

Evening Expression of *Arabidopsis GIGANTEA* Is Controlled by Combinatorial Interactions among Evolutionarily Conserved Regulatory Motifs ©W|OPEN

Markus C. Berns, Karl Nordström,¹ Frédéric Cremer, Réka Tóth,² Martin Hartke, Samson Simon, Jonas R. Klasen, Ingmar Bürstel,³ and George Coupland⁴

Max Planck Institute for Plant Breeding Research, D-50829 Cologne, Germany

ORCID ID: 0000-0001-5291-1082 (M.C.B.)

Diurnal patterns of gene transcription are often conferred by complex interactions between circadian clock control and acute responses to environmental cues. *Arabidopsis thaliana* GIGANTEA (GI) contributes to photoperiodic flowering, circadian clock control, and photoreceptor signaling, and its transcription is regulated by the circadian clock and light. We used phylogenetic shadowing to identify three evolutionarily constrained regions (conserved regulatory modules [CRMs]) within the GI promoter and show that CRM2 is sufficient to confer a similar transcriptional pattern as the full-length promoter. Dissection of CRM2 showed that one subfragment (CRM2-A) contributes light inducibility, while another (CRM2-B) exhibits a diurnal response. Mutational analysis showed that three ABA RESPONSE ELEMENT LIKE (ABREL) motifs in CRM2-A and three EVENING ELEMENTS (EEs) in CRM2-B are essential in combination to confer a high amplitude diurnal pattern of expression. Genome-wide analysis identified characteristic spacing patterns of EEs and 71 *A. thaliana* promoters containing three EEs. Among these promoters, that of FLAVIN BINDING KELCH REPEAT F-BOX1 was analyzed in detail and shown to harbor a CRM functionally related to GI CRM2. Thus, combinatorial interactions among EEs and ABRELS confer diurnal patterns of transcription via an evolutionarily conserved module present in GI and other evening-expressed genes.

INTRODUCTION

Plants are sensitive to a wide range of environmental stimuli that influence the transcription of distinct sets of genes. Many of these responses, such as those to light and low temperature, are also regulated by the circadian clock, which is an internal time-keeping mechanism that generates a rhythm of ~24 h. The clock controls transcription of genes, mediating such responses to ensure that they are transcribed at particular times of day. In addition, their transcriptional induction by environmental stimuli is often limited by the clock to specific times, a process called gating (Harmer, 2009; de Montaigu et al., 2010). Thus, the diurnal patterns of transcription of clock-regulated genes are mediated by complex combinations of clock and environmental control (Millar and Kay, 1996). Here, we use an evolutionary approach to define the

mechanism by which light and clock signals converge on a promoter in *Arabidopsis thaliana* and show that this is based on interactions among evolutionarily conserved motifs.

The structure of the circadian clock of *A. thaliana* was defined by molecular-genetic and modeling approaches (Harmer, 2009; Nagel and Kay, 2012). Recent models represent this circuit as two major feedback loops in the morning and the evening that interact via additional loops (Pokhilko et al., 2012). The morning loop mainly comprises the MYB-like transcription factors LATE ELONGATED HYPOCOTYL (LHY) and CIRCADIAN CLOCK ASSOCIATED1 (CCA1) as well as the DNA binding PSEUDO RESPONSE REGULATORS (PRRs) PRR7 and PRR9 (Schaffer et al., 1998; Wang and Tobin, 1998; Makino et al., 2000). By contrast, the evening loop includes EARLY FLOWERING3 (ELF3), ELF4, and LUX ARRHYTHMO, which act together in a transcriptional complex termed the EVENING COMPLEX (EC) (Hicks et al., 2001; Doyle et al., 2002; Dixon et al., 2011; Helfer et al., 2011; Nusinow et al., 2011; Herrero et al., 2012). This complex feeds back to repress its own transcription as well as repressing the expression of the other evening loop genes, including *TIMING OF CAB EXPRESSION1*, which encodes a PRR, and *GIGANTEA (GI)*, which encodes a large plant-specific protein (Fowler et al., 1999; Park et al., 1999; Strayer et al., 2000). As well as acting within the clock mechanism, many of these components regulate sets of other genes that contribute to a wide range of phenotypic traits, and these are then expressed at specific times within the circadian cycle (de Montaigu et al., 2010; Nagel and Kay, 2012). Around 30 to 40% of the *A. thaliana* transcriptome is transcriptionally regulated by the circadian clock in this way (Harmer, 2009).

¹ Current address: Universität des Saarlandes, Postfach 151150, D-66041 Saarbrücken, Germany.

² Current address: Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

³ Current address: Department of Chemical Engineering and Bioengineering, Stanford University, Stanford, CA 94305.

⁴ Address correspondence to coupland@mpipz.mpg.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: George Coupland (coupland@mpipz.mpg.de).

Some figures in this article are displayed in color online but in black and white in the print edition.

Online version contains Web-only data.

Articles can be viewed online without a subscription.

www.plantcell.org/cgi/doi/10.1105/tpc.114.129437

In addition to acting as a component of the evening loop, *GI* contributes to many other processes. Mutations in *GI* cause late flowering under long days (LDs) (Rédei, 1962), a long hypocotyl (Araki and Komeda, 1993), impaired phytochrome signaling (Huq et al., 2000), short circadian period (Park et al., 1999; Mizoguchi et al., 2005), resistance to oxidative stress (Kurepa et al., 1998), increased starch and sucrose content (Eimert et al., 1995; Dalchau et al., 2011), as well as sensitivity to exposure to low temperatures (Cao et al., 2005). *GI* mRNA abundance is circadian clock controlled, showing a peak around 8 to 10 h after dawn (Fowler et al., 1999; Park et al., 1999). Transcription of *GI* is also activated by exposure to light, and this response was found to be gated to the morning, out of phase with the major peak in *GI* transcription (Patiel et al., 2006). *GI* has no homology to proteins of known function, precluding predictions of how it might influence such a wide range of phenotypes at the biochemical level. However, the observation that *GI* interacts with and stabilizes the ubiquitin ligase ZEITLUPE (*ZTL*), which contributes to the regulation of the circadian clock, suggested how it might modulate circadian rhythms (Kim et al., 2007). Furthermore, FLAVIN BINDING, KELCH REPEAT, F-BOX1 (*FKF1*), which is closely related to *ZTL*, is also stabilized by interaction with *GI* (Sawa et al., 2007; Fornara et al., 2009). This complex is required for the degradation of a family of DOF transcription factors called CYCLING DOF FACTORS (CDFs), which are transcriptional repressors of the flowering-time gene *CONSTANS* (*CO*) (Imaizumi et al., 2005; Fornara et al., 2009). Subsequently, *GI*, *FKF1*, and the CDFs were also implicated in direct regulation of *FLOWERING LOCUS T* (*FT*), the flowering time gene directly activated by *CO* soon after the peak in *GI* expression (Sawa and Kay, 2011; Song et al., 2012). Circadian clock-controlled transcription of *GI* thus facilitates the formation of active protein complexes with *FKF1* and *ZTL* proteins around 10 to 12 h after dawn, contributing to photoperiodic flowering and the evening loop of the circadian clock.

The mechanisms by which *GI* transcription is controlled to create the gated response to light and the coregulation of *GI* and *FKF1* have not been extensively analyzed. Mutations in the morning loop components *LHY* and *CCA1* cause *GI* transcription to occur earlier in the diurnal and circadian cycles, suggesting that when these proteins are expressed in the morning, they repress *GI* transcription (Mizoguchi et al., 2002). This suggestion was supported by the demonstration that in the morning, *CCA1* binds directly to the *GI* promoter in vivo (Lu et al., 2012). *LHY* and *CCA1* bind to several motifs closely related in sequence. These include the EVENING ELEMENT (EE; AAAATATCT), which was identified in the promoters of *A. thaliana* clock-controlled genes that are coexpressed late in the day (Harmer et al., 2000; Harmer and Kay, 2005). The EE is closely related to the previously identified *CCA1* BINDING SITE (CBS), which confers transcriptional activation in response to light in the morning (Wang et al., 1997). These motifs are also recognized by related MYB transcription factors, including *LHY* and *REVEILLE8* (*REV8*) (Alabadí et al., 2001; Rawat et al., 2011). In the promoter of the EC gene *ELF4*, *LHY* and *CCA1* bind to EEs where they act as repressors of light induction that is conferred through *FHY1-FAR1* binding sites (CACCGCG) and ACGT-containing elements (Li et al., 2012). In addition to light regulation, EEs are also enriched in the promoters of cold-induced genes (Mikkelsen and Thomashow, 2009; Dong et al.,

2011), and *CCA1* binds directly to the promoters of *C-REPEAT BINDING FACTOR* genes (Dong et al., 2011), which encode transcription factors that activate a major cold-responsive regulon. In cold-responsive promoters, EEs are often found in proximity to ABA RESPONSE ELEMENT LIKE (ABREL) (AGCTG) motifs (Mikkelsen and Thomashow, 2009). These motifs are predicted to bind bZIP transcription factors and were first identified in stress-induced genes (Busk and Pagès, 1998; Choi et al., 2000; Uno et al., 2000; Suzuki et al., 2005).

Systematic comparison of promoters of orthologous genes isolated from several closely related species, so-called phylogenetic shadowing, can be a powerful means of identifying functionally important promoter motifs (Cliften et al., 2001; Boffelli et al., 2003). This approach has allowed the computational prediction of enhancers in the promoters of genes from a broad range of organisms, including insects and vertebrates (Stark et al., 2007; Lindblad-Toh et al., 2011; Yáñez-Cuna et al., 2013). In *A. thaliana*, it has been applied to the analysis of developmental processes. Examination of the long *AGAMOUS* second intron identified conserved motifs required for its spatial pattern of transcription (Hong et al., 2003). Similarly, motifs in the *CRABS CLAW* promoter that contribute to expression in nectaries, in the *FT* promoter that contribute to photoperiodic response, and in the *LEAFY* promoter that confer responsiveness to auxin were identified (Lee et al., 2005; Adrian et al., 2010; Yamaguchi et al., 2013). In circadian clock-controlled transcriptional regulation, functionally important motifs identified by other means were shown to be evolutionarily conserved (Michael et al., 2008; Spensley et al., 2009; Herrero et al., 2012).

The temporal pattern of transcription of *GI* shows extreme evolutionary conservation, exhibiting a peak in expression around 8 to 12 h after dawn in several distantly related plant species (Hayama et al., 2003; Dunford et al., 2005; Zhao et al., 2005; Hecht et al., 2007). Here, we use evolutionary conservation for phylogenetic shadowing to identify conserved regulatory modules (CRMs) in the promoters of *GI* orthologs. Within CRM2, three EEs and three ABREL motifs are present in functionally separable subfragments but act in combination to confer diurnal, circadian, and light-inducible patterns of transcription. Genome-wide analysis of the distribution of EE showed that they are often clustered as in *GI* and that genes containing three EEs in their promoters are enriched for clock-regulated genes. This set includes *FKF1*, whose promoter contains a CRM related to that of CRM2 of *GI*, but the functionally important motifs in *GI* CRM2 are shuffled in *FKF1*, suggesting a relatively distant evolutionary relationship.

RESULTS

Identification of Evolutionarily Conserved Modules in the *GI* Promoter

A comparative evolutionary approach was employed to analyze transcriptional regulation of *GI* by identifying conserved regions within the promoters of orthologs from other *Brassicaceae* species. Around 2.5 kb upstream of the translational start site of *GI* confers on a luciferase marker gene the same circadian pattern of expression as shown by *GI* mRNA (Onai et al., 2004; Ding et al.,

2007; Edwards et al., 2010). Similar regions upstream of *Gl* homologs were extracted from a further seven *Brassicaceae* species and compared using VISTA (Figure 1A; see Methods). To support the idea that the compared sequences conferred a similar pattern of expression to *Gl*, the promoter of the *Gl* gene of *Arabidopsis thaliana* (*Aa-Gl*), which is most distantly related to *A. thaliana Gl* among the species analyzed, was fused to the *LUCIFERASE (LUC)* gene and introduced into *A. thaliana*. Under diurnal cycles, luciferase activity in these transgenic plants showed an almost identical rhythmic pattern to plants carrying *Gl:LUC* (Figures 1B and 1C).

The VISTA plot using *Aa-Gl* promoter as the base identified three major regions exhibiting more than 75% similarity between *Aa-Gl* promoter and the *Gl* ortholog promoters isolated from the other species. These three regions were named CRM1, CRM2, and CRM3 (Figure 1A) and are defined in Supplemental Table 1.

Analysis of Activity of Promoter Fusion Constructs Demonstrates That CRM2 but Not CRM1 or CRM3 Is Sufficient to Confer Rhythmic Expression

To assess the independent activity of each of the three CRMs isolated from *A. thaliana Gl*, they were fused individually to a minimal promoter (*pNOPALINE SYNTHASE [pNOS]*) attached to the luciferase marker gene (*LUC*). To employ this modular strategy, a 105-bp fragment of *pNOS* that contains a transcriptional start site but shows very weak transcriptional activity (Puente et al., 1996) was first fused to *LUC* downstream of a Gateway recombination site. The activities of different promoter modules were then tested via their effect on the activity of *pNOS*. First, the activity of each construct was determined by measuring the average absolute luminescence of the *CRM-NOS:LUC* transgenic plants under diurnal cycles of 8 h light/16 h dark (8L:16D). The *CRM2-NOS:LUC* plants showed ~50% of the luciferase activity of *Gl:LUC* plants, indicating that CRM2 was sufficient to drive high rates of transcription from a minimal promoter (Figure 2A). By contrast, *CRM1-NOS:LUC* and *CRM3-NOS:LUC* gave rise to much less than 10% of the luciferase activity of *Gl:LUC*. These experiments suggested that CRM2 was the most important module in conferring the *Gl* expression pattern, and all subsequent experiments focused on that region.

The pattern of *LUC* activity was then tested under diurnal cycles (Figure 2B). *CRM2-NOS:LUC* showed a very similar pattern of expression to *Gl:LUC* conferring a peak in activity toward the end of the diurnal cycle (Figure 2B). Thus, CRM2 contains all the information necessary to drive a rhythmic expression pattern similar to that of the full-length *Gl* promoter.

Deletions of the *Gl* promoter were then employed to assess the contribution of CRM2 to the regulation of *Gl* transcription. Deletions of the 2.5-kb *Gl* promoter constructed the 1.8-kb promoter, which retained CRM2 but ended immediately 5' of it, and the 1.1-kb promoter in which CRM2 was deleted. Transformants carrying the 2.5-, 1.8-, and 1.1-kb fusions were tested for absolute luminescence (Supplemental Figure 1), as described previously for the CRM fusions to *NOS:LUC* (Figure 2A). The 2.5- and 1.8-kb fusions showed similar activity, whereas plants carrying the 1.1-kb *Gl:LUC* fusion were reduced both in amplitude and in absolute *LUC* activity but retained a diurnal peak in expression in the evening (Figures 2C and 2D). These results suggest that CRM2 contains elements required for high amplitude *Gl* promoter activity and demonstrate that

both the proximal 1.1-kb *Gl* promoter and CRM2 can drive diurnal expression with an evening peak, indicating functional redundancy within the full-length promoter.

To test whether CRM2 is sufficient to confer *Gl* biological activity, a fusion of CRM2 to the *Gl* cDNA was constructed. The *CRM2:Gl* fusion was introduced into the *gi-2* mutant and the flowering time of the transgenic plants compared with the wild type, *Gl* overexpressing *35S:Gl*, and the *gi-2* mutant. Under long days, *CRM2:Gl* plants flowered much earlier than *gi-2* mutants and at a similar time, only around three to four leaves later, to wild-type plants (Figures 2E and 2F). Under short days (SDs), *CRM2:Gl* flowered at a similar time to the wild type and *gi-2*, but much later than *35S:Gl* (Figure 2F), indicating that CRM2 did not confer ectopic *Gl* activity causing early flowering under non-inductive conditions. These results suggest that CRM2 confers biological activity similar to that of the full-length promoter.

Overall, these experiments demonstrate that the 698-bp conserved CRM2 region confers a diurnal pattern of transcriptional regulation similar to that of *Gl* and that this is sufficient to drive functional *Gl* activity during control of photoperiodic flowering.

Subdivision of CRM2 Identifies Separate Promoter Regions Conferring Different Aspects of Diurnal Regulation

To increase the resolution of analysis of CRM2, the region was divided into three smaller fragments that each contained conserved motifs (Figure 3A; Supplemental Table 1). These fragments, CRM2-A, CRM2-B, and CRM2-C, were fused to the *NOS* minimal promoter and introduced into Landsberg *erecta (Ler)*. The resulting transgenic plants were analyzed for luciferase expression pattern during a diurnal cycle.

Transgenic plants carrying *CRM2-A-NOS:LUC* growing under diurnal cycles showed a steep rise in *LUC* activity at dawn that then remained high during the photoperiod and fell gradually during the following dark period (Figure 3B). Therefore, activity of *CRM2-A-NOS:LUC* increased in expression much earlier during the photoperiod than *Gl:LUC* or *CRM2-NOS:LUC*, and the steep rise in expression at dawn suggests that it confers light inducibility (Figures 2B and 3B). *CRM2-B-NOS:LUC* activity rose slightly later during the photoperiod than *Gl:LUC* activity, reached a later peak during the dark period, and then fell in activity (Figure 3C). CRM2-B therefore contains all of the information required for a similar diurnal pattern to CRM2 but rises in activity less quickly and reaches peak levels later. By contrast, with CRM2-A and CRM2-B, CRM2-C did not confer rhythmic *LUC* activity, but rather *CRM2-C-NOS:LUC* was expressed constantly throughout the diurnal cycle (Figure 3D). Taken together, the analysis of these three subfragments of CRM2 suggests that this conserved region is modular with the sequences that confer rhythmic activity mainly contained within CRM2-B and those within CRM2-A largely contribute rapid induction at dawn, which might be an acute response to light.

Contribution of Conserved ABREs and EEs to the Diurnal Regulation Conferred by CRM2

To identify potential transcription factor binding sites within CRM2-A and CRM2-B that might confer their different regulatory

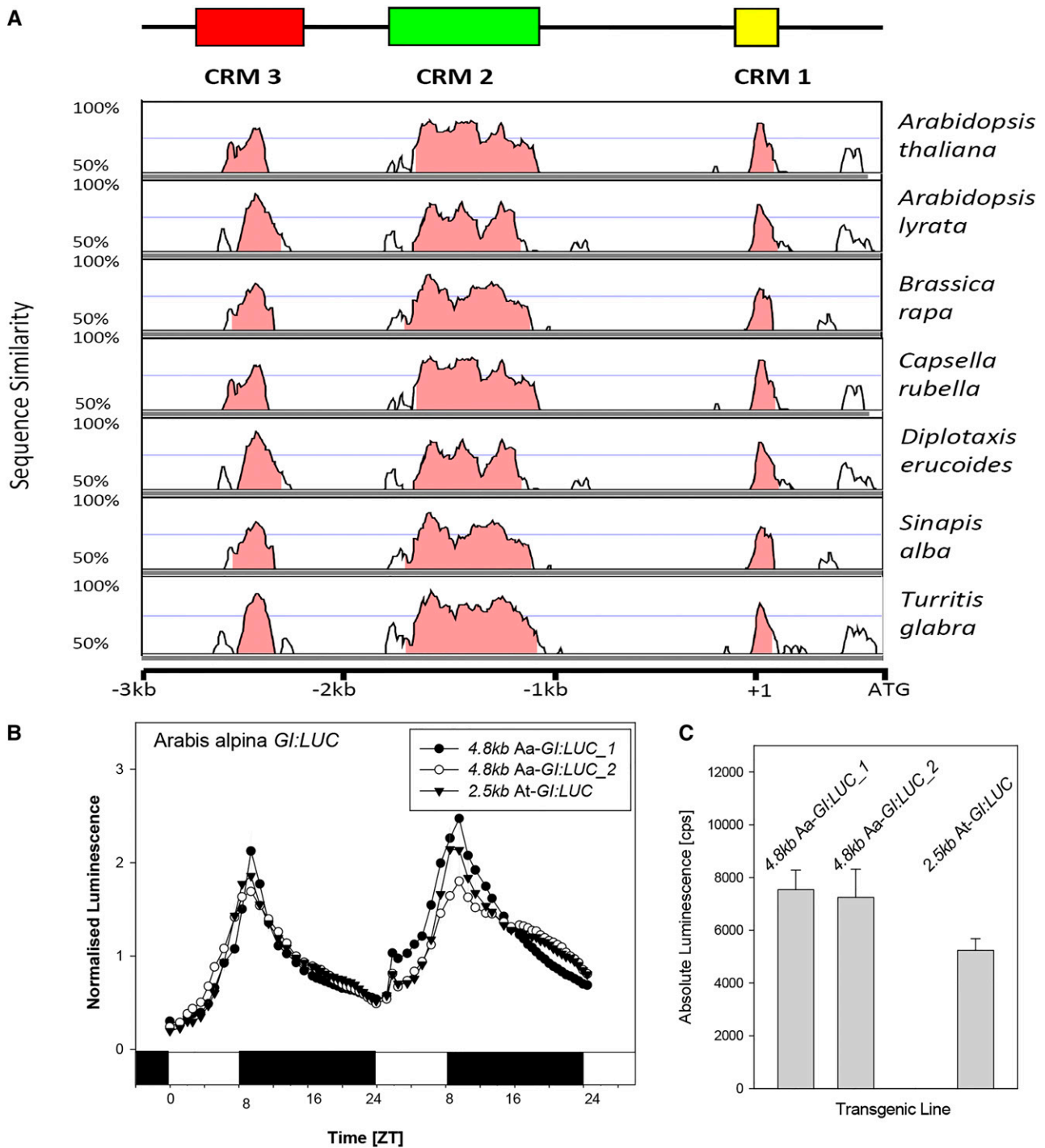


Figure 1. Phylogenetic Shadowing Identifies Three Evolutionarily Conserved Blocks within the *Gl* Promoter.

(A) Phylogenetic shadowing of the promoters of *Gl* orthologs isolated from eight *Brassicaceae* species. The *Gl* promoters of *A. thaliana*, *A. lyrata*, *Brassica rapa*, *Capsella rubella*, *Diplotaxis erucoides*, *Sinapis alba*, and *Turritis glabra* were each aligned to the *Gl* promoter of *A. alpina*. Three CRMs were identified based on this analysis (indicated as CRM1, CRM2, and CRM3). All pairwise alignments are presented as VISTA plots. Pink color indicates regions with at least 70% of conservation on a 100-bp sliding window.

(B) and **(C)** Comparison of the expression patterns of 2.5-kb *AtGl:LUC* and two independent homozygous lines of 4.8-kb *AaGl:LUC* in *A. thaliana*. Seven-day-old plants carrying 2.5-kb *AtGl:LUC* and 4.8-kb *AaGl:LUC* were grown under cycles of SDs (8 h light/16 h dark) for 48 h. Luciferase activity was continuously measured with a TopCount.

activities, conserved sequence motifs in these regions of all eight analyzed promoters were identified by creating multiple sequence alignments of CRM2 using ClustalX (Larkin et al., 2007) and were subsequently visualized with WebLogo (Figure 3A) (Crooks et al., 2004). An overview of all conserved *cis*-regulatory elements within CRM2 can be found in Supplemental Table 2.

Interestingly, two types of conserved *cis*-regulatory elements each occurred three times within CRM2, and we therefore focused on characterizing their contribution to the diurnal pattern of transcription conferred by this module. CRM2-A contained three conserved sequences (ACGTG) that are identical to the ABREL motif (Figure 3A) (Busk and Pagès, 1998; Suzuki et al., 2005; Mikkelsen and Thomashow, 2009). To test the contribution of these sequences to the activity of CRM2, all three were mutated to generate the triple ABREL mutant module, CRM2-[TA]. This mutant fragment was then inserted adjacent to the minimal NOS promoter to generate the construct CRM2-[TA]-NOS:LUC. Transgenic *Ler* plants carrying this construct were tested for LUC activity under diurnal cycles (Figure 3E; Supplemental Figures 2A and 3A). LUC activity in these plants rose slightly later than in CRM2-NOS:LUC controls and did not reach as high an amplitude. These results indicate that the ABREL motifs make a contribution to the steep rise in *Gl* expression at peak time and to its amplitude under diurnal cycles.

CRM2-B contained three conserved EEs (Figure 3A) (Harmer et al., 2000; Harmer and Kay, 2005). All three EEs were mutated in the context of CRM2 to generate the triple EE mutant construct CRM2-[TEE]-NOS:LUC. Transgenic *Ler* plants containing this construct showed an immediate rise in LUC activity at dawn that remained at the same level during the photoperiod and then fell in the dark (Figure 3F; Supplemental Figures 2B and 3B). Therefore, the EEs contribute to suppressing CRM2 activity at dawn and to increasing amplitude of induction in the evening.

To test the combined activity of the ABREs and the EEs, all six motifs were disrupted in the context of CRM2 to generate the CRM2-[TA-TEE]-NOS:LUC construct. Transgenic plants containing this construct showed no diurnal pattern of expression under diurnal cycles, indicating that these six motifs are essential for the diurnal rhythmicity conferred by CRM2 (Figure 3G; Supplemental Figures 2C and 3C). Thus, the combined activity of these motifs is required for diurnal regulation of transcription conferred by CRM2.

CRM2 Is Required and Sufficient for Gated Light Induction of the *Gl* Promoter

In addition to circadian clock regulation, *Gl* transcription shows an acute response to exposure to light (Paltiel et al., 2006). To monitor this in more detail, 2.5kb *Gl*:LUC and 1.8 kb *Gl*:LUC plants were grown under diurnal cycles, transferred to continuous darkness, and then different batches of plants were exposed to 30-min pulses

of white light at 2.5-h intervals for 25 h. In this gating assay, the amplitude of the light response as measured by LUC activity depended on the time during subjective night at which plants were exposed to light (Figure 4, Table 1; Supplemental Figure 4 and Supplemental Table 3). The highest amplitude response occurred when the light pulse was given in the evening of the subjective day, around 12 h after subjective dawn. Therefore, 2.5kb *Gl*:LUC and 1.8 kb *Gl*:LUC exhibit clock-gated induction by light in a similar phase to its circadian peak of expression. The 1.1kb *Gl*:LUC plants were then tested for clock-gated light induction. These transgenic plants showed no high amplitude response in subjective evening (Figure 4B) but exhibited a similar reduced induction by light at all times tested. This experiment demonstrated therefore that CRM2 is necessary for the gated light induction of the *Gl* promoter.

To test whether CRM2 was sufficient to confer gating of the light response, the CRM2-NOS:LUC plants were tested in the same assay. A higher response to light was detected in subjective evening than subjective morning (Figure 4C), and statistical analysis demonstrated that CRM2 is sufficient for clock-regulated gating of light induction (Table 1; Supplemental Figure 4B and Supplemental Table 3; see Methods). Taken together, these experiments indicate that CRM2 is both required and sufficient for the light-gated induction of *Gl* expression in the evening.

Contribution of Subregions of CRM2 to the Gated Light Response

The steep rise in CRM2-A-NOS:LUC expression at dawn under diurnal conditions (Figure 3B) suggested that the CRM2-A region contributed to the light induction conferred by CRM2. To test this idea, CRM2-A-NOS:LUC plants were analyzed in the gating assay described above (Figure 4D). CRM2-A-NOS:LUC activity was found to respond strongly to light with similar amplitude at each time point. Therefore, CRM2-A retained the light responsiveness of *Gl*:LUC but did not show a gated response at the end of the subjective day. In addition, CRM2-B-NOS:LUC showed induction in response to light pulses with highest amplitude toward the end of the subjective day (Figure 4E), as observed for CRM2-NOS:LUC (Figure 4C). These results suggest that CRM2-B is sufficient to confer a gated light response, although the absolute expression level was only 10% of that found after light induction of CRM2-A-NOS:LUC (Supplemental Figure 1). CRM2-C-NOS:LUC did not show light induction (Figure 4F). Taken together, these experiments demonstrate that CRM2 confers a gated light response to the *Gl* promoter, and this is mediated by separable subfragments within CRM2 that act together combinatorially. Whereas CRM2-A contributes to a high amplitude of light induction, CRM2-B contributes clock-controlled expression and CRM2-C enhances expression. Therefore, these three regions act in combination to provide the high amplitude gated response to light conferred by CRM2.

Figure 1. (continued).

(B) Normalized Luciferase activity is shown for 48 h. Data represent means of at least eight measured seedlings for each line. Black and white boxes indicate dark/light cycles. Error bars indicate SE. ZT is Zeitgeber time and here represents time from lights on at dawn.

(C) Absolute Luciferase activity, averaged above 48 h. Data represent means of at least eight measured seedlings for each line. Error bars indicate SE.

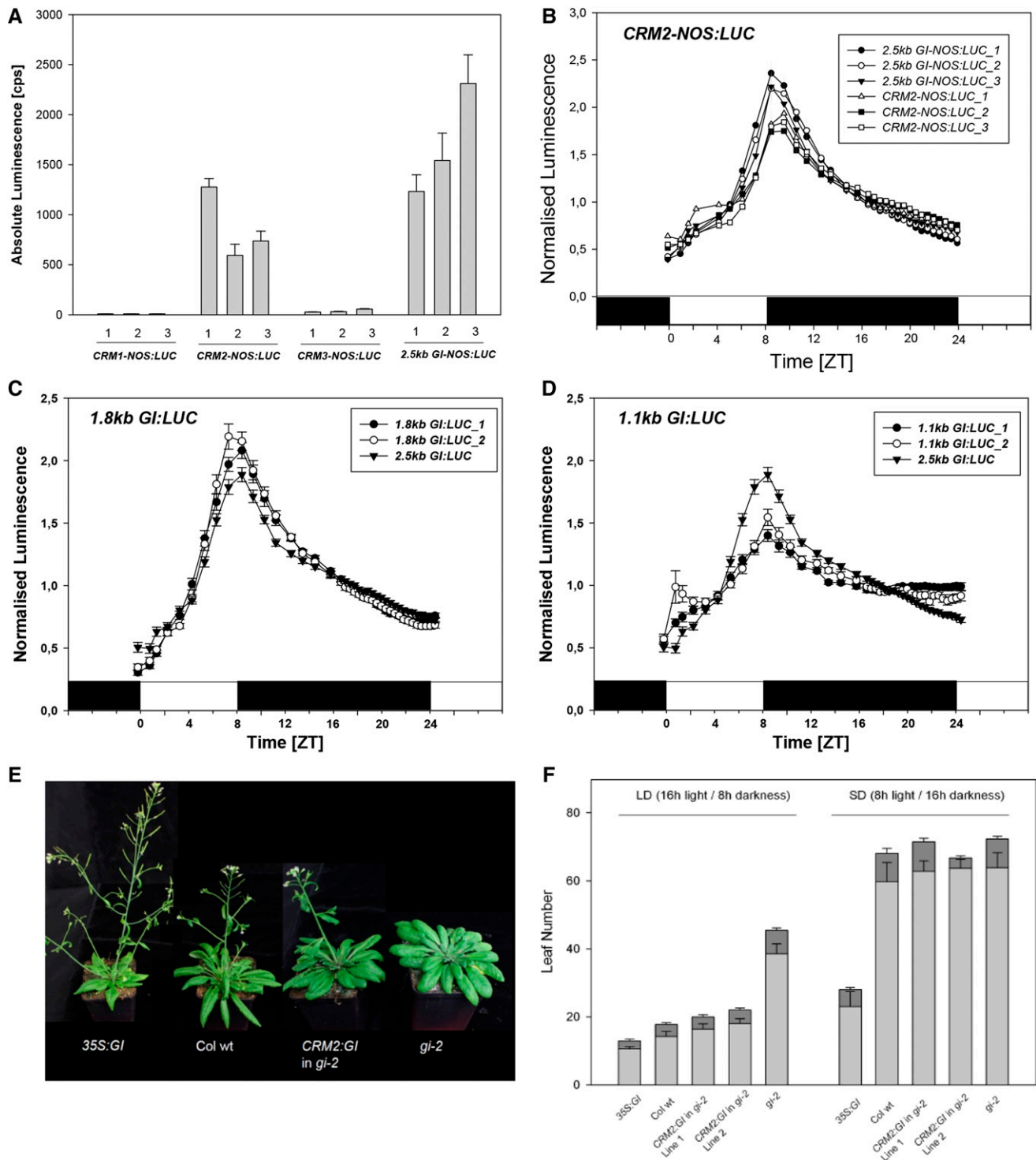


Figure 2. Analysis of Three CRMs within the *A. thaliana* *Gl* Promoter.

(A) Absolute luciferase activity conferred by three CRMs within the *Gl* promoter. CRM1, CRM2, CRM3, and the 2.5-kb *Gl* promoter were fused to a *NOS* minimal promoter and cloned upstream of the luciferase coding sequence. Homozygous lines derived from three independent transformants for each construct were measured in a TopCount for 24 h under SD conditions (8 h light/16 h dark). Average luciferase expression of eight individual 7-d-old seedlings is shown for each line. Error bars indicate *se*.

Contribution of ABREL and EEs to the Gated Light Response Conferred by CRM2

The roles of the conserved ABREs and EEs in CRM2 to the gated light response were then tested using the mutant constructs. *CRM2-[TA]-NOS:LUC* plants showed a response to light at the end of the subjective day (Figure 4G; Supplemental Figure 4C), as observed for *Gl:LUC* or *CRM2-NOS:LUC*. Statistical analysis confirmed the gated light response conferred by this construct (Table 1; Supplemental Table 3). By contrast, mutation of all three EEs in the *CRM2-[TEE]-NOS:LUC* plants dramatically reduced the gated light response (Figure 4H). No induction at the end of the subjective day was detected, but a weak increase around subjective dawn was found, consistent with the pattern under diurnal conditions (Figure 3F). However, statistical analysis did not detect a significant gated response ($P < 0.01$) (Table 1; Supplemental Figure 4D and Supplemental Table 3). These results indicate that the EEs are essential for the gated light response of CRM2 at dusk.

To test the combined activity of the ABREs and the EEs on the gated light response, the *CRM2-[TA-TEE]-NOS:LUC* construct was tested. This construct did exhibit a response to light but no statistically significant gating of the response was detected (Figure 4I, Table 1; Supplemental Figure 4E and Supplemental Table 3). Also, mutation of the ABREL motifs abolished any residual gating response observed around subjective dawn in the *CRM2-[TEE]-NOS:LUC* plants. The absence of gating of the light response in these plants is consistent with their lack of a diurnal pattern (Figure 3G) and demonstrates that these six motifs are essential for all diurnal and light-gated responses conferred by CRM2.

Relationship between LHY CCA1 and the EEs in Regulation of the *Gl* Promoter

The MYB transcription factor CCA1 has been shown to bind to the *Gl* promoter (Lu et al., 2012), while CCA1 and LHY are functionally closely related, exhibiting genetically redundant effects on circadian clock function (Alabadí et al., 2002; Mizoguchi et al., 2002). Whether the *cis*-acting triple EE mutation had an identical effect on *CRM2-NOS:LUC* expression to inactivation of the *trans*-acting factors LHY and CCA1 was therefore tested (Figure 5).

CRM2-NOS:LUC and *2.5kb Gl:LUC* were introduced into the *lhy-12 cca1-1* double mutant and LUC activity scored through a diurnal cycle. Both constructs conferred similar diurnal patterns of LUC activity, with expression rising steeply from dawn and peaking toward the end of the photoperiod (Figure 5A). Although the pattern was similar for both constructs, *CRM2-NOS:LUC* showed a lower amplitude peak. These patterns were distinct from those conferred by the same constructs in the *Ler* background, where no rise in LUC activity was observed at dawn (Figure 2). These results suggest that LHY and CCA1 suppress expression of LUC from the *Gl* and *CRM2-NOS* promoters early in the day so that in the double mutant, the peak in LUC activity is shifted earlier in the diurnal cycle.

The *CRM2-[TEE]-NOS:LUC* construct was also introduced into the *lhy cca1* double mutant. In these plants, the diurnal pattern of LUC activity did not rise steeply at dawn and produced a lower amplitude peak than *CRM2-NOS:LUC* (Figure 5B). The pattern of LUC activity for *CRM2-[TEE]-NOS:LUC* was therefore similar in *Ler* and *lhy cca1* plants but more strongly impaired than *CRM2-NOS:LUC* in *lhy cca1* background (Figures 3F and 5B). These results demonstrate that *lhy cca1* has a weaker effect on *CRM2-NOS:LUC* expression than the TEE mutations, suggesting that other factors also regulate transcription from CRM2 through the EEs.

The effect of the *lhy cca1* background on gated light induction of *2.5kb Gl:LUC* was also tested. In *Ler*, *2.5kb Gl:LUC* shows a strongly gated response to light in the subjective evening (Figure 5C). In *lhy cca1*, *2.5kb Gl:LUC* responded weakly to light pulses during the first 12 h after transfer to darkness, as observed in *Ler*, but was then strongly induced by light at all times tested for the remaining 36 h in the dark (Figure 5D). This pattern suggests that gating of *Gl:LUC* can occur during the first cycle in the dark in the absence of LHY CCA1 but that it is impaired in later cycles when circadian rhythms dampen in the mutant.

Genome-Wide Distribution of EEs and ABREs Suggests Cooperative Interactions in the Regulation of Many *A. thaliana* Genes

The analysis of the *Gl* promoter described above demonstrated that evolutionarily conserved EEs and ABREs combine to generate a high amplitude peak in *Gl* expression under diurnal conditions. To test whether clusters of EEs and ABREs are frequently

Figure 2. (continued).

(B) Diurnal expression pattern of CRM2. Each three independent homozygous lines of *CRM2-NOS:LUC* and *2.5kb Gl-NOS:LUC* were measured in a TopCount for 24 h under SD conditions (8 h light/16 h dark). Black and white boxes indicate dark/light cycles. At least eight 7-d-old seedlings were scored for each line and the normalized luciferase activity is shown. Error bars indicate *se*. ZT is Zeitgeber time and here represents time from lights on at dawn.

(C) and (D) Diurnal expression pattern of constructs with retained CRM2 (*1.8kb Gl:LUC*) or deleted CRM2 (*1.1kb Gl:LUC*). Each two independent homozygous lines of *1.8kb Gl:LUC* (C) and *1.1kb Gl:LUC* (D) were measured in a TopCount for 24 h under SD conditions (8 h light/16 h dark). Black and white boxes indicate dark/light cycles. At least 12 7-d-old seedlings were scored for each line and the normalized luciferase activity is shown. Error bars indicate *se*.

(E) Flowering phenotype of a CRM2 complementation line. CRM2 was fused to the *Gl* cDNA and introduced into the *gi-2* mutant. Plants shown were grown for 5 weeks under LD (16 h light/8 h dark) conditions.

(F) Flowering time of CRM2 complementation lines. Flowering time of two independent homozygous *CRM2:Gl* lines under LD (16 h light/8 h dark) and SD (8 h light/16 h dark) conditions, compared with *35S:Gl*, *gi-2*, and Columbia wild-type plants. Error bars indicate *sd* ($n = 10$). Both experiments were repeated twice with similar results.

[See online article for color version of this figure.]

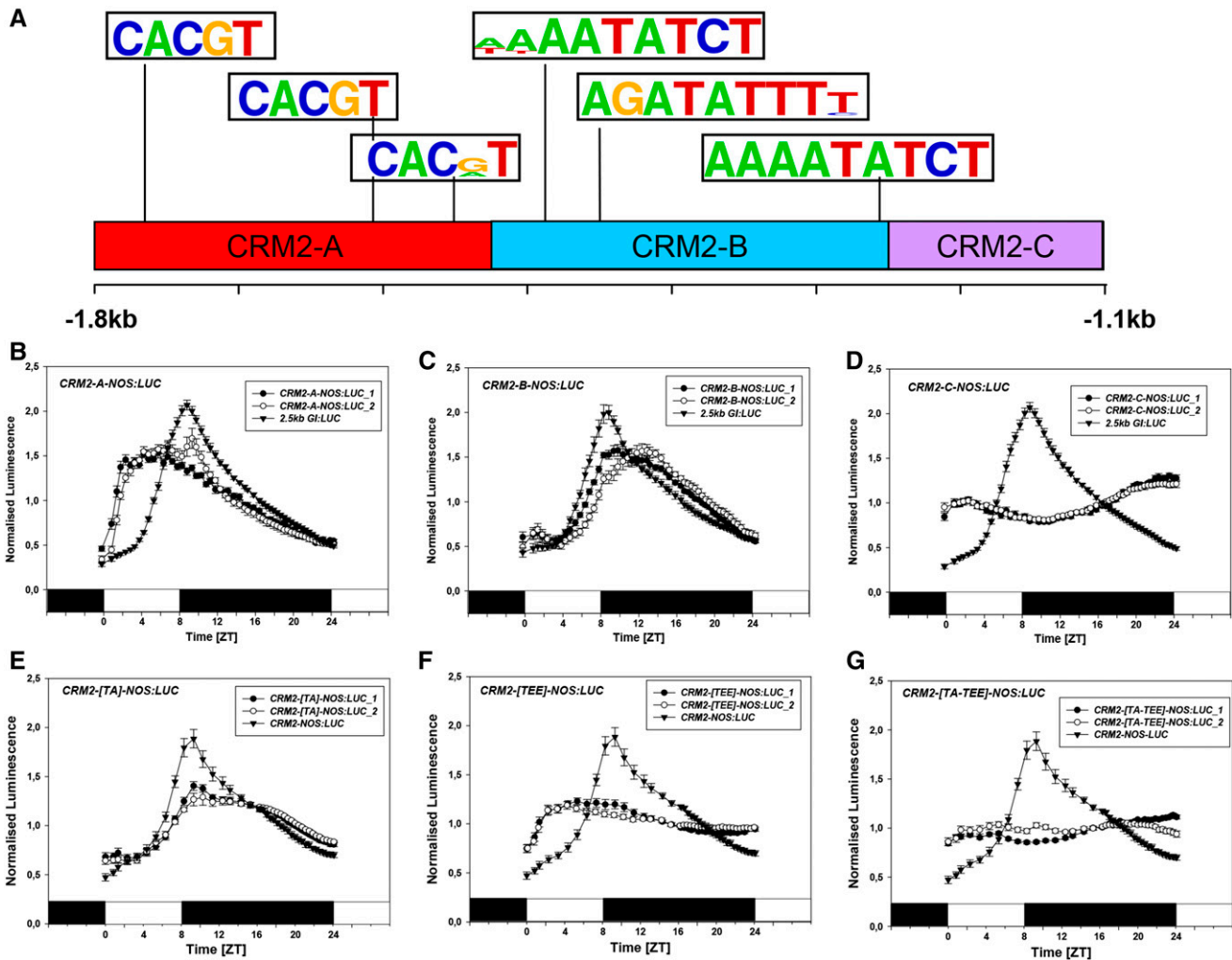


Figure 3. Combinatorial Interaction of Distinct Motifs Confers Diurnal Regulation of the *G1* Promoter.

(A) Highly conserved ABREL and EE motifs are distributed widely across conserved subfragments within CRM2. Conservation of ABREs and EEs is shown as WebLogos; larger letters indicate more highly conserved bases (see Methods for details).

(B) to (G) Diurnal expression patterns of different *G1* promoter fragments. Plants were grown for 7 d under SDs (8 h light/16 h dark) and then measured under this light regime for 24 h in a TopCount. Two independent homozygous lines of the indicated construct were compared with *CRM2-NOS:LUC* or *2.5kb G1:LUC*. All experiments represent the average of ~15 seedlings and were done at least twice with similar results. Black and white boxes indicate dark/light cycles. Error bars indicate SE. TA, triple ABREL mutation; TEE, triple EE mutation. ZT represents Zeitgeber time and here refers to time from lights on at dawn.

found in promoters of other *A. thaliana* genes, a whole-genome analysis was performed. First, a genome-wide analysis was performed to classify the promoters of all *A. thaliana* genes based on the number of EE sequences they contain identical to those found in the *G1* promoter (AAAATATCT). Only 99 genes, or 0.3% of those screened, contained three or more EEs within 3 kb of their translational start sites (Figure 6A). By contrast, 1.7% of genes contained two EEs and 13.7% harbored one EE. After manual inspection (see Methods), a list of 71 genes containing at least three EEs in their promoters was used for further analysis (list is provided in Supplemental Table 4). All of these promoters were then analyzed for the number of ABREs present. No difference in the average number of ABREs was detected for promoters

containing zero, one, or two EEs. However, those promoters containing three or more EEs also contained on average a significantly higher number of ABREs compared with promoters containing two or fewer EEs ($P = 0.02998$) (Figure 6B), indicating that there is a tendency for multiple EEs and ABREs to occur in the same promoters. In particular, only 4 of the 71 promoters containing at least three EEs harbored no ABREs (defined as ACGTG), whereas 40 contained at least three ABREs (Supplemental Table 4). Thus, the majority of promoters containing three EEs also contained multiple ABREs.

To test whether EEs (including strict EEs AAAATATCT [Harmer et al., 2000], EVENING ELEMENT-LIKE [AATATCT; Mikkelsen and Thomashow, 2009], and CBS [AAAAAATCT; Wang and Tobin, 1998]) tend to cluster across the whole genome, their positions

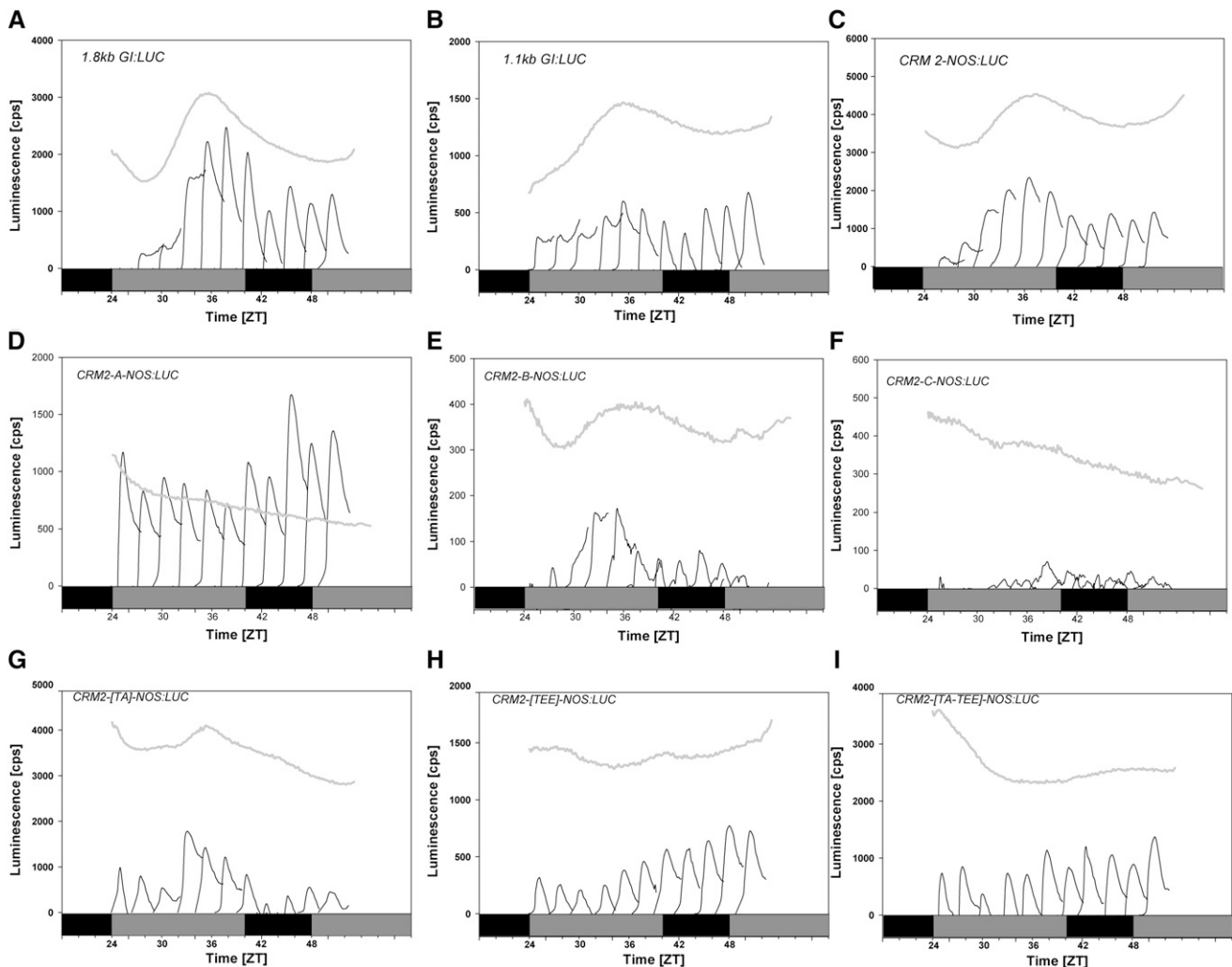


Figure 4. Combinatorial Interaction of Distinct Motifs Confers a Gated Response to the *Gl* Promoter.

Gating assay with different *Gl* promoter constructs. Seedlings were grown for 7 d under LD conditions (16 h light/8 h dark) and then transferred to the dark. Replicate samples were exposed to a white light pulse of 30 min every 2.5 h and luciferase activity was monitored in a TopCount. Values represent the difference between measured mean values and the untreated dark sample (gray curve). Representative lines for the indicated constructs are shown. Black and gray boxes indicate subjective days and nights, respectively. All experiments represent the average of ~15 seedlings and were done at least twice with similar results. TA, triple ABREL mutation; TEE, triple EE mutation. ZT is Zeitgeber time and here represents time from lights on at dawn.

relative to one another in all *A. thaliana* promoters were determined. EEs on the same strand were found to be highly overrepresented within 100 bp of each other (Figure 6C). Within this short distance, several spacing distances occurred more frequently, the furthest of which was 84 bp. A similar analysis was performed for EEs on the opposite strand, and an interesting pattern was observed (Supplemental Figure 5). Only one distance was overrepresented within 100 bp, but a strong overrepresentation was detected at a spacing of between 255 and 260 bp. Taken together, these data suggest that EEs tend to be clustered, with a preference for being within 50 bp on the same strand.

Whether ABRELS (ACGTG) are clustered and tend to occur in the vicinity of EEs across the whole genome was then tested, as previously done specifically for cold-induced genes (Mikkelsen

and Thomashow, 2009). ABRELS were found to frequently occur at different positions within 100 bp of each other (Figure 6D). As ABRELS are overrepresented in EE-rich promoters (Figure 6B), we also investigated the relationship of EEs and ABRELS on a genome-wide level. Strikingly, ABRELS occurred at high frequency at defined positions within 250 bp of EEs (Figure 6E), demonstrating clustering of ABRELS and EEs across the genome and constraints on their spacing. As a control, the LUX BINDING SITE (LBS; defined as CGAATC) was used (Figure 6F), which is also found in the CRM2 of *Gl* (Supplemental Table 2). However, this element was not over represented in the vicinity of EEs (Figure 6F).

These analyses indicate that across the whole genome, promoters containing multiple EEs are more likely to harbor higher numbers of ABRELS and EEs tend to be clustered relative to

Table 1. Gating Scores of Constructs with Point Mutations

Construct	P Value	Gated
2.5 kb <i>Gl:LUC</i>	2.42E-11	Yes
<i>CRM2-NOS:LUC</i>	1.33E-08	Yes
<i>CRM2-[TA]-NOS:LUC</i>	0.000104	Yes
<i>CRM2-[TEE]-NOS:LUC</i>	0.025900	No
<i>CRM2-[TA-TEE]-NOS:LUC</i>	0.265000	No

Gating assays were performed as described in Figure 4, but each with three independent homozygous transgenic lines (number of individual seedlings: three to six for each line). All values were normalized to their own 24 h mean, and the deviation from this mean was calculated for each time point. P values are a cumulative calculation from two independent experiments with one homozygous line as shown in Figure 4 and one additional experiment with three independent homozygous lines. Statistical significance was determined in an ANOVA setting; P values < 0.01 indicate a gated response.

other EEs and to occur close to ABREs, suggesting that these motifs may be functionally interrelated, as mentioned above and in the Discussion for the *Gl* promoter.

***FKF1* Promoter Also Contains a Cluster of Three EEs and Three ABREs within a Functional CRM**

The whole-genome analysis described above indicated that EEs and ABREs tend to occur in clusters within promoters and that in *A. thaliana*, the promoters of 71 genes contain three EEs and on average a higher number of ABREs. Gene ontology analysis indicated that they were enriched for genes showing diurnal patterns of expression and for those involved in cold responses (Supplemental Figure 6A). The diurnal expression patterns of these 71 genes were analyzed using previously published microarray data (Mockler et al., 2007) (Supplemental Figure 6B). Among these 71 genes, the proportion with peaks in expression 10 and 12 h after dawn was significantly ($P < 0.05$) higher than in the genome as a whole (Supplemental Figure 6B) and those with peaks at 8 and 13 h were also overrepresented ($P < 0.1$). One of the genes containing three EEs in its promoter and exhibiting a strong diurnal rhythm similar to that of *Gl* was examined in detail to determine whether it contains a CRM analogous to CRM2 of *Gl*. For this analysis, *FKF1* was selected because its coexpression with *Gl* allows formation of a complex containing both proteins (Sawa et al., 2007) and the *FKF1* promoter harbors EEs and ABREs (Supplemental Figure 7).

Comparison of the promoters of the *FKF1* orthologs of *A. alpina* and *A. thaliana* revealed blocks of homology and, notably, a 485-bp region around the transcriptional start site (Figure 7A). This region termed CRM-*FKF1* contained three EEs and three ABREs, as observed for CRM2 of *Gl* (Supplemental Figure 7), although the locations and order of these regions within *FKF1* CRM and *Gl* CRM2 are not strictly conserved (Figure 7B). Regions 3.6 and 0.5 kb upstream of the translational start codon of *FKF1* were fused to *LUC* and transgenic plants constructed. *LUC* activity peaked around 12 h after dawn under diurnal conditions of 16 h light/8 h dark, as predicted from previously reported mRNA analysis (Figure 7C) (Nelson et al., 2000; Imaizumi et al., 2003). Strikingly, the diurnal and absolute pattern conferred by 0.5kb *FKF1:LUC* appeared identical to that of 3.6kb *FKF1:LUC*, indicating that the conserved proximal 0.5-kb region contains all the

information required to confer the *FKF1* expression pattern (Figure 7C; Supplemental Figures 1A and 1C). Comparison of the diurnal patterns of 2.5 kb *Gl:LUC* and 3.6 kb *FKF1:LUC* showed that these peak at similar times, particularly under LDs of 12 or 16 h (Figures 7D and 7E). Finally, light inducibility of *FKF1:LUC* was tested in a gating assay similar to that used previously to analyze *Gl:LUC* (Figure 4). *FKF1* showed peak response to light 10 h after subjective dawn (Figure 7F), at the same time as *Gl:LUC*, although the amplitude of induction of *FKF1:LUC* was lower (Figure 7G). Taken together, these experiments indicate that *Gl*-CRM2 and *FKF1*-CRM are functionally related and contain identical multiple EE and ABRE motifs (Supplemental Figures 7A and 7B).

DISCUSSION

A phylogenetic shadowing approach was used to systematically identify CRMs in the promoter of *Gl*. Gain- and loss-of-function experiments showed that CRM2 was required and sufficient to confer diurnal, circadian, and light-inducible patterns of transcription similar to those conferred by the full-length promoter. CRM2 was divisible into functionally separable subfragments that act combinatorially to confer high-amplitude expression in the evening. These subfragments contain clusters of three EEs and three ABRE motifs that were mutagenized, and these six motifs were essential for the diurnal pattern of expression conferred by CRM2. Related patterns of motifs are found in CRMs present in coregulated genes in *A. thaliana*. Our data indicate how combinatorial interactions among physically separable regions confer complex transcriptional regulation of *Gl* and contribute to our understanding of the evolution and identity of the transcriptional code controlling clock-regulated gene transcription in the *Brassicaceae*.

Identification and Structure of CRM2

Phylogenetic shadowing is based on comparing the promoters of orthologs isolated from relatively closely related species that are nevertheless sufficiently diverged so that neutrally evolving sequences differ while functionally important sequences are constrained (Cliften et al., 2001; Boffelli et al., 2003; Stark et al., 2007). Using genes from several related species increases the power of the analysis by increasing the amount of evolutionary divergence being exploited. The availability of genomic sequences from several *Brassicaceae* species has simplified this approach for studying promoters of *A. thaliana* genes (Hu et al., 2011; Wang et al., 2011). Also, the rate of neutral evolution in *A. thaliana* was estimated at around 7×10^{-9} /base/generation, which is high enough to allow variation at almost all nonconstrained bases of a typical promoter in at least one of the eight species analyzed here (Ossowski et al., 2010). In the case of *Gl*, comparison of the *A. thaliana* and *A. alpina* genes provided sufficient resolution to identify the CRMs, suggesting that the divergence time of these species was appropriate for these comparisons. However, the species of choice for such comparisons may depend on the gene under study, as for the *FT* promoter, comparison between *A. thaliana* and *Arabidopsis lyrata* already conferred high resolution (Adrian et al., 2010).

The CRMs we identified in the *Gl* promoter are conceptually similar to enhancers in promoters of metazoan genes, which are

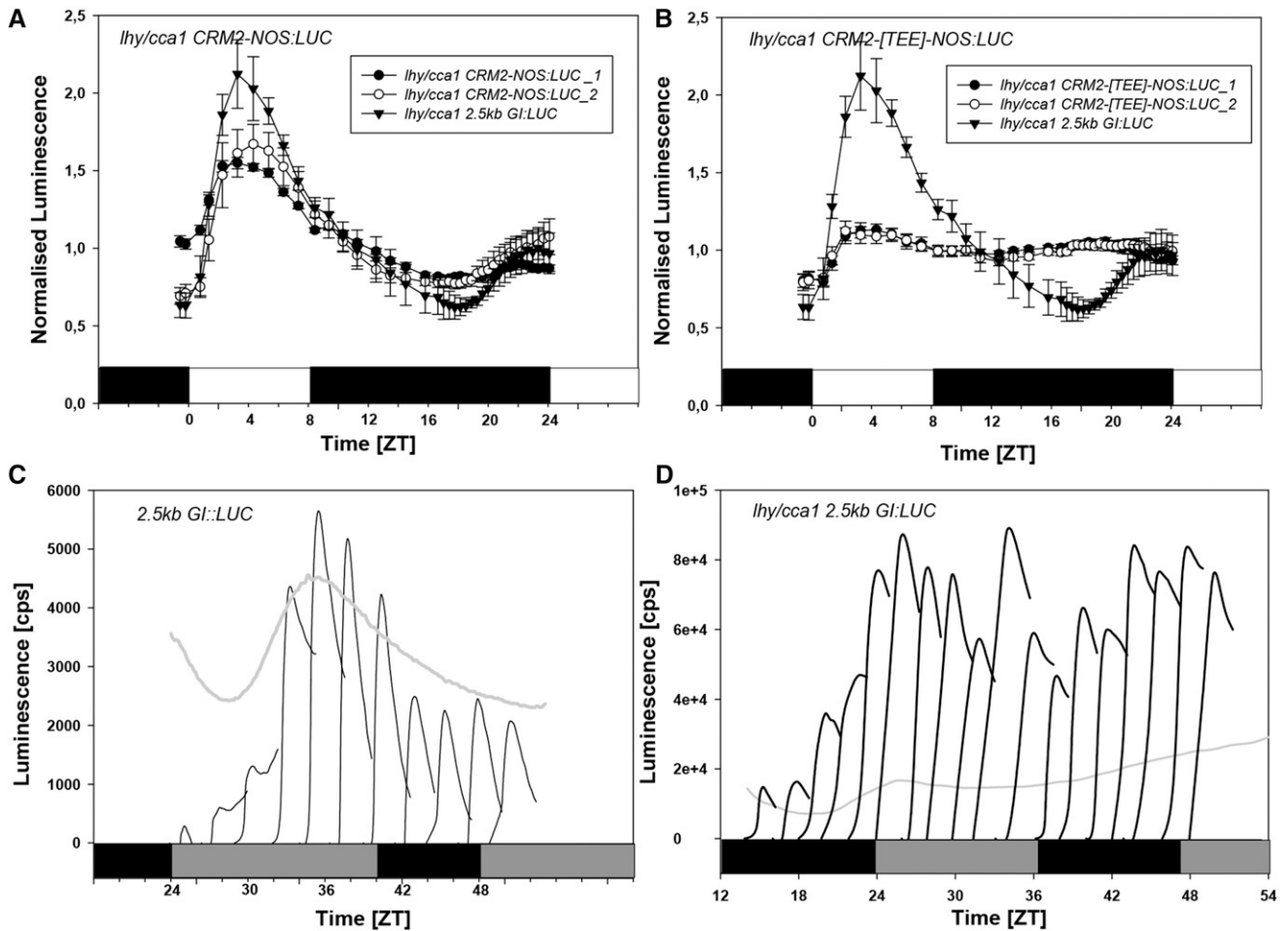


Figure 5. LHY/CCA1 and Other Factors Regulate the Expression of *Gl*.

(A) and (B) Diurnal expression patterns of different *Gl* promoter constructs in the *lhy/cca1* background. Plants were grown for 7 d under SDs (8 h light/16 h dark) and then measured under the same light regime for 24 h in a TopCount. Two independent homozygous lines of the CRM2-NOS:LUC (A) and CRM2-[TEE]-NOS:LUC (B) constructs compared with 2.5kb *Gl*::LUC are shown. All experiments represent the average of ~15 seedlings and were done at least twice with similar results. Black and white boxes indicate dark/light cycles. Error bars indicate se. ZT is Zeitgeber time and here represents time from lights on at dawn.

(C) and (D) Gating assay with *Gl*::LUC in *Ler* and *lhy/cca1*. Seedlings were grown for 7 d under LD conditions (16 h light/8 h dark, in the *Ler* background) or 10 d under LD conditions (12 h light/12 h dark, in the *lhy/cca1* background) and then transferred to the dark. Replicate samples were exposed to a white light pulse of 30 min every 2.5 h, and luciferase activity was monitored in a TopCount. Values represent the difference between measured mean values and the untreated dark sample (gray curve). Black and gray boxes indicate subjective days and nights. All curves represent the average of ~15 seedlings and were done twice with similar results. ZT is Zeitgeber time and here represents time from lights on at dawn.

functionally important, evolutionarily conserved regions containing densely packed transcription factor binding sites (Tautz, 2000). These enhancers are also often surrounded by long regions of DNA that are apparently not functionally important. The CRMs in *Gl* also coincide with regions of hypersensitivity to DNase 1 (Zhang et al., 2012), consistent with them containing binding sites for transcription factors and are therefore also similar to high occupancy target regions (Nègre et al., 2011).

CRM2 exhibits a modular structure and could be divided into three regions conferring distinct aspects of *Gl* promoter activity. CRM2-A showed strong light induction at all times of day and a diurnal pattern that featured a strong rise in expression at dawn. By contrast, CRM2-B conferred an evening peak of expression

that is delayed compared with that controlled by *Gl* promoter or CRM2, and its light induction is of low amplitude gated to the evening. CRM2-C shows no diurnal or light inducibility in expression but a constant level of expression at all times. Thus, the activity of CRM2 can be considered a composite of functionally separable subdomains.

Combinatorial Activity of ABREs and EEs in *Gl* Transcription

Three conserved ABREs were present in CRM2-A, suggesting that these could contribute to the strong light inducibility conferred by CRM2-A and the enhanced amplitude of expression

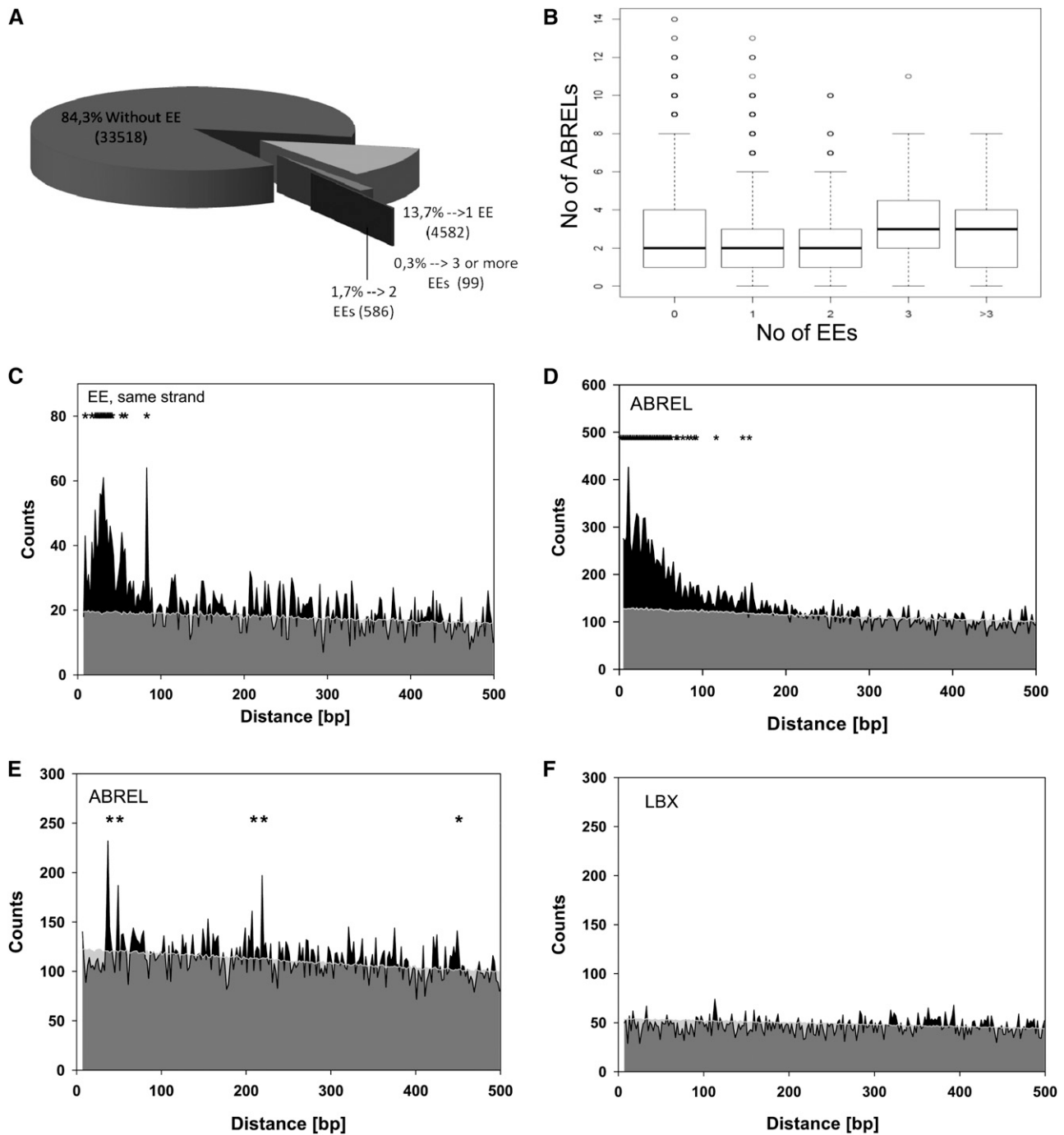


Figure 6. Genome-Wide Analysis of EE and ABREL Positions.

(A) Overview of EE-containing promoters. *A. thaliana* gene promoters that carry different numbers of EEs (AAAATATCT) within a 3-kb region upstream of the translational start site were identified. A subset of 99 promoters contains at least three EEs.

(B) Distribution of ABREs within EE-rich promoters. The number of ABREL motifs in genes with two or fewer EE motifs are significantly fewer compared with the group of genes with three or more EE motifs (*t* test, *P* value = 0.02998). Ninety-five percent of the data lie within the whiskers and the circles represent the outliers.

(C) and **(D)** EEs (containing the EE, the CBS, and the SEE) and ABREs were mapped within 3-kb regions upstream of start codons of all *A. thaliana* genes. A random background was generated and compared with the actual EE or ABREL distribution (details in the Methods section). *x* axis shows the distance of an EE/ABREL to the closest neighboring EE/ABREL. Gray shade indicates the calculated background distribution. Asterisks indicate statistically significant overrepresentations (*P* < 0.05).

CRM2-A contributes to CRM2. Similarly, three conserved EEs present in CRM2-B could confer the evening regulation of CRM2 and its light inducibility. Mutational analysis showed that EEs and ABREs act in combination to produce the characteristic evening pattern of transcription conferred by CRM2. Studying CRM2 in isolation from the remainder of the *G1* promoter allowed the contribution of these motifs to be more clearly defined by avoiding redundancy with other regions of the promoter.

Inactivation of the three EEs of CRM2 creates a pattern of expression in which LUC activity rises prematurely at dawn, remains high throughout the light period, and then falls in darkness. This pattern is likely based on light-induced transcription because in the gating assay, the CRM2-[TEE] construct also conferred light inducibility at all times of day, with a statistically significant enhancement at subjective dawn. By contrast, light induction of transcription from wild-type CRM2 was gated to the evening. This result suggests that light induction contributes to the main diurnal peak in *G1* transcription and that the EE elements contribute to restricting light inducibility to this time. Such a conclusion is consistent with *G1* mRNA tracking dusk by falling rapidly at this time under SDs (Fowler et al., 1999; Edwards et al., 2010). However, these results are in contrast to previous reports that light induction of *G1* is gated to the morning (Patiel et al., 2006). Although these conclusions are difficult to reconcile, gated expression in the evening is consistent with the major diurnal peak in *G1* expression and with our comparison of the behavior of CRM2-[TEE] and CRM2, which suggests that EEs are required to restrict light inducibility to the evening by inhibiting the response to light in the morning. This role of the EEs likely involves LHY and CCA1, which bind EEs and related sequences (Wang et al., 1997; Alabadi et al., 2001) and are expressed in the morning when *G1* transcription is inhibited (Schaffer et al., 1998; Wang and Tobin, 1998). Such a role of LHY CCA1 is supported by the observations that CCA1 binds directly to *G1* promoter in the morning (Lu et al., 2012) and that in *lhy cca1* double mutants, *G1* transcription also rises prematurely at dawn (Mizoguchi et al., 2002). The approach used here of testing the effect of mutagenesis of the EEs in vivo is complementary to that of analyzing expression of *G1* in *lhy cca1* double mutants and more precise because it precludes indirect effects of *lhy cca1* influencing *G1* expression through altering expression of other clock-regulated genes. Nevertheless, the EEs of *G1* are likely to have additional functions because in the *lhy cca1* double mutant, their inactivation in the CRM2-[TEE] construct reduced amplitude of LUC expression relative to CRM2. Thus, even in the *lhy cca1* double mutant, the EEs have a function in contributing to the amplitude of transcription conferred by CRM2. These results suggest that the EEs are required both to repress *G1* transcription in the morning and to increase amplitude in the evening. That EEs have more than one function was suggested previously based on analysis of mutant oligomerized EEs

(Harmer and Kay, 2005). Also, REV8 was shown to promote transcription of genes late in the day through binding to EEs (Rawat et al., 2011; Hsu et al., 2013). Our results indicate that the EEs in CRM2 of the *G1* promoter are used at least twice in the circadian cycle to repress expression in the morning and to activate expression later in the day, presumably through the activity of different MYB-like transcription factors such as REV8.

The ABREs are located in CRM2-A, which contributes to light induction and amplitude of expression of the *G1* promoter. The CRM2-[TA] construct showed a similar diurnal pattern to CRM2 but reached peak levels more slowly and the amplitude of its peak in expression was reduced. The ABREs are predicted to be recognized by bZIP transcription factors, several of which are implicated in light induction of gene transcription (Chattopadhyay et al., 1998; Jakoby et al., 2002). However, CRM2-[TA] exhibited a similar pattern of gated light induction to CRM2 with a similar amplitude, so if the ABREs are involved in light induction then in this assay, their role is redundant with other motifs present in CRM2. Nevertheless, analysis of the CRM2-[TAA]-NOS:LUC plants indicated that in the absence of the EEs, the ABREs are required for increased transcription at dawn in diurnal cycles and a weak response to light at subjective dawn in the gating assay. Thus, mutational analysis of ABREs in CRM2 suggests that their major contribution is to the amplitude of expression in the evening under diurnal cycles, and this might involve a response to light that is only revealed in the absence of the EEs.

Mutation of all six motifs within CRM2, the three EEs and the three ABREs, abolished any diurnal rhythm in expression. Therefore, the combined activity of these motifs is essential for the diurnal pattern conferred by CRM2. Similarly, the ABREs are required for the peak at dawn observed in CRM2-[TEE]-NOS:LUC plants, while the EEs are required for the evening peak observed in CRM2-[TAA]-NOS:LUC plants. Thus, the diurnal pattern of expression conferred by CRM2 involves the combined activities of EEs and ABREs. In the gating assay, mutation of all six motifs did not abolish light induction, although this was not gated. Therefore, CRM2 must contain additional motifs that can confer light inducibility even in the absence of the EEs and ABREs.

EEs and ABREs were also previously shown to act in a combinatorial manner to confer cold activation of transcription on two genes and to be overrepresented in the promoters of cold-induced genes (Mikkelsen and Thomashow, 2009). In this case, both had a positive effect on cold activation, although their effect on diurnal transcription was not directly measured. Nevertheless, they were proposed to confer cold-induced transcription soon after dawn because that is the phase of expression of the genes analyzed. Also, cold-induced genes were overrepresented among the 71 *A. thaliana* genes identified as containing three EEs in their promoters (Supplemental Figure 6A). Whether CRM2 confers cold induction has not yet been tested, although *G1* mRNA was found

Figure 6. (continued).

(E) and **(F)** The position of all ABREs **(E)** and LBXs **(F)** relative to EEs was mapped within 3-kb upstream regions within the *A. thaliana* genome. A random background was generated and compared with the actual distributions (details in the Methods section). *x* axis shows the distance of an ABREL or LBS element from the closest neighboring EE. Gray shade indicates the calculated background distribution. Asterisks are inserted above peaks representing statistically significant overrepresentations ($P < 0.01$).

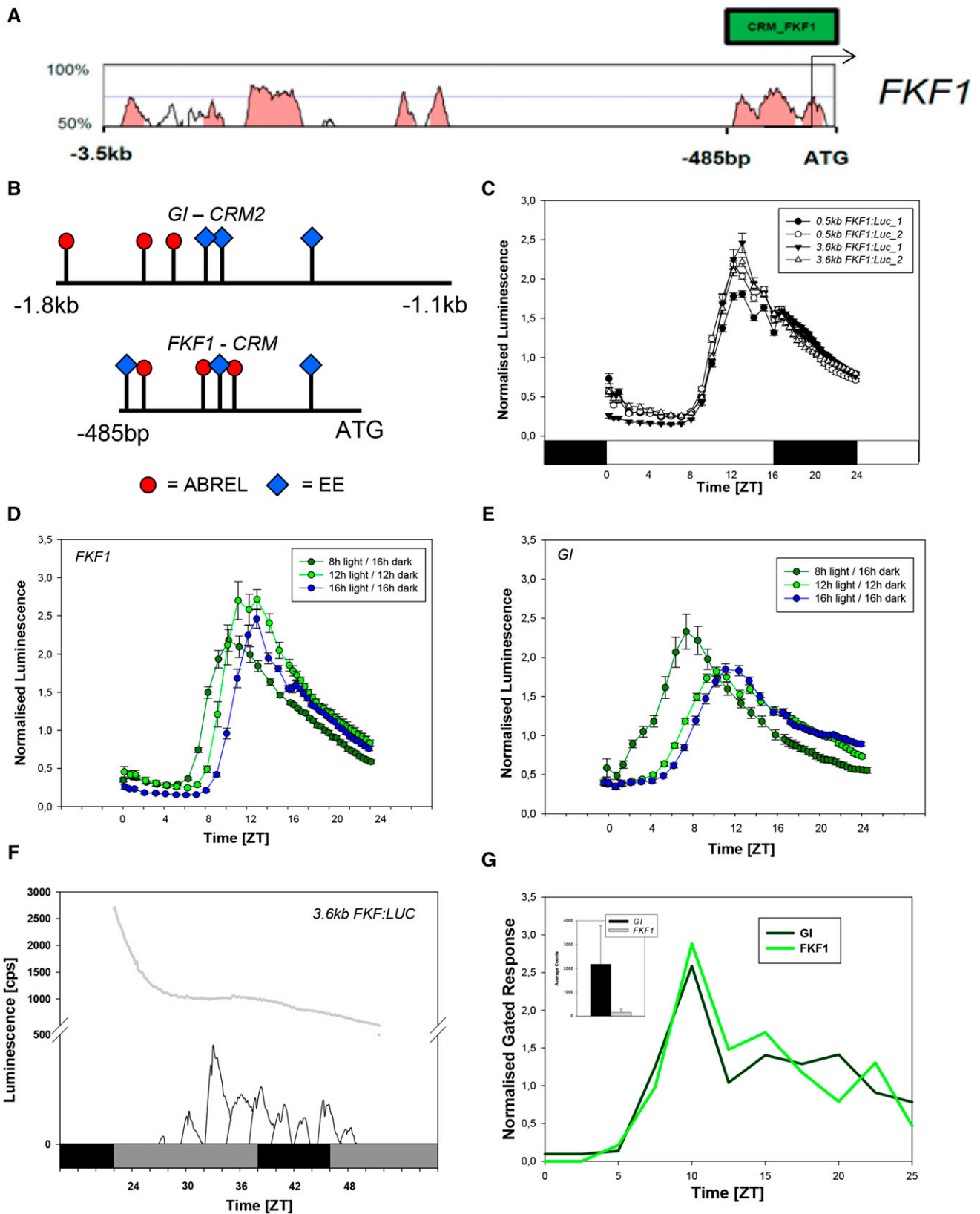


Figure 7. *GI* and *FKF1* Are Coregulated.

to be increased in abundance in cold (Fowler and Thomashow, 2002). In the case of *Gl*, the EEs and ABREs are interpreted differently from these cold-induced genes, conferring a high amplitude diurnal peak in expression in the evening.

Genome-Wide Analysis of Promoters Containing EEs and ABREs

CRM2 contains three EEs and three ABREs, and whole-genome analysis demonstrated that clustering of these motifs is commonly observed. Potentially this could be a form of genetic redundancy to create robustness of the effect of these motifs on *Gl* transcription (Frankel et al., 2010). However, our observation that particular distances between these elements are strongly statistically overrepresented suggests that they act combinatorially in the regulation of transcription, perhaps by allowing interaction between protein complexes bound to these motifs. Notably, the frequency of ABREs relative to EEs were represented as two pairs of high amplitude peaks, and the distance between these pairs was approximately equivalent to one nucleosome (146 bp), suggesting that protein complexes bound to the EE and the ABREs might interact on the same side of the DNA.

Comparison of orthologs efficiently identified CRMs in the *Gl* promoter; however, these could not be recognized by comparing coexpressed genes in *A. thaliana*. Therefore, in this case, phylogenetic shadowing was more effective in identifying functionally important regions than comparison of promoters of coregulated genes. Nevertheless, analysis of *FKF1*, which is transcriptionally coregulated with *Gl* allowing interaction of their protein products (Sawa et al., 2007), demonstrated that *FKF1* promoter also contains a CRM harboring EEs, ABREs, and other similar motifs to *Gl* CRM2. However, *FKF1* CRM could not be identified by direct alignment of the *FKF1* and *Gl* promoters. These CRMs might be an example of convergent evolution

where the CRM in each gene arose independently. *ELF4* represents another example of convergent evolution to confer an evening peak in expression similar to *Gl*, but the motifs conferring light induction as well as the number and spacing of sites in the *ELF4* promoter are different from those in the *FKF1* or *Gl* promoters (Li et al., 2011). Alternatively, the CRMs of *Gl* and *FKF1* might be derived from the same ancestral CRM, but the *FKF1* CRM and *Gl* CRM2 might have diverged over a longer evolutionary time than those of the *Gl* orthologs. Such divergence could have occurred through mutation and stabilizing selection as described for enhancers in fruit fly (*Drosophila melanogaster*) (Ludwig et al., 2000).

Definition of a Circadian Transcriptional Code

A major current aim of biology is to be able to predict the expression patterns of genes from their regulatory sequences, as it is possible to predict the protein sequence from the open reading frame (Yáñez-Cuna et al., 2013). Combining evolutionary comparisons, whole-genome information, and experimental dissection of promoters, as performed here for *Gl*, is a powerful methodology for defining combinations of motifs that confer particular transcriptional patterns. In plants, diurnal regulation of environmental responses is strongly controlled at the transcriptional level. Our work on *Gl* defines combinations of motifs that confer temporally regulated light induction, and together with other recent applications of related methods contributes to the definition of a plant circadian transcriptional code (Harmer and Kay, 2005; Michael et al., 2008; Spensley et al., 2009; Helfer et al., 2011; Li et al., 2011). In addition, the observation that light inducibility and circadian control are contributed by short, defined, separable modules fuels the design of synthetic promoters constructed by combining such fragments from different promoters and thereby conferring complex transcriptional patterns (Rushton et al., 2002; Liu et al., 2011).

Figure 7. (continued).

(A) Phylogenetic shadowing analysis of the *FKF1* promoter. The *FKF1* promoter was compared between *A. thaliana* (base genome) and *A. alpina*. Pink color indicates regions with more than 70% conservation based on a 100-bp sliding window. A highly conserved region is highlighted with a green box and was named CRM_FKF1.

(B) Comparison of highly conserved promoter modules between *Gl* and *FKF1*. Both contain three highly conserved EEs and ABREs.

(C) Diurnal expression patterns of different *FKF1* promoter fusions. A 483-bp fragment (CRM_FKF1) of the *FKF1* promoter was fused to Luciferase (resulting in 0.5-kb *FKF1:LUC*) and compared with the 3.6-kb *FKF1:LUC* fusion. Around 20 seedlings of two independent homozygous lines were grown for 7 d under LDs (16 h light/8 h dark) and then measured for 24 h in a TopCount. Error bars indicate \pm SE. ZT is Zeitgeber time and here represents time from lights on at dawn.

(D) and **(E)** *Gl* and *FKF1* expression in *A. thaliana* tracks dusk under different photoperiods. Around 20 seedlings of 3.6kb *FKF1:LUC* **(D)** or 2.5kb *Gl:LUC* **(E)** were entrained for 7 d under three different photoperiods (16 h light/8 h dark, 12 h light/12 h dark, and 8 h light/16 h dark) and measured under the respective condition in a TopCount for 24 h. Error bars indicate \pm SE. ZT is Zeitgeber time and here represents time from lights on at dawn.

(F) Gating assay with the 3.6kb *FKF1:LUC* construct. Batches of three to six seedlings were grown for 7 d under LDs (16 h light/8 h dark) and then transferred to the dark. Replicate samples were exposed to a white light pulse of 30 min every 2.5 h and luciferase activity was monitored in a TopCount. Values represent the difference between measured mean values and the untreated dark sample (gray curve). Black and gray boxes indicate subjective days and nights, respectively. Experiment was done twice with similar results. ZT is Zeitgeber time and here represents time from lights on at dawn.

(G) Gated response to a light pulse shown by *Gl:LUC* and *FKF1:LUC*. Peak data points from **(F)** were normalized to their 25 h mean expression and plotted to a respective 2.5kb *Gl:LUC* control. Small panel shows absolute luciferase expression averaged over 24 h from the same experiment. Error bars represent \pm SE. The data are based on an average of 6 to 12 seedlings for each time point. For *FKF1:LUC*, two independently replicated experiments were performed, whereas for *Gl:LUC*, five independently replicated experiments were performed. ZT is Zeitgeber time and here represents time from lights on at dawn.

METHODS

Phylogenetic Analysis

GI promoter sequences were identified by BLAST from different sources. Sequences from *Arabidopsis thaliana* and *Arabidopsis lyrata* were obtained from TAIR (<http://www.arabidopsis.org>) or Phytozome (<http://www.phytozome.net>). The *Capsella rubella* sequence was assembled from raw sequence reads available at NCBI. *Arabis alpina* sequences were identified from the *A. alpina* sequencing project at MIPZ Cologne. The *Brassica rapa* sequence was obtained from the Brassica Genome Project (brassica.info). Sequences from *Turritis glabra*, *Diplotaxis erucoides*, and *Sinapis alba* were amplified from genomic DNA using degenerated primers.

All pairwise alignments were done with Shuffle LAGAN (Brudno et al., 2003) using default settings. VISTA Plots were made with the VISTA Browser (Mayor et al., 2000), with a calculation window of 100 bp and a consensus identity of 70%. Multiple sequence alignments were done with DIALIGN (Morgenstern, 2004) and ClustalW (Larkin et al., 2007) using default parameters. Conserved *cis*-regulatory elements were visualized with WebLogo (Crooks et al., 2004). Conserved blocks (CRMs) were defined based on conservation between *A. thaliana* and *A. alpina*. A region containing a stretch of 100 bp or longer that showed at least 70% conservation was considered to be a CRM. Gene Ontology term enrichment analysis was performed with FatiGO from the Babelomics server (Medina et al., 2010).

Plasmid Constructions

Promoter fragments were amplified with primer pairs containing Gateway recombination sites using a proofreading polymerase. Mutagenic primers were used for site directed mutagenesis. All fragments with the respective primer combinations are described in the primer table (Supplemental Table 5).

Amplified and mutated promoter fragments were recombined with entry vector pDONR 207 (Invitrogen) in a BP reaction according to the user's manual. To create LUC fusions, plasmids were recombined with the binary vectors pGWLuc (GenBank: AM295157.1) or pGWLuc_nos (pGWLuc with an additional 105-bp nos minimal promoter fragment [Puente et al., 1996] inserted into the *HindIII* restriction site upstream of the luciferase fragment of pGWLuc) in a LR reaction (Invitrogen) according to the user's manual.

Binary vectors were introduced into *Agrobacterium tumefaciens* strain GV3101 containing the helper plasmid pSOUP via electroporation. Transformed bacteria were selected on Luria-Bertani medium containing the appropriate antibiotics and presence of the plasmid was verified by colony PCR.

Plant Genotypes and Generation of Transgenic Plants

The 35S:*GI A. thaliana* plants were derived by backcrossing the *Ler* lines described by Mizoguchi et al. (2005) six times into Columbia. The *gi-2* mutant was described by Rédei (1962). Binary plasmids were transformed into *A. thaliana* plants by the floral dip method (Clough and Bent, 1998) or by a simplified floral dip method (Davis et al., 2009).

T1 plants carrying the plasmid were selected on soil based on their resistance to Basta (Bayer). T2 were identified based on segregation analysis on half-strength Murashige and Skoog medium supplemented with 1% sucrose and containing 12 $\mu\text{g}/\text{mL}$ phosphinotricin. Lines that showed a segregation ratio between 1:2 and 1:4 (based on 40 to 60 seedlings) were considered for further analysis. Out of 5 to 10 analyzed T2 lines, at least two independent homozygous lines (two representative lines were chosen, based on all T2 lines) were established in the T3 generation for all described constructs.

Luminescence Measurements and Statistical Analysis

Seeds were surface sterilized, stratified for 3d at 4°C, and grown for 7d 10 d for the gating experiments with *lhy cca1* plants) under cool white light (~70 μE) on half-strength Murashige and Skoog medium supplemented with 1% sucrose. Seedlings were then manually transferred into 96-well opaque microtiter plates with each 200 μL of half-strength Murashige and Skoog medium supplemented with 1% sucrose. Twenty microliters of 1 mM D-luciferin was added for each plant, and luminescence was measured from the next day in a TopCount scintillation and luminescence counter (Perkin-Elmer). For diurnal measurements, each plate was manually transferred every 30 min or every 60 min from cool white light (~70 μE) to the TopCount to measure luminescence.

Absolute values were calculated by averaging 8 to 12 seedlings from independent homozygous lines over a 24-h light-dark cycle. Normalization was done by relating each single time point of 8 to 12 seedlings to the aforementioned 24 h mean of each independent line.

For all gating experiments, replicate plates were kept in the dark and exposed to a white light pulse of 30 min (~70 μE) and then transferred back into darkness. For all gating measurements as shown in Figures 4 and 5, expression of 12 seedlings of an established homozygous line was measured for 2 to 3 h and absolute data are shown with the subtracted values of the untreated sample. Every experiment was done at least twice independently with comparable results. Additional gating experiments with three independent heterozygous T2 lines of constructs with point mutations were done with three to six seedlings in the same experimental setup.

To test for a gated response, a statistical test was developed that compared the deviation of induction at each time point from the mean of all time points over the 25 h of measurement. Therefore, the peak values for each induction curve were calculated and the nonlinear relationships between response variables (the peak induction) and the predictor variable (the respective time) were modeled using a generalized additive model (GAM) with smoothing splines (Hastie and Tibshirani, 1990). The fitting of the model was tested in an ANOVA setting. All analyses were done using R (R Core Team, 2014) and the GAM package. The commands used for the statistical test are available in Supplemental Data Set 1.

Co-Occurrence of *cis*-Regulatory Elements

The TAIR9 data set containing 3000 bp upstream of each translation start was downloaded from the TAIR website (www.arabidopsis.org) and used for all calculations. Using this criterion, 99 genes were identified that contained three EEs within 3 kb of the translational start site. Manual annotation reduced this number to 71 because in some cases EEs were within 3 kb of the translational start site of more than one gene and manual inspection allowed the additional genes to be removed. Matches for each motif were identified on both strands of all promoter regions. To compare two groups of motifs, the absolute distances between all valid pairs in each promoter were calculated. Furthermore, the total number of motifs for each motif and promoter were counted.

A random background was generated by bootstrapping. The number of motifs for each group was sampled from the true distribution of counts, and then the given number of positions was sampled from the true distribution of positions. If the strand was considered, the sampled positions were randomized to either strand with equal probability. This process was repeated 1,000,000 or 2,000,000 times, depending on whether the strand was considered or not. For each random promoter, the absolute distances between all valid pairs were calculated.

For distances below 500 bp, counts for bins with the size of two base pairs were generated for both the true distribution and the randomized background. For each bin, the probability is to find at least as many motifs at the given distances as found in the true distribution. This was done with a binomial distribution. The probability of success in a single draw equaled the percentage of distances in the current bin in the randomized

background. The number of draws was the total number of distances in the true distribution. The calculations were made in the statistical software R and each probability was adjusted with the internal method `p.adjust`. Values below 0.05 (EE comparison) or 0.01 (bHLH-bZIP-LBS comparison) were considered to be significant. Diurnal and circadian microarray data sets (Mockler et al., 2007) were analyzed using Phaser.

Flowering Time Measurements

All *A. thaliana* plants were in the Columbia background. Plants were grown at 20°C under LD (16 h light/8 h dark) or SD (8 h light/16 h dark) conditions after stratification at 4°C for 3 d. Light was provided by cool-white fluorescent tubes (~70 µE). Flowering time was measured by scoring the number of rosette and cauline leaves on the main stem of 8 to 12 individuals. Experiments were done three times with similar results and one representative data set is shown.

Accession Numbers

Sequence data from this article can be found in GenBank/EMBL under accession numbers KM497444 to KM497445 (*FKF1* promoters) and KM497446 to KM497453 (*Gl* promoters).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Absolute Gene Expression Patterns in All Lines Used in This Study.

Supplemental Figure 2. Diurnal Expression Patterns of *Gl* Promoter Constructs with Point Mutations in Multiple Independent Lines.

Supplemental Figure 3. Absolute Expression Levels of *Gl* Promoter Constructs with Point Mutations in Multiple Independent Lines.

Supplemental Figure 4. Statistical Analysis of Light Gating in Different Constructs with *cis*-Element Mutations.

Supplemental Figure 5. Genome-Wide Analysis of EE Positions.

Supplemental Figure 6. Promoters with Multiple EEs and ABREs Confer Evening Expression and Cold Responsiveness.

Supplemental Figure 7. Highly Conserved Modules within the *Gl* and *FKF1* Promoters Contain Multiple EEs and ABREs.

Supplemental Table 1. Positions and Lengths of Three CRMs and CRM2 Subfragments in the *Gl* Promoter as Described in the Article.

Supplemental Table 2. Conserved *cis*-Regulatory Elements in the *Gl* Promoter of *A. thaliana*.

Supplemental Table 3. “Gating Scores” of All Constructs Used in This Study.

Supplemental Table 4. Seventy-One *A. thaliana* Genes Have at Least Three EEs in Their Promoters.

Supplemental Table 5. Primer Table.

Supplemental Data Set 1. Commands for Statistical Test of Light-Gating Data.

ACKNOWLEDGMENTS

We thank Amaury de Montaigu for critical reading of the article and advice, the MPIPZ gardeners Monika Bürgel and Heike Eichler for excellent greenhouse service, as well as Lisa Hallau and Julia Werres for excellent support with luciferase experiments. We thank Hailong An

for the original 2.5-kb *Gl:LUC* construct. Hugo Konijn kindly provided the *Gl* cDNA vector. This work was funded by an International Max Planck Research School studentship to M.C.B. and a core grant from the Max Planck Society to G.C.

AUTHOR CONTRIBUTIONS

M.C.B. and R.T. performed research. M.C.B., F.C., R.T., and G.C. designed research. M.C.B., K.N., F.C., M.H., S.S., J.R.K., and I.B. contributed new analytic, computational, or experimental tools. M.C.B. and K.N. analyzed data. M.C.B. and G.C. wrote the article.

Received July 7, 2014; revised September 11, 2014; accepted October 12, 2014; published October 31, 2014.

REFERENCES

- Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G., and Turck, F. (2010). *cis*-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. *Plant Cell* **22**: 1425–1440.
- Alabadi, D., Yanovsky, M.J., Más, P., Harmer, S.L., and Kay, S.A. (2002). Critical role for CCA1 and LHY in maintaining circadian rhythmicity in *Arabidopsis*. *Curr. Biol.* **12**: 757–761.
- Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Más, P., and Kay, S.A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* **293**: 880–883.
- Araki, T., and Komeda, Y. (1993). Analysis of the role of the late-flowering locus, *Gl*, in the flowering of *Arabidopsis thaliana*. *Plant J.* **3**: 231–239.
- Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L., and Rubin, E.M. (2003). Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**: 1391–1394.
- Budno, M., Malde, S., Poliakov, A., Do, C.B., Couronne, O., Dubchak, I., and Batzoglou, S. (2003). Glocal alignment: finding rearrangements during alignment. *Bioinformatics* **19** (suppl. 1): i54–i62.
- Busk, P.K., and Pagès, M. (1998). Regulation of abscisic acid-induced transcription. *Plant Mol. Biol.* **37**: 425–435.
- Cao, S., Ye, M., and Jiang, S. (2005). Involvement of GIGANTEA gene in the regulation of the cold stress response in Arabidopsis. *Plant Cell Rep.* **24**: 683–690.
- Chattopadhyay, S., Ang, L.H., Puente, P., Deng, X.W., and Wei, N. (1998). Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *Plant Cell* **10**: 673–683.
- Choi, H., Hong, J., Ha, J., Kang, J., and Kim, S.Y. (2000). ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.* **275**: 1723–1730.
- Cliften, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish, W.R., Waterston, R.H., and Johnston, M. (2001). Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res.* **11**: 1175–1186.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**: 735–743.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**: 1188–1190.

- Dalchau, N., Baek, S.J., Briggs, H.M., Robertson, F.C., Dodd, A.N., Gardner, M.J., Stancombe, M.A., Haydon, M.J., Stan, G.B., Gonçalves, J.M., and Webb, A.A.R.** (2011). The circadian oscillator gene *GIGANTEA* mediates a long-term response of the *Arabidopsis thaliana* circadian clock to sucrose. *Proc. Natl. Acad. Sci. USA* **108**: 5104–5109.
- Davis, A.M., Hall, A., Millar, A.J., Darrah, C., and Davis, S.J.** (2009). Protocol: Streamlined sub-protocols for floral-dip transformation and selection of transformants in *Arabidopsis thaliana*. *Plant Methods* **5**: 3.
- de Montaigu, A., Tóth, R., and Coupland, G.** (2010). Plant development goes like clockwork. *Trends Genet.* **26**: 296–306.
- Ding, Z., Millar, A.J., Davis, A.M., and Davis, S.J.** (2007). TIME FOR COFFEE encodes a nuclear regulator in the *Arabidopsis thaliana* circadian clock. *Plant Cell* **19**: 1522–1536.
- Dixon, L.E., Knox, K., Kozma-Bognar, L., Southern, M.M., Pokhilko, A., and Millar, A.J.** (2011). Temporal repression of core circadian genes is mediated through EARLY FLOWERING 3 in *Arabidopsis*. *Curr. Biol.* **21**: 120–125.
- Dong, M.A., Farré, E.M., and Thomashow, M.F.** (2011). Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **108**: 7241–7246.
- Doyle, M.R., Davis, S.J., Bastow, R.M., McWatters, H.G., Kozma-Bognár, L., Nagy, F., Millar, A.J., and Amasino, R.M.** (2002). The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature* **419**: 74–77.
- Dunford, R.P., Griffiths, S., Christodoulou, V., and Laurie, D.A.** (2005). Characterisation of a barley (*Hordeum vulgare* L.) homologue of the *Arabidopsis* flowering time regulator *GIGANTEA*. *Theor. Appl. Genet.* **110**: 925–931.
- Edwards, K.D., Akman, O.E., Knox, K., Lumsden, P.J., Thomson, A.W., Brown, P.E., Pokhilko, A., Kozma-Bognar, L., Nagy, F., Rand, D.A., and Millar, A.J.** (2010). Quantitative analysis of regulatory flexibility under changing environmental conditions. *Mol. Syst. Biol.* **6**: 424.
- Eimert, K., Wang, S.-M., Lue, W.I., and Chen, J.** (1995). Monogenic recessive mutations causing both late floral initiation and excess starch accumulation in *Arabidopsis*. *Plant Cell* **7**: 1703–1712.
- Fornara, F., Panigrahi, K.C., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J.A., and Coupland, G.** (2009). *Arabidopsis* DOF transcription factors act redundantly to reduce *CONSTANS* expression and are essential for a photoperiodic flowering response. *Dev. Cell* **17**: 75–86.
- Fowler, S., and Thomashow, M.F.** (2002). *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* **14**: 1675–1690.
- Fowler, S., Lee, K., Onouchi, H., Samach, A., Richardson, K., Morris, B., Coupland, G., and Putterill, J.** (1999). *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *EMBO J.* **18**: 4679–4688.
- Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L.** (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**: 490–493.
- Harmer, S.L.** (2009). The circadian system in higher plants. *Annu. Rev. Plant Biol.* **60**: 357–377.
- Harmer, S.L., and Kay, S.A.** (2005). Positive and negative factors confer phase-specific circadian regulation of transcription in *Arabidopsis*. *Plant Cell* **17**: 1926–1940.
- Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A.** (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**: 2110–2113.
- Hastie, T., and Tibshirani, R.** (1990). *Generalized Additive Models*. (London: Chapman and Hall).
- Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K.** (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**: 719–722.
- Hecht, V., Knowles, C.L., Vander Schoor, J.K., Liew, L.C., Jones, S.E., Lambert, M.J.M., and Weller, J.L.** (2007). Pea *LATE BLOOMER1* is a *GIGANTEA* ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant Physiol.* **144**: 648–661.
- Helper, A., Nusinow, D.A., Chow, B.Y., Gehrke, A.R., Bulyk, M.L., and Kay, S.A.** (2011). *LUX ARRHYTHMO* encodes a nighttime repressor of circadian gene expression in the *Arabidopsis* core clock. *Curr. Biol.* **21**: 126–133.
- Herrero, E., et al.** (2012). *EARLY FLOWERING4* recruitment of *EARLY FLOWERING3* in the nucleus sustains the *Arabidopsis* circadian clock. *Plant Cell* **24**: 428–443.
- Hicks, K.A., Albertson, T.M., and Wagner, D.R.** (2001). *EARLY FLOWERING3* encodes a novel protein that regulates circadian clock function and flowering in *Arabidopsis*. *Plant Cell* **13**: 1281–1292.
- Hong, R.L., Hamaguchi, L., Busch, M.A., and Weigel, D.** (2003). Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* **15**: 1296–1309.
- Hsu, P.Y., Devisetty, U.K., and Harmer, S.L.** (2013). Accurate time-keeping is controlled by a cycling activator in *Arabidopsis*. *eLife* **2**: e00473.
- Hu, T.T., et al.** (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**: 476–481.
- Huq, E., Tepperman, J.M., and Quail, P.H.** (2000). *GIGANTEA* is a nuclear protein involved in phytochrome signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **97**: 9789–9794.
- Imaizumi, T., Tran, H.G., Swartz, T.E., Briggs, W.R., and Kay, S.A.** (2003). *FKF1* is essential for photoperiodic-specific light signalling in *Arabidopsis*. *Nature* **426**: 302–306.
- Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A., and Kay, S.A.** (2005). *FKF1* F-box protein mediates cyclic degradation of a repressor of *CONSTANS* in *Arabidopsis*. *Science* **309**: 293–297.
- Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F., bZIP Research Group** (2002). *bZIP* transcription factors in *Arabidopsis*. *Trends Plant Sci.* **7**: 106–111.
- Kim, W.Y., Fujiwara, S., Suh, S.S., Kim, J., Kim, Y., Han, L., David, K., Putterill, J., Nam, H.G., and Somers, D.E.** (2007). *ZEITLUPE* is a circadian photoreceptor stabilized by *GIGANTEA* in blue light. *Nature* **449**: 356–360.
- Kurepa, J., Smalle, J., Van Montagu, M., and Inzé, D.** (1998). Oxidative stress tolerance and longevity in *Arabidopsis*: the late-flowering mutant *gigantea* is tolerant to paraquat. *Plant J.* **14**: 759–764.
- Larkin, M.A., et al.** (2007). ClustalW and ClustalX version 2.0. *Bioinformatics* **23**: 2947–2948.
- Lee, J.Y., Baum, S.F., Alvarez, J., Patel, A., Chitwood, D.H., and Bowman, J.L.** (2005). Activation of *CRABS CLAW* in the nectaries and carpels of *Arabidopsis*. *Plant Cell* **17**: 25–36.
- Li, G., et al.** (2011). Coordinated transcriptional regulation underlying the circadian clock in *Arabidopsis*. *Nat. Cell Biol.* **13**: 616–622.
- Li, J., Terzaghi, W., and Deng, X.W.** (2012). Genomic basis for light control of plant development. *Protein Cell* **3**: 106–116.
- Lindblad-Toh, K., et al., Broad Institute Sequencing Platform and Whole Genome Assembly Team, Baylor College of Medicine**

- Human Genome Sequencing Center Sequencing Team, Genome Institute at Washington University** (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**: 476–482.
- Liu, W., Mazarei, M., Rudis, M.R., Fethe, M.H., and Stewart, C.N., Jr.** (2011). Rapid in vivo analysis of synthetic promoters for plant pathogen phyto-sensing. *BMC Biotechnol.* **11**: 108.
- Lu, S.X., Webb, C.J., Knowles, S.M., Kim, S.H., Wang, Z., and Tobin, E.M.** (2012). CCA1 and ELF3 Interact in the control of hypocotyl length and flowering time in *Arabidopsis*. *Plant Physiol.* **158**: 1079–1088.
- Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M.** (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**: 564–567.
- Makino, S., Kiba, T., Imamura, A., Hanaki, N., Nakamura, A., Suzuki, T., Taniguchi, M., Ueguchi, C., Sugiyama, T., and Mizuno, T.** (2000). Genes encoding pseudo-response regulators: insight into His-to-Asp phosphorelay and circadian rhythm in *Arabidopsis thaliana*. *Plant Cell Physiol.* **41**: 791–803.
- Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I.** (2000). VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**: 1046–1047.
- Medina, I., et al.** (2010). Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res.* **38**: W210–3.
- Michael, T.P., et al.** (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.* **4**: e14.
- Mikkelsen, M.D., and Thomashow, M.F.** (2009). A role for circadian evening elements in cold-regulated gene expression in *Arabidopsis*. *Plant J.* **60**: 328–339.
- Millar, A.J., and Kay, S.A.** (1996). Integration of circadian and phototransduction pathways in the network controlling CAB gene transcription in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **93**: 15491–15496.
- Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carré, I.A., and Coupland, G.** (2002). *LHY* and *CCA1* are partially redundant genes required to maintain circadian rhythms in *Arabidopsis*. *Dev. Cell* **2**: 629–641.
- Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., and Coupland, G.** (2005). Distinct roles of *GIGANTEA* in promoting flowering and regulating circadian rhythms in *Arabidopsis*. *Plant Cell* **17**: 2255–2270.
- Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C., Kay, S.A., and Chory, J.** (2007). The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* **72**: 353–363.
- Morgenstern, B.** (2004). DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res.* **32**: W33–6.
- Nagel, D.H., and Kay, S.A.** (2012). Complexity in the wiring and regulation of plant circadian networks. *Curr. Biol.* **22**: R648–R657.
- Nègre, N., et al.** (2011). A cis-regulatory map of the *Drosophila* genome. *Nature* **471**: 527–531.
- Nelson, D.C., Lasswell, J., Rogg, L.E., Cohen, M.A., and Bartel, B.** (2000). *FKF1*, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*. *Cell* **101**: 331–340.
- Nusinow, D.A., Helfer, A., Hamilton, E.E., King, J.J., Imaizumi, T., Schultz, T.F., Farré, E.M., and Kay, S.A.** (2011). The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* **475**: 398–402.
- Onai, K., Okamoto, K., Nishimoto, H., Morioka, C., Hirano, M., Kami-Ike, N., and Ishiura, M.** (2004). Large-scale screening of *Arabidopsis* circadian clock mutants by a high-throughput real-time bioluminescence monitoring system. *Plant J.* **40**: 1–11.
- Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M.** (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**: 92–94.
- Paltiel, J., Amin, R., Gover, A., Ori, N., and Samach, A.** (2006). Novel roles for *GIGANTEA* revealed under environmental conditions that modify its expression in *Arabidopsis* and *Medicago truncatula*. *Planta* **224**: 1255–1268.
- Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A., and Nam, H.G.** (1999). Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* *GIGANTEA* gene. *Science* **285**: 1579–1582.
- Pokhilko, A., Fernández, A.P., Edwards, K.D., Southern, M.M., Halliday, K.J., and Millar, A.J.** (2012). The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol. Syst. Biol.* **8**: 574.
- Puente, P., Wei, N., and Deng, X.W.** (1996). Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in *Arabidopsis*. *EMBO J.* **15**: 3732–3743.
- R Core Team** (2014). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).
- Rawat, R., Takahashi, N., Hsu, P.Y., Jones, M.A., Schwartz, J., Salemi, M.R., Phinney, B.S., and Harmer, S.L.** (2011). *REVEILLE8* and *PSEUDO-RESPONSE REGULATOR5* form a negative feedback loop within the *Arabidopsis* circadian clock. *PLoS Genet.* **7**: e1001350.
- Rédei, G.P.** (1962). Supervital mutants of *Arabidopsis*. *Genetics* **47**: 443–460.
- Rushton, P.J., Reinstädler, A., Lipka, V., Lippok, B., and Somssich, I.E.** (2002). Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. *Plant Cell* **14**: 749–762.
- Sawa, M., and Kay, S.A.** (2011). *GIGANTEA* directly activates Flowering Locus T in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **108**: 11698–11703.
- Sawa, M., Nusinow, D.A., Kay, S.A., and Imaizumi, T.** (2007). *FKF1* and *GIGANTEA* complex formation is required for day-length measurement in *Arabidopsis*. *Science* **318**: 261–265.
- Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carré, I.A., and Coupland, G.** (1998). The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**: 1219–1229.
- Song, Y.H., Smith, R.W., To, B.J., Millar, A.J., and Imaizumi, T.** (2012). *FKF1* conveys timing information for *CONSTANS* stabilization in photoperiodic flowering. *Science* **336**: 1045–1049.
- Spensley, M., Kim, J.Y., Picot, E., Reid, J., Ott, S., Helliwell, C., and Carré, I.A.** (2009). Evolutionarily conserved regulatory motifs in the promoter of the *Arabidopsis* clock gene *LATE ELONGATED HYPOCOTYL*. *Plant Cell* **21**: 2606–2623.
- Stark, A., et al., Harvard FlyBase curators, Berkeley Drosophila Genome Project** (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**: 219–232.
- Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Más, P., Panda, S., Kreps, J.A., and Kay, S.A.** (2000). Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* **289**: 768–771.

- Suzuki, M., Ketterling, M.G., and McCarty, D.R.** (2005). Quantitative statistical analysis of cis-regulatory sequences in ABA/VP1- and CBF/DREB1-regulated genes of *Arabidopsis*. *Plant Physiol.* **139**: 437–447.
- Tautz, D.** (2000). Evolution of transcriptional regulation. *Curr. Opin. Genet. Dev.* **10**: 575–579.
- Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K., and Yamaguchi-Shinozaki, K.** (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc. Natl. Acad. Sci. USA* **97**: 11632–11637.
- Wang, X., et al., Brassica rapa Genome Sequencing Project Consortium** (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**: 1035–1039.
- Wang, Z.Y., and Tobin, E.M.** (1998). Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**: 1207–1217.
- Wang, Z.Y., Kenigsbuch, D., Sun, L., Harel, E., Ong, M.S., and Tobin, E.M.** (1997). A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis* Lhcb gene. *Plant Cell* **9**: 491–507.
- Yamaguchi, N., Wu, M.F., Winter, C.M., Berns, M.C., Nole-Wilson, S., Yamaguchi, A., Coupland, G., Krizek, B.A., and Wagner, D.** (2013). A molecular framework for auxin-mediated initiation of flower primordia. *Dev. Cell* **24**: 271–282.
- Yáñez-Cuna, J.O., Kvon, E.Z., and Stark, A.** (2013). Deciphering the transcriptional cis-regulatory code. *Trends Genet.* **29**: 11–22.
- Zhang, W., Zhang, T., Wu, Y., and Jiang, J.** (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* **24**: 2719–2731.
- Zhao, X.Y., Liu, M.S., Li, J.R., Guan, C.M., and Zhang, X.S.** (2005). The wheat TaGl1, involved in photoperiodic flowering, encodes an *Arabidopsis* Gl ortholog. *Plant Mol. Biol.* **58**: 53–64.