGGGGGACTTTTTGATAATTCGAGCCAAATATTTATATTTGCGA
solyc07g008330-R	ATGACGTCCTCGGAGGAGGCCATCTGCTCCTCATCTGAGTAGTTGAACTTGAC
Solyc02g023990-aTIS-F	ATGCATTTTTCAGTTGCATACGGAAAAG
Solyc02g023990_R	CCAAGGAATTTCTTGACCATCCCAG
Solyc02g023990-dTIS-F	ATGAAGCTGAATTTTGTTACTCAGCG
Solyc02g023990_dTISmutation_F	CCAAAAGCTTAAGCTGAATTTTGTTACTCAG
Solyc02g023990_dTISmutation_R	TTACTGAATCCCAAATCTGACACACACTC
Sloyc09g007540-F	GGAACTTCATCTAAACTTCATAACGGC
Solyc09g007540-aTIS-R	AGCCTTAGACCTTTTCTCATCCAT
Solyc09g007540-uTIS-R	ATTATCGGGTTTAGAAGAGAAAATACGACGTC
Solyc09g007540-uTISmutation-F	CTTCATATAGGCTTGCCTTAGGGTAC
Solyc09g007540-uTISmutation-R	TTTAGATGAAGTTCCAAGGGCGAATTC
Solyc03g044470-aTIS-F	GACAGTCCTCCTGCTATGGCTG
Solyc03g044470-R	GGAAGTTGATGACGACGAGCAAC
Solyc03g044470-dTIS-R	ACTGTTATAGATTTCCCTTAGTACTCCTAAAACCTG
Solyc03g044470-aTISmutation-F	TCCTGCTTAGGCTGTTCCACTC
Solyc03g044470-aTISmutation-R	GGACTGTCAAGGGCGAATTCG