Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis*

Yukio Kurihara[a,1], Yuko Makita[a,1], Mika Kawashima[a], Tomoya Fujita[b,c], Shintaro Iwasaki[b,d], and Minami Matsui[a,2]

[a]Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science, 230-0045 Yokohama, Japan; [b]RNA Systems Biochemistry Laboratory, RIKEN Cluster for Pioneering Research, 351-0198 Wako, Japan; [c]Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, 226-8503 Yokohama, Japan; and [d]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 351-0198 Wako, Japan

Edited by José M. Alonso, North Carolina State University, Raleigh, NC, and accepted by Editorial Board Member Joseph R. Ecker May 24, 2018 (received for review March 23, 2018)

Plants adapt to alterations in light conditions by controlling their gene expression profiles. Expression of light-inducible genes is transcriptionally induced by transcription factors such as HY5. However, few detailed analyses have been carried out on the control of transcription start sites (TSSs). Of the various wavelengths of light, it is blue light (BL) that regulates physiological responses such as hypocotyl elongation and flowering time. To understand how gene expression is controlled not only by transcript abundance but also by TSS selection, we examined genome-wide TSS profiles in *Arabidopsis* seedlings after exposure to BL irradiation following initial growth in the dark. Thousands of genes use multiple TSSs, and some transcripts have upstream ORFs (uORFs) that take precedence over the main ORF (mORF) encoding proteins. The uORFs often function as translation inhibitors of the mORF or as triggers of nonsense-mediated mRNA decay (NMD). Transcription from TSSs located downstream of the uORFs in 220 genes is enhanced by BL exposure. This type of regulation is found in *HY5* and *HYH*, major regulators of light-dependent gene expression. Translation efficiencies of the genes showing enhanced usage of these TSSs increased upon BL exposure. We also show that transcripts from TSSs upstream of uORFs in 45 of the 220 genes, including *HY5*, accumulated in a mutant of NMD. These results suggest that BL controls gene expression not only by enhancing transcriptions but also by choosing the TSS, and transcripts from downstream TSSs evade uORF-mediated inhibition to ensure high expression of light-regulated genes.

TSS | CAGE | uORF | ribosome profiling | NMD

Light is the most important environmental factor for plants not only for photosynthesis but also for photomorphogenesis (1). When etiolated seedlings are transferred to the light, they start deetiolation, and dynamic changes in their gene expression profiles occur at both the transcriptional and translational levels. Several transcription factors (TFs), such as HY5 and HYH, are involved in this process (2–7). HY5 perceives most signals from different wavelengths, but the *hy5* mutant is insensitive to blue, red, and far-red light and shows strong repression of signal transduction (5, 8). An *HY5* homolog, *HYH*, has also been reported for its role in photomorphogenesis (9).

In contrast to red/far-red light signals, information about the blue light (BL) signal is limited, although it is known that BL controls several important processes such as deetiolation, flowering time, and circadian rhythm through the BL photoreceptors, cryptochromes, and phototropins (1, 10). Expression of genes involved in photosynthesis is up-regulated by BL irradiation, and it is generally accepted that induction of these genes is controlled at the transcriptional level. It was recently reported that red light influences transcription start site (TSS) selection (11). However, there has been very little research on the control of TSSs upon BL exposure.

Cap analysis of gene expression (CAGE) was developed for precise and genome-wide identification of capped 5′ ends of mRNAs corresponding to TSSs of genes (12, 13). TSSs defined by CAGE analysis are classified in accordance with the shape of the CAGE tag cluster (CTC) into single, steep, and broad types (14–16). These CTC types approximately correlate with expression levels. For example, in maize, genes with broad CTCs are likely to show higher expression than genes with single or steep CTCs (14). Most genes in *Arabidopsis* and other organisms have multiple CTCs, including a major CTC and alternative CTCs (16–18).

More than 30% of genes in *Arabidopsis* contain small stretches of upstream ORFs (uORFs) in the 5′ untranslated regions (5′ UTRs) of the mRNAs (19, 20). It is known that these often function as translation repressors of the main ORFs (mORFs) located downstream of the uORFs (20–22). Not all but some uORFs are also reported to depress mRNA stability acting as triggers of nonsense-mediated mRNA decay (NMD) (23–26). NMD is caused by a protein complex, and one of its essential components is an RNA helicase, UP Frameshift 1 (UPF1) (27, 28).

Significance

The upstream ORFs (uORFs) in the 5′UTRs of mRNA often function as repressors of main ORF translation or triggers of nonsense-mediated mRNA decay. In this study, we report on transcription start site (TSS) selection when etiolated *Arabidopsis* seedlings are exposed to blue light, and reveal that transcription from uORF-avoiding TSSs is induced throughout the genome. It is possible that transcripts arising from TSSs downstream of uORFs evade uORF-mediated inhibition of gene expression. Thus, uORF-avoiding transcription starts are an important mechanism of gene expression regulation during a plant's response to environmental changes.

Author contributions: Y.K., Y.M., and M.M. designed research; Y.K. performed research; Y.K., Y.M., T.F., and S.I. contributed new reagents/analytic tools; Y.K., Y.M., M.K., T.F., and S.I. analyzed data; and Y.K., Y.M., T.F., S.I., and M.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.M.A. is a guest editor invited by the Editorial Board.

Data deposition: Sequence data from this article have been deposited in DDBJ/EMBL/Genbank BioProject, https://www.ncbi.nlm.nih.gov/bioproject (accession no. PRJDB5794) and Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE109122).

See Commentary on page 7654.

[1]Y.K. and Y.M. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: minami@riken.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1804971115/-/DCSupplemental.

Published online June 18, 2018.

After transcription, mRNAs are translated into proteins, and a recently developed technology called ribosome profiling enables monitoring of the actively translating ribosomes. Ribosome profiling is dependent on deep sequencing of ribosome-protected RNA fragments (so-called ribosome footprints) generated by RNase digestion (29–34). This methodology allows identification of which genes within the genome are translated and monitoring of changes in the translation efficiency (TE) of each gene in vivo.

To understand the nature of TSSs during BL exposure in *Arabidopsis* seedlings, we captured genome-wide changes in TSSs using CAGE and also measured translational control by ribosome profiling. We observed that BL causes enhanced usage of TSSs downstream of uORFs and increases the TEs of the corresponding genes. In addition, transcripts from TSSs upstream of uORFs in only a small fraction of genes are stabilized in the *UPF1* mutation. These observations suggest that BL enables evasion of uORF-mediated inhibition of gene expression by influencing the selection of a TSS.

Results

CAGE Analysis Revealed Diversity of TSSs After BL Exposure. To understand the diversity of TSSs within the whole genome after BL exposure, we performed CAGE analysis on 3-d-old wild-type (WT) and *hy5 Arabidopsis* seedlings grown in darkness following 1 h of BL exposure. We chose this time point as *HY5* mRNA increases and reaches a maximum level after 1 h of exposure (*SI Appendix*, Fig. S1*A*). In CAGE analysis, the level of accumulation is represented as tag per million reads (TPM). We detected a total of 48,207 CTCs (TPM > 0.5), which corresponds to 17,627 annotated genes (TAIR10) in the four treatments (WT_Dark, WT_Blue, *hy5*_Dark, and *hy5*_Blue), and 24,270 CTCs and 15,152 genes overlapped in all four (Fig. 1*A* and *SI Appendix*, Fig. S1*B*).

Positional analysis of CTCs along with the corresponding genes indicates that more than 70% of the CTCs mapped "Upstream" and "5′UTR," which are defined in the TAIR10 gene model (Fig. 1*B*). We did not obtain any significant differences in the distribution of CTC positions between all four conditions. We also observed that thousands of CTCs mapped in coding sequences (CDSs), exons of noncoding RNAs (ncRNAs), and 3′UTRs, indicating the possibility of alternative TSSs in these regions.

CAGE analysis gives information about not only the locations of TSSs but also the amount of transcript arising from a TSS. To compare accumulation levels from CAGE analysis with those of RNA sequencing (RNA-Seq), we performed RNA-Seq using the same RNA samples as used for the CAGE analysis. TPM values correlated well with fragments per kilobase of exon per million reads (FPKMs) of RNA-Seq analysis with R^2 scores of 0.91, 0.92, 0.90, and 0.90 in WT_Dark, WT_Blue, *hy5*_Dark, and *hy5*_Blue, respectively (Fig. 1*C*). This result indicates that our CAGE analysis evaluates TSSs quantitatively in the four treatments.

We noticed that the widths of the CTCs were different between genes. We detected a correlation between these widths and the accumulation of CAGE tags in all conditions (Fig. 1*D*). Based on their widths, we classified the clusters into three types: 1 nt as "Single," 2 nt to 10 nt as "Steep," and more than 10 nt as "Broad" (Fig. 1*E*, *Left*). There were few differences in the distribution of the three CTC types between the four conditions (Fig. 1*E*, *Right*). We found that Broad was more enriched in the 5′UTR of genes than the other types, and Single was more enriched in CDSs (*SI Appendix*, Fig. S1*C*). This tendency may be dependent on core promoter architecture, for example, if regulatory sequences such as TATA boxes are present (16). Indeed, promoters around 5′UTRs often possess such regulatory sequences.

We examined 602 up-regulated and 274 down-regulated genes after exposure to BL in WT (RNA-Seq, q value < 0.05, fold change >3 or <1/3). Accumulation of 381 of the up-regulated genes and 195 of the down-regulated genes was not significantly changed in the *hy5* mutant. In both, the proportion of Broad CTCs mainly contributed to these changes (Fig. 1*F*). This observation correlates well with high expression of Broad CTCs (Fig. 1*D*).

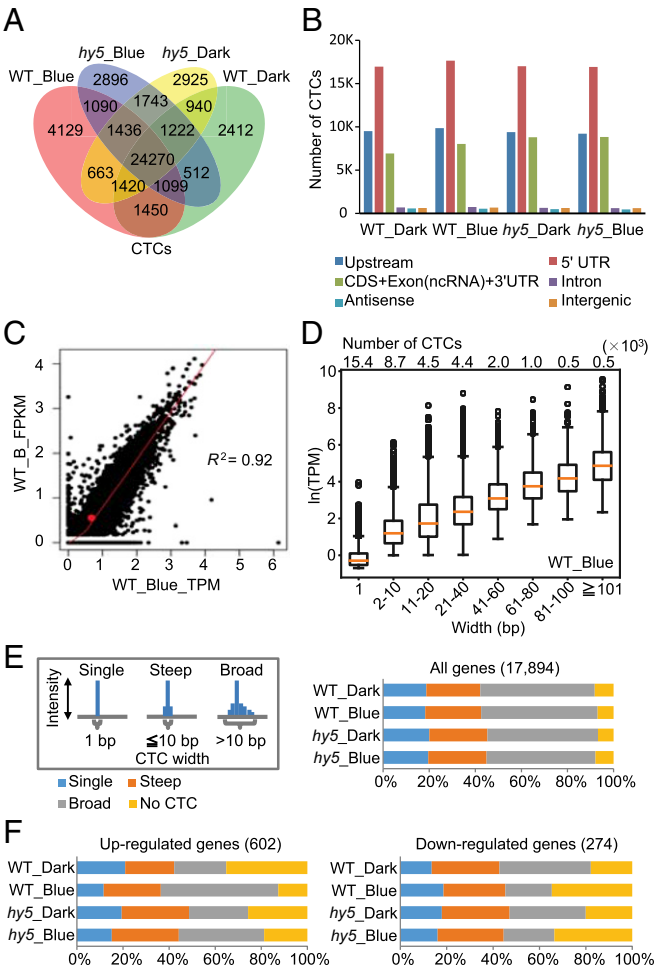

Fig. 1. CAGE analysis reveals genome-wide TSSs and characterizes them. (*A*) Venn diagram showing the overlap of CTCs between WT_Dark, WT_Blue, *hy5*_Dark, and *hy5*_Blue. (*B*) Distribution of positions of CTCs. (*C*) Dot plot showing the correlation at gene level between TPM (CAGE) and FPKM (RNA-Seq). Logarithmic values (base is 10) are used. Data of WT_Blue is shown as representative. (*D*) Box plot showing the relationship between the width and accumulation (TPM) of CTCs. Numbers of CTCs used are shown at the top of each plot. The box plot of WT_Blue is shown as representative data. (*E* and *F*) Distribution of the three types (Single, Steep, and Broad) of CTCs (*E*, *Left*) in groups of (*E*, *Right*) all and (*F*) up-regulated and down-regulated genes (q value < 0.05 and fold change >3 or <1/3) in response to BL. CTCs were classified based on their width into single (1 nt), steep (≤10 nt), and broad types (>10 nt).

Control of Gene Expression by Multiple TSSs. In a CTC, there is a peak (Fig. 2*A*, asterisk). We noted the position of this peak as a point of reference for changes caused by BL exposure. Most of the BL up-regulated genes have the same main peak before and after BL exposure. We noticed that, in 93 of the 602 genes, the position of the main peak shifted downstream or upstream of the original position (Fig. 2*A*). The peak position also changed in 19 of the 274 down-regulated genes (*SI Appendix*, Fig. S2*A*).

Apart from these small fluctuations in the main peak position, we also observed more dynamic changes in TSSs caused by BL exposure. As exemplified in Fig. 2*B*, *HPR1*, which encodes hydroxypyruvate reductase that regulates photorespiratory flux

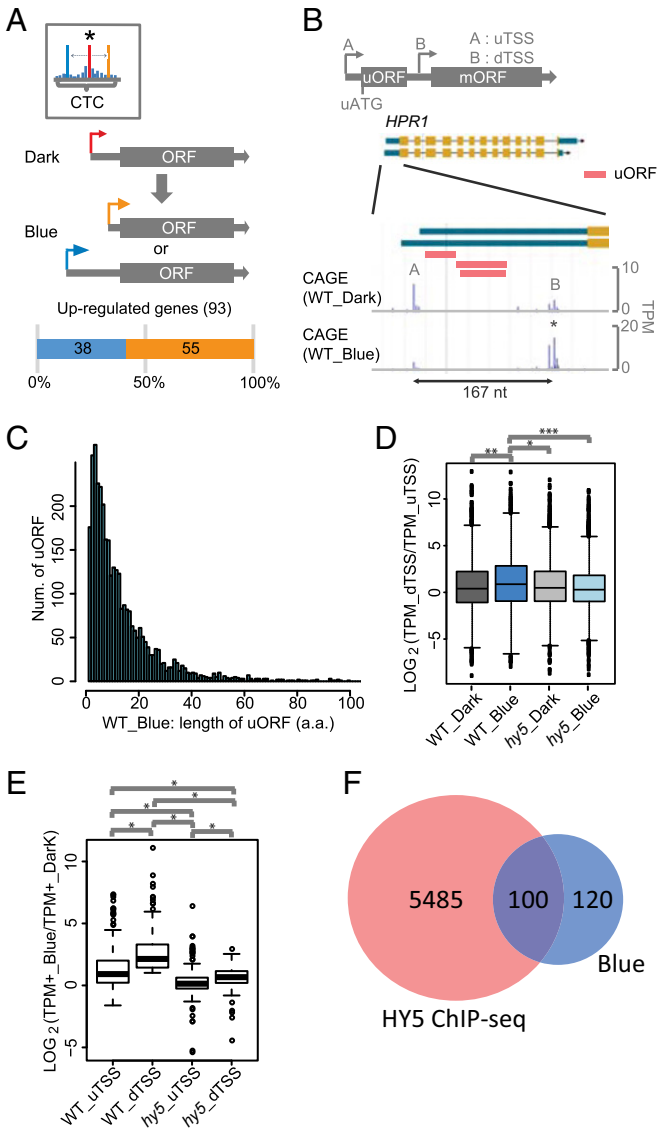

Fig. 2. The uORF-avoiding transcription occurs throughout genome. (*A*) Positional shifts of peaks of CTCs in 93 up-regulated genes upon BL exposure. The asterisk in the upper illustration indicates an example of a peak in a CTC. (*B*) An example (*HPR1*) of a gene that undergoes uORF-avoiding transcription upon BL exposure. Position "A" indicates a uTSS located upstream of an upstream ATG (uATG) codon, while position "B" indicates a dTSS located downstream of a uATG codon as illustrated above. The asterisk indicates significant difference (Welch's *t* test, *: *P* value < 0.01). (*C*) Distribution of lengths of uORFs located between uTSSs and dTSSs. Lengths of WT_Blue are shown as representative data. (*D*) Box plot showing logarithmic values of fold differences (TPM_dTSS/TPM_uTSS) of 2,308, 2,779, 2,311, and 2,384 genes that use both uTSSs and dTSSs in WT_Dark, WT_Blue, *hy5*_Dark, and *hy5*_Blue, respectively. Asterisks indicate significant differences (Tukey's test; $*P < 10^{-2}$; $**P < 10^{-3}$; $***P < 10^{-6}$). (*E*) Box plot showing logarithmic values of fold changes (TPM+_Blue/TPM+_Dark) of uTSSs and dTSSs of the 220 genes that showed enhanced uORF-avoiding transcription from dTSSs upon BL exposure. Asterisks indicate significant differences (Tukey's test; $*P < 0.05$). To avoid a denominator of zero when calculating fold change, additional values (10^{-2}) were added to the TPM (TPM+). (*F*) Venn diagram of the overlap between the 5,585 HY5 target genes identified by ChIP-seq (red) and the 220 genes showing enhanced uORF-avoiding transcription upon BL (blue).

(35), has three uORFs in its 5′UTR. It has two TSSs before and after the uORFs (Fig. 2*B*, positions A and B). The expression level from each TSS changes according to the light conditions. In the dark, the TSS located upstream of the uORFs (position A in Fig. 2*B*) is mainly used, but, after BL exposure, transcription from this TSS is decreased and transcription from another TSS located downstream of the uORFs increases to ensure BL induction of the *HPR1* gene. It is reported that a uORF in the 5′ UTR sometimes interferes with translation of a downstream mORF. This alternative shift of TSS position from upstream to downstream of a uORF may function to relieve negative control by uORFs.

To understand this kind of regulation, we examined CAGE data of genes that have uORFs in their 5′UTR and also have both upstream (uTSS) and downstream (dTSS) TSSs. More precisely, we defined a uTSS as a TSS located upstream of the first ATG methionine codon of a uORF, and we define dTSS as a TSS located downstream of this first ATG (not necessarily downstream of all uORFs).

Using these criteria, we estimated that there are 2,308, 2,779, 2,311, and 2,384 genes from the WT_Dark, WT_Blue, *hy5*_Dark, and *hy5*_Blue treatments, respectively (Dataset S1). Nearly half of these genes possess multiple uORFs. Most uORF lengths are shorter than 40 amino acids, consistent with a previous report (Fig. 2*C*) (19).

We compared the ratios of TPMs of dTSSs to uTSSs between the four conditions. After BL exposure, this ratio increased in WT, indicating transcription from dTSSs is enhanced (Fig. 2*D*). However, such an increase was not observed in the *hy5* mutant. Of the 602 BL-up-regulated genes, 39, 114, 46, and 79 genes have both uTSSs and dTSSs in WT_Dark, WT_Blue, *hy5*_Dark, and *hy5*_Blue, respectively. Transcriptional increase from their dTSSs by BL was more enhanced than the result shown in Fig. 2*D* (*SI Appendix*, Fig. S2*B*).

To understand light-controlled TSS selection, we looked for genes that significantly undergo BL-induced dTSS transcription in WT and defined 220 genes [TPM_dTSS (WT_Blue) vs. TPM_dTSS (WT_Dark), TPM (WT_Blue) > 1, Welch's *t* test *P* value < 0.05, fold change (WT_Blue/WT_Dark) > 2] (Dataset S2 and Fig. 2*E*, WT_uTSS vs. WT_dTSS). Gene ontology analysis for these 220 genes showed that light response-related terms are highly enriched (Dataset S2). Note that all transcripts from multiple TSSs in a gene contribute to the total accumulation and that the genes that showed increased transcription from dTSSs after BL exposure are not necessarily BL-up-regulated genes identified by RNA-Seq. In fact, only 89 of the 220 genes overlap with the 602 BL-up-regulated genes.

The increase in transcription from dTSSs in the 220 genes was strongly impaired in *hy5* mutants, although weak induction still occurred in *hy5*, indicating HY5 is partly responsible for induction from dTSSs (Fig. 2*E*, WT_dTSS vs. *hy5*_dTSS). For example, *HRT1* has the binding site of HY5 upstream of two TSSs; this was identified by chromatin-immunoprecipitation sequencing (ChIP-seq) analysis and an in vitro binding assay (*SI Appendix*, Fig. S2*C* and Dataset S2) (36). In addition, ChIP-seq analysis of HY5 showed that 100 genes out of the 220 genes have HY5-binding sites in their promoters, suggesting the possibility that transcription from their dTSSs is directly regulated by HY5 (Fig. 2*F* and Dataset S2).

More precisely, 58 of the 220 genes showed enhanced transcription from dTSSs without induction from uTSSs [TPM_uTSS (WT_Dark) vs. TPM_uTSS (WT_Blue), fold change (WT_Blue/WT_Dark) <1.2] (*SI Appendix*, Fig. S2*D*). Of these genes, ChIP-seq analysis identified 22 that have HY5-binding sites, implying that transcription from dTSSs but not uTSSs is directly regulated by HY5.

uORFs Regulate Downstream Translation. We examined translational regulation by uORFs that are avoided by dTSS usage upon BL exposure using an in vitro wheat germ translation system. We used five light signal-related genes whose transcription from dTSSs is induced by BL exposure [*HY5*, *HYH*, *HPR1*, *High*

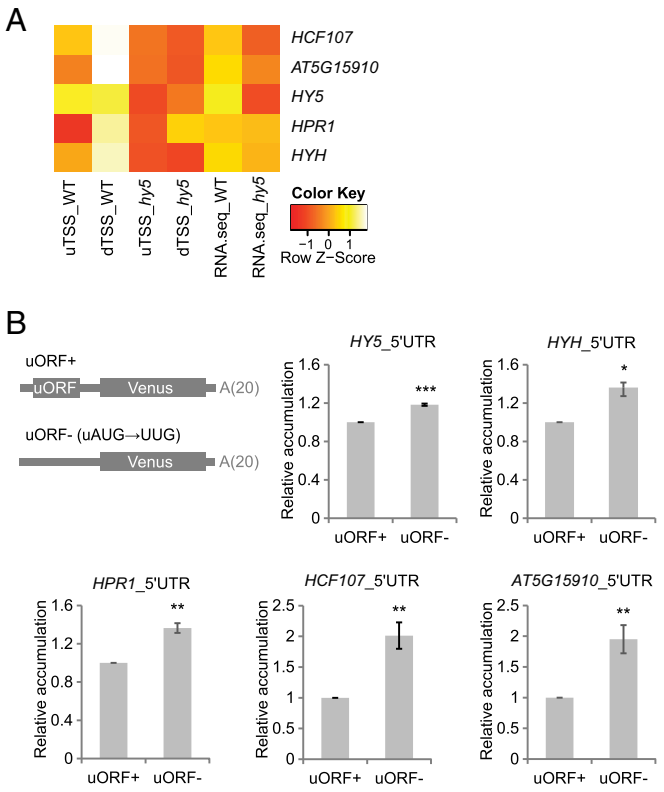

Fig. 3. The uORFs regulate downstream mORF translation. (*A*) Heat map showing logarithmic values of fold changes (TPM+_Blue/TPM+_Dark and FPKM+_Blue/FPKM+_Dark) of uTSSs and dTSSs, and RNA-Seq of the five genes whose 5′UTRs were selected for in vitro protein synthesis assays. To avoid a denominator of zero when calculating fold change, additional values (10^{-2}, 10^{-4}) were added to the TPM and FPKM (TPM+ and FPKM+), respectively. (*B*) In vitro protein (*Venus*) synthesis assays using wheat germ extract. *Venus* mRNAs with or without uORFs in the 5′UTR were transcribed in vitro and used for translation. In uORF− mRNAs, uAUG was replaced with UUG. Western blot analysis was performed to quantify synthesized Venus protein. Data were averaged and normalized to the values from uORF+ mRNAs. Asterisks indicate significant differences ($n = 3$, Student's t test; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$). Error bars denote SD.

Chlorophyll Fluorescent 107 (*HCF107*), and *AT5G15910*] (Fig. 3*A* and *SI Appendix*, Fig. S3). This induction is impaired by the *hy5* mutation, and there are HY5-binding sites in the promoters of *HYH*, *HPR1*, *HCF107*, and *AT5G15910* (Dataset S2). The 5′UTRs of these genes were fused to the *Venus* reporter gene. For each construct, we introduced a point mutation in the ATG of the uORF, changing it to TTG to remove the uORFs (uORF−) (Fig. 3*B*). In all cases, the introduction of this point mutation increased accumulation of the Venus protein, indicating an inhibitory effect of these uORFs.

Transcription from dTSS After BL Exposure Enhances Translational Efficiency. The most straightforward outcome of the emergence of a uORF along the transcript is translational inhibition. To directly measure the impact on translation after BL exposure, we performed ribosome profiling. Our ribosome footprints showed a three-nucleotide phase along the ORF, which represents ribosome movement at every codon, and is a hallmark of typical ribosome profiling data (*SI Appendix*, Fig. S4*A*). We noted that our data cover two distinct sizes of footprint: a major one, 27 nt to 29 nt, and a minor one, 20 nt to 21 nt (*SI Appendix*, Fig. S4*B*). In a previous study with *Saccharomyces cerevisiae* (37, 38), two similar peaks of footprint size were also monitored and suggested to be generated by differential nuclease accessibility by distinct ribosome conformations.

Ribosome profiling showed a reverse correlation in translation between the uORF and the downstream mORF upon BL exposure. Consistent with the loss of the uORF by increased dTSS usage, the TEs (overrepresentation or underrepresentation of ribosome footprints over conventional RNA-Seq) of most of the genes showing uORF-avoiding transcription from dTSSs were increased by BL exposure (Fig. 4*A*, black points and Fig. 4*B*). Concomitantly, translation from their uORFs was down-regulated (Fig. 4*C*).

Part of Inhibitory Effect of uORF Is Controlled by UPF1. It is known that uORFs located in the 5′UTR of mRNAs induce NMD (22–25). We speculated that mRNAs arising from uTSSs might be negatively controlled by NMD to prevent unnecessary translation in dark conditions. We tested this possibility using the *upf1-1* mutant that shows impaired NMD (39).

Eleven-day-old light-grown seedlings of WT and the *upf1-1* mutant were used for CAGE analysis. The analysis identified 876 transcriptions from uTSSs that were significantly enhanced in *upf1-1* compared with WT [TPM_uTSS (*upf1-1*) vs. TPM_uTSS (WT), TPM_uTSS > 0.5, Student's t test P value < 0.05, fold change (*upf1-1*/WT) >1.5] (Fig. 5*A* and Dataset S3). Of these, 357 genes showed enhanced transcription from uTSSs

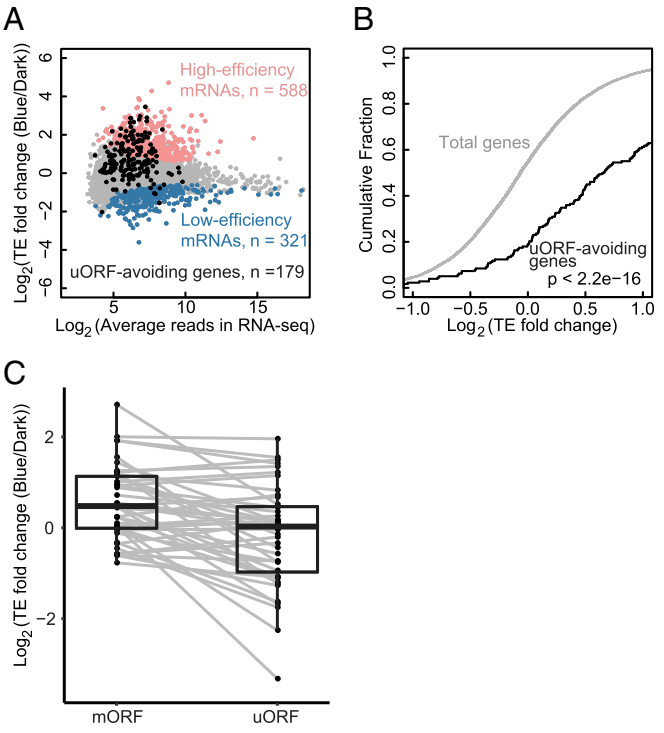

Fig. 4. Ribosome profiling reveals that BL-enhanced transcription from dTSSs evades uORF-mediated translation repression. (*A*) Log ratios and mean average (MA) plot of mean reads in RNA-Seq versus TE fold change of mORF after BL exposure (TE_Blue/TE_Dark). TEs were defined by overrepresentation or underrepresentation of footprint counts (sequenced reads ≥ 2) in ribosome profiling over RNA accumulation in RNA-Seq. Red, blue, black, and gray dots indicate significantly up-regulated, significantly down-regulated, uORF-avoiding, and other genes, respectively. Of the 220 uORF-avoiding genes, 179 genes were detected in this analysis. Significant differences were determined by negative binominal tests with DESeq (50) (q < 0.01). (*B*) Cumulative curve of TE fold change (TE_Blue/TE_Dark) of BL-enhanced uORF-avoiding genes (black) and all genes (gray). Significant differences were determined by the Mann–Whitney u test. (*C*) Box plot showing TE fold change of the 44 uORFs detected and the corresponding mORFs of BL-enhanced uORF-avoiding genes.

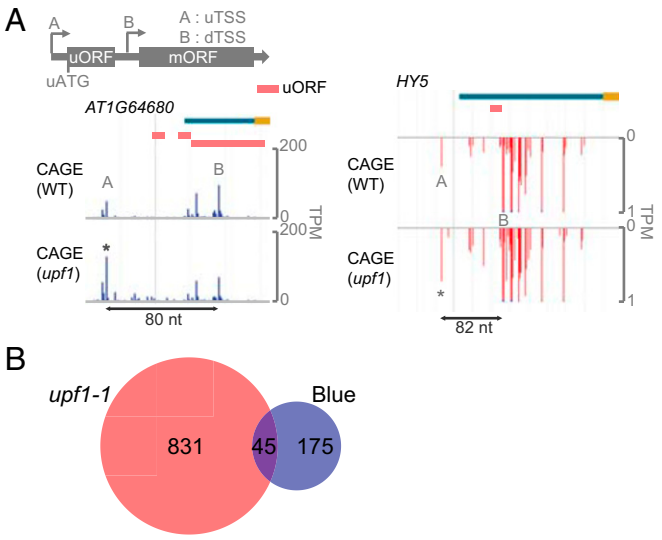

Fig. 5. Transcripts derived from uTSSs accumulate in *upf1-1*. (*A*) Two examples of genes (*AT1G64680* and *HY5*) that show enhanced accumulation of transcripts from uTSSs in *upf1-1*. Position "A" indicates uTSSs located upstream of uATG codons, while position "B" indicates dTSSs located downstream of uATG codons. Heights of dTSSs and their peripheries of the *HY5* loci are saturated in this diagram. Asterisks indicate significant differences (Welch's *t* test, *: *P* value < 0.01). Blue (*AT1G64680*) and red (*HY5*) CAGE signals are from positive and negative strands, respectively. (*B*) Venn diagram of the overlap between the 876 genes showing enhanced accumulation of transcripts from uTSSs in *upf1-1* (red) and the 220 genes showing enhanced uORF-avoiding transcription upon BL (blue).

but not from dTSSs [TPM_dTSS (*upf1-1*) vs. TPM_dTSS (WT), fold change (*upf1-1*/WT) < 1.2]. We compared these genes with those that underwent enhancement of transcription from dTSSs after BL exposure (Fig. 5*B*). Only 45 of 813 genes overlapped with the 220 genes showing BL-enhanced transcription from dTSSs. *HY5* and *AT1G64680* were among these 45 genes (Fig. 5*A*). The transcript from the uTSS of the *HPR1* gene is also likely to accumulate in *upf1-1* (*SI Appendix*, Fig. S5). These results suggest that at least a small subset of transcripts from uTSSs of the uORF-avoiding genes induced by BL is negatively regulated by NMD.

Discussion

In general, gene expression is regulated at the transcriptional and posttranscriptional levels to allow adaptation to environmental changes. Gene expression is controlled by several regulatory mechanisms such as tissue-specific transcription activation, alternative splicing, and mRNA editing (7, 40–43). The mRNAs possess the information not only for protein coding but also for the regulation of stability and translation efficiency (44). The uORFs located in the 5′UTR of mRNA are such regulatory elements and modulate translation efficiency (20–22).

Plants perceive light not only for photosynthesis but also for photomorphogenesis. Light irradiation causes many physiological changes in plants, especially at the seedling stage. BL mainly controls hypocotyl elongation and flowering time (1, 10). We examined gene expression in etiolated *Arabidopsis* seedlings after irradiation with BL. As there has been little research on precise TSSs, we performed CAGE analysis to investigate the response to BL exposure. This analysis identified a subset of genes in which the main TSS undergoes a shift from upstream to downstream of the uORF in response to BL (Fig. 2). This shift allows evasion of uORF-mediated inhibition of gene expression, such as translational inhibition in light-responsive genes (*SI Appendix*, Fig. S6).

Genes undergoing uORF-avoiding transcription include the light-inducible genes, *HY5*, *HYH*, and *HPR1* (Figs. 2 and 3*A*). HY5 and HYH are bZIP-type TFs that are important for light signaling and photomorphogenesis through inhibition of hypocotyl elongation (7), while *HPR1* encodes hydroxypyruvate reductase that is involved in photorespiration (35). We hypothesize that expression of these genes is fine-tuned by employing a uORF-avoiding TSS to ensure protein translation. HY5 has two TSSs upstream and downstream of the uORF. BL exposure enhanced transcription from the dTSS, and this induction is cancelled by the *hy5* mutation, indicating autoregulation of its transcription. Transcription from the *HY5* uTSS accumulated in the *upf1* mutation, UPF1 being a major component of NMD (Fig. 5*A*). This result suggests that *HY5* expression is very tightly controlled at the mRNA level through placement of the uORF, and this uORF may be targeted by NMD for removal of residual mRNA in the dark. Indeed, HY5 is also regulated at the protein level via degradation by the COP1/SPA E3 ubiquitin ligase complex in darkness (5, 8, 45). Thus, multiple layers of regulatory mechanisms ensure tight control of light-regulated genes to prevent unwanted expression of photosynthetic genes in the dark and allow an immediate increase in their expression upon exposure to BL.

We have demonstrated that uORFs placed before the *Venus* reporter gene caused inhibition of translation in vitro (Fig. 3). We have also shown, using ribosome profiling, that transcription downstream of a uORF increases translational efficiency. It has recently been reported that the R motif, which comprises mainly purines, influences translation efficiency during pathogen recognition (34). In addition, it is speculated that non-AUG uORFs, which initiate with non-AUG start codons, may regulate gene expression as AUG uORFs do (46). It will be intriguing to investigate the positional relationship between TSSs and regulatory sequences such as the R motif and non-AUG uORFs.

Another finding is that BL-enhanced transcription from dTSSs is not accompanied by increased transcription from uTSSs in a group of genes (*SI Appendix*, Fig. S2*D*). This result indicates different transcriptional regulation mechanisms between uTSSs and dTSSs. In the case of the *HPR1* gene, whose promoter possesses a G-box sequence for HY5 binding, only transcription from the dTSS and not the uTSS is dependent on HY5 (Fig. 2*B* and *SI Appendix*, Fig. S2*C*). It is known that other TFs like HYH and GBF1 also bind to the G-box sequence and regulate the light response (47, 48). They might affect the selection of TSSs as does HY5. In this context, it is possible that TF binding is the main determinant of transcriptional induction from alternative TSSs. Alternatively, other factors may be involved in the selection of TSSs; DNA methylation has been reported to be one of these, suggesting epigenetic regulation is also important (49).

Previous ribosome profiling work in *Arabidopsis* showed that light exposure regulates translation of mORFs through uORFs but did not demonstrate how the uORFs exert their influence (33). Here, we suggest that the TSS shifts from an uTSS to a dTSS to skip sequences of the uORF, and mRNAs transcribed from dTSSs evade uORF-mediated translational inhibition and/or mRNA decay.

We also found that about 20% of TSSs are located in the CDS and 3′UTR (Fig. 1*B*). It is reported that these TSSs can produce isoforms of proteins and, in particular, can generate ones with or without intracellular localization signals at their *N* termini (11, 14, 49). Recently, it has also been reported that red light exposure produces such modified proteins that have altered intracellular localization (11). Thus, not only does uORF-avoiding transcription from dTSSs serve as a regulator of gene expression but so do other mechanisms of selecting the position of the TSS.

Materials and Methods

Details of plant growth conditions, CAGE, ribosome profiling, RNA-Seq analyses, experimental procedures for the construction of templates for in vitro transcription, in vitro protein synthesis assays and Western blot analysis are described in *SI Appendix, SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Prof. Ferenc Nagy for providing seeds of *HY5pro:HY5-YFP* transgenic plants and Prof. Yuichiro Watanabe for his comments on this paper. Computations were partly supported by Manabu Ishii, Itoshi Nikaido, and the Bioinformatics Analysis Environment Service on RIKEN Cloud at RIKEN Advance Center. This study was supported by Grant-in Aid JP25113006 (to M.M.) from the Ministry of Education, Culture, Sports, Science and Technology in Japan; in part by Grant-in-Aid JP17H05679 for Scientific Research on Innovative Areas "Nascent Chain Biology"; Grant-in-Aid JP17H04998 (to S.I.) for Young Scientists from the Japan Society for the Promotion of Science; the Pioneering Projects "Cellular Evolution" (S.I.); All RIKEN Multidiscipline Research Program "Disease and Epigenome" (S.I.); the Aging Project from RIKEN (S.I.), and Takeda Science Foundation (S.I.).

1. Wu SH (2014) Gene expression regulation in photomorphogenesis from the perspective of the central dogma. *Annu Rev Plant Biol* 65:311–333.
2. Lee J, et al. (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 19:731–749.
3. Liu H, et al. (2008) Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. *Science* 322:1535–1539.
4. Singh A, Ram H, Abbas N, Chattopadhyay S (2012) Molecular interactions of GBF1 with HY5 and HYH proteins during light-mediated seedling development in *Arabidopsis thaliana*. *J Biol Chem* 287:25995–26009.
5. Gangappa SN, Botto JF (2016) The multifaceted roles of HY5 in plant growth and development. *Mol Plant* 9:1353–1365.
6. Pedmale UV, et al. (2016) Cryptochromes interact directly with PIFs to control plant growth in limiting blue light. *Cell* 164:233–245.
7. Oyama T, Shimura Y, Okada K (1997) The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. *Genes Dev* 11:2983–2995.
8. Osterlund MT, Hardtke CS, Wei N, Deng XW (2000) Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. *Nature* 405:462–466.
9. Holm M, Ma LG, Qu LJ, Deng XW (2002) Two interacting bZIP proteins are direct targets of COP1-mediated control of light-dependent gene expression in *Arabidopsis*. *Genes Dev* 16:1247–1259.
10. Lin C (2002) Blue light receptors and signal transduction. *Plant Cell* 14:S207–S225.
11. Ushijima T, et al. (2017) Light controls protein localization through phytochrome-mediated alternative promoter selection. *Cell* 171:1316–1325.e12.
12. Shiraki T, et al. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA* 100:15776–15781.
13. Kodzius R, et al. (2006) CAGE: Cap analysis of gene expression. *Nat Methods* 3: 211–222.
14. Mejía-Guerra MK, et al. (2015) Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. *Plant Cell* 27:3309–3320.
15. Morton T, et al. (2014) Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures. *Plant Cell* 26:2746–2760.
16. Yamamoto YY, et al. (2009) Heterogeneity of *Arabidopsis* core promoters revealed by high-density TSS analysis. *Plant J* 60:350–362.
17. Nepal C, et al. (2013) Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Res* 23: 1938–1950.
18. Miura F, et al. (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci USA* 103:17846–17851.
19. Kim BH, Cai X, Vaughn JN, von Arnim AG (2007) On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome Biol* 8: R60.
20. von Arnim AG, Jia Q, Vaughn JN (2014) Regulation of plant translation by upstream open reading frames. *Plant Sci* 214:1–12.
21. 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 USA* 106:7507–7512.
22. Ebina I, et al. (2015) Identification of novel *Arabidopsis thaliana* upstream open reading frames that control expression of the main coding sequences in a peptide sequence-dependent manner. *Nucleic Acids Res* 43:1562–1576.
23. Kurihara Y, et al. (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. *Proc Natl Acad Sci USA* 106:2453–2458.
24. Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, Silhavy D (2009) Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol* 71:367–378.
25. Uchiyama-Kadokura N, et al. (2014) Polyamine-responsive ribosomal arrest at the stop codon of an upstream open reading frame of the AdoMetDC1 gene triggers nonsense-mediated mRNA decay in *Arabidopsis thaliana*. *Plant Cell Physiol* 55:1556–1567.
26. Malabat C, Feuerbach F, Ma L, Saveanu C, Jacquier A (2015) Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* 4:06722.
27. Kervestin S, Jacobson A (2012) NMD: A multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* 13:700–712.
28. Lykke-Andersen S, Jensen TH (2015) Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* 16:665–677.
29. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324:218–223.
30. Iwasaki S, Ingolia NT (2017) The growing toolbox for protein synthesis studies. *Trends Biochem Sci* 42:612–624.
31. Bazin J, et al. (2017) Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proc Natl Acad Sci USA* 114:E10018–E10027.
32. Hsu PY, et al. (2016) Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proc Natl Acad Sci USA* 113:E7126–E7135.
33. Liu MJ, et al. (2013) Translational landscape of photomorphogenic Arabidopsis. *Plant Cell* 25:3699–3710.
34. Xu G, et al. (2017) Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature* 545:487–490.
35. Timm S, et al. (2013) Serine acts as a metabolic signal for the transcriptional control of photorespiration-related genes in *Arabidopsis*. *Plant Physiol* 162:379–389.
36. Kurihara Y, et al. (2014) Next-generation sequencing of genomic DNA fragments bound to a transcription factor *in vitro* reveals its regulatory potential. *Genes (Basel)* 5:1115–1131.
37. Lareau LF, Hite DH, Hogan GJ, Brown PO (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* 3:e01257.
38. Matsuo Y, et al. (2017) Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat Commun* 8:159.
39. Yoine M, Ohto MA, Onai K, Mita S, Nakamura K (2006) The lba1 mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in *Arabidopsis*. *Plant J* 47:49–62.
40. Drechsel G, et al. (2013) Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the *Arabidopsis* steady state transcriptome. *Plant Cell* 25:3726–3742.
41. Hartmann L, et al. (2016) Alternative splicing substantially diversifies the transcriptome during early photomorphogenesis and correlates with the energy availability in *Arabidopsis*. *Plant Cell* 28:2715–2734.
42. Shikata H, et al. (2014) Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proc Natl Acad Sci USA* 111:18781–18786.
43. Song L, et al. (2016) A transcription factor hierarchy defines an environmental stress response network. *Science* 354:aag1550.
44. Brar GA (2016) Beyond the triplet code: Context cues transform translation. *Cell* 167: 1681–1692.
45. Saijo Y, et al. (2003) The COP1-SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. *Genes Dev* 17:2642–2647.
46. Spealman P, et al. (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res* 28:214–222.
47. Ram H, Priya P, Jain M, Chattopadhyay S (2014) Genome-wide DNA binding of GBF1 is modulated by its heterodimerizing protein partners, HY5 and HYH. *Mol Plant* 7: 448–451.
48. Binkert M, et al. (2014) UV-B-responsive association of the *Arabidopsis* bZIP transcription factor ELONGATED HYPOCOTYL5 with target genes, including its own promoter. *Plant Cell* 26:4200–4213.
49. Pumplin N, et al. (2016) DNA methylation influences the expression of DICER-LIKE4 isoforms, which encode proteins of alternative localization and function. *Plant Cell* 28:2786–2804.
50. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106.