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 d, 1, and Minami Matsui 1, 2

1Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science, 230-0045 Yokohama, Japan; 2RNA Systems Biochemistry Laboratory, RIKEN Cluster for Pioneering Research, 351-0198 Wako, Japan; 3Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, 226-8503 Yokohama, Japan; and 4Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 351-0198 Wako, Japan

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 genomewide 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 transcription but also by choosing the TSS, and transcripts from downstream TSSs evade uORF-mediated inhibition to ensure high expression of light-regulated genes.

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, UPF1 (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. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

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.

1Y.K. and Y.M. contributed equally to this work.
2To 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.
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 TE 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. S1A). 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. 1A and SI Appendix, Fig. S1B).

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. 1B). 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. 1C). 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. 1D). 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. 1E, Left). There were few differences in the distribution of the three CTC types between the four conditions (Fig. 1E, 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. S1C). 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. 1F). This observation correlates well with high expression of Broad CTCs (Fig. 1D).

Control of Gene Expression by Multiple TSSs. In a CTC, there is a peak (Fig. 2A, 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. 2A). The peak position also changed in 19 of the 274 down-regulated genes (SI Appendix, Fig. S2A).

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. 2B, HPR1, which encodes hydroxypyruvate reductase that regulates photorespiratory flux
(35), has three uORFs in its 5' UTR. It has two TSSs before and after the uORFs (Fig. 2B, 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. 2B) 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 HPRI 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. 2C) (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. 2D). However, such an increase was not observed in the hy5 mutant. Of the 602 BL-up-regulated genes, 39, 46, and 79 genes have both uTSS and dTSS 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. 2D (SI Appendix, Fig. S2B).

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 vs. TPM_uTSS; (WT_Dark), (WT_Blue)] that transcription from their dTSSs is directly regulated by HY5 (Fig. 2E, 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. 2E, WT_uTSS vs. hy5_uTSS). 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. S2C 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. 2F 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. S2D). 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, HPRI, High
Ribosome profiling reveals that BL-enhanced transcription from dTSSs
Kurihara et al.

... to be generated by differential nucleosome 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. 4A, black points and Fig. 4B). Concomitantly, translation from their uORFs was down-regulated (Fig. 4C).

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 < 0.05, fold change (upf1-1/WT) > 1.5 (Fig. 5A and Dataset S3). Of these, 557 genes showed enhanced transcription from uTSSs.

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. S4A). 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. S4B). In a previous study with Saccharomyces cerevisiae (37, 38), two similar peaks of footprint size were also monitored and suggested...
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. 5B). 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. 5A). The transcript from the uTSS of the HPRI 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 HPRI (Figs. 2 and 3A). HY5 and HYH are bZIP-type TFs that are important for light signaling and photomorphogenesis through inhibition of hypocotyl elongation (7), while HPRI encodes hydroxypyruvate reductase that is involved in photosynthesis (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 enhances 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. 5A). 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. S2D). This result indicates different transcriptional regulation mechanisms between uTSSs and dTSSs. In the case of the HPRI 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. 2B and SI Appendix, Fig. S2C). 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. 1B). 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 HY5; Prof. Shinya Yamanaka for transgenic plants and Prof. Yuichiro Watanabe for his comments on this paper. Computations were partly supported by Manabu Ishii, Itsushi Nikiado, 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 JP17H04989 (to S.I.) for Young Scientists from the Japan Society for the Promotion of Science; the Pioneering Projects “Cellular Evolution” (S.I.); All RIKEN Multidisciplinary Research Program “Disease and Epigenome” (S.I.); the Aging Project from RIKEN (S.I.), and Takeda Science Foundation (S.I.).