

UC Riverside

UC Riverside Previously Published Works

Title

Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis

Permalink

<https://escholarship.org/uc/item/1w6583vk>

Journal

Proceedings of the National Academy of Sciences of the United States of America, 112(34)

ISSN

0027-8424

Authors

Nagel, Dawn H
Doherty, Colleen J
Pruneda-Paz, Jose L
et al.

Publication Date

2015-08-25

DOI

10.1073/pnas.1513609112

Peer reviewed

Genome-wide identification of CCA1 targets uncovers an expanded clock network in *Arabidopsis*

Dawn H. Nagel^{a,1}, Colleen J. Doherty^{a,b,1,2}, Jose L. Pruneda-Paz^{c,d}, Robert J. Schmitz^{e,f,3}, Joseph R. Ecker^{e,f,9}, and Steve A. Kay^{a,4}

^aMolecular and Computational Biology Section, University of Southern California, Los Angeles, CA 90089; ^bSan Diego Center for Systems Biology, University of California, San Diego, La Jolla, CA 92093; ^cSection of Cell and Developmental Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; ^dCenter for Chronobiology, University of California, San Diego, La Jolla, CA 92093; ^eGenomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037; ^fPlant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037; and ⁹Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037

Contributed by Steve A. Kay, July 13, 2015 (sent for review July 1, 2015)

The circadian clock in *Arabidopsis* exerts a critical role in timing multiple biological processes and stress responses through the regulation of up to 80% of the transcriptome. As a key component of the clock, the *Myb*-like transcription factor *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) is able to initiate and set the phase of clock-controlled rhythms and has been shown to regulate gene expression by binding directly to the evening element (EE) motif found in target gene promoters. However, the precise molecular mechanisms underlying clock regulation of the rhythmic transcriptome, specifically how clock components connect to clock output pathways, is poorly understood. In this study, using ChIP followed by deep sequencing of *CCA1* in constant light (LL) and diel (LD) conditions, more than 1,000 genomic regions occupied by *CCA1* were identified. *CCA1* targets are enriched for a myriad of biological processes and stress responses, providing direct links to clock-controlled pathways and suggesting that *CCA1* plays an important role in regulating a large subset of the rhythmic transcriptome. Although many of these target genes are evening expressed and contain the EE motif, a significant subset is morning phased and enriched for previously unrecognized motifs associated with *CCA1* function. Furthermore, this work revealed several *CCA1* targets that do not cycle in either LL or LD conditions. Together, our results emphasize an expanded role for the clock in regulating a diverse category of genes and key pathways in *Arabidopsis* and provide a comprehensive resource for future functional studies.

genome-wide | circadian clock | clock-controlled outputs | transcriptional regulation

In anticipation of the daily changes in light and temperature, the plant's internal circadian clock regulates a large portion of the transcriptome, nearly 80% in rice, poplar, and *Arabidopsis* (1, 2). The clock provides temporal coordination of multiple biological functions to ensure optimal efficiency. For example, before dawn, photosynthetic transcripts accumulate in anticipation of the expected sunlight. Even when shifted to constant light (LL) and temperature, in the absence of day/night cues, transcripts associated with photosynthesis are still up-regulated before the subjective day (1–4). Not only are light- and energy-related activities restricted to particular times, but other aspects of plant growth, including water use, hormone activity, UVB response, and low-temperature response, show this gating effect (5–9). This temporal partitioning of biological responses ultimately provides an optimal integration of the plant's organismal functions with the environment.

The circadian clock in plants involves a posttranscriptional component and a well-studied transcriptional–translation feedback regulation between the *Myb*-like transcription factors *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) and *LATE ELONGATED HYPOCOTYL* (*LHY*) and a member of the *PSEUDO RESPONSE REGULATOR* (*PRR*) family, *TIMING OF CAB2 EXPRESSION 1* (*TOC1*) (10–16). Together these three components contribute to the rhythmic expression of several other circadian regulators

and output genes (17–19). *CCA1*, along with *LHY* and 11 other members of REVEILLE (RVE) proteins, belongs to a subfamily of MYB-domain-containing transcription factors. Primarily, *CCA1* functions as a transcriptional repressor that binds to the evening element (EE), a motif enriched in promoters of genes with peak expression in the evening (3, 13, 20–22). *CCA1* mRNA accumulates around dawn and thus represses the expression of its evening-phased targets (13, 18). However, as *CCA1* protein levels decrease later in the day, this repression is relaxed, and the expression of *CCA1* targets can accumulate, peaking in the evening (13). Although considered a bona fide transcriptional repressor, *CCA1* also can activate the expression of the clock genes *PSEUDO-RESPONSE REGULATOR 9* (*PRR9*) and *PRR7* by binding directly to their promoters (23). *CCA1* is considered a master regulator of clock function, and its misexpression causes severe clock phenotypes that result in altered plant growth and stress responses in *Arabidopsis* (11, 22, 24, 25).

To gain a better understanding of the role that the circadian clock plays in the control of these multiple biological processes through *CCA1*, either directly or through indirect regulation of downstream factors, we performed ChIP sequencing (ChIP-Seq) analyses to

Significance

The circadian clock, an endogenous time-keeping mechanism common to most species, allows organisms to coordinate biological processes with specific times of day. In plants, the role of the clock extends to almost every aspect of growth and development, including responses to biotic and abiotic stresses. The core molecular components and circuits of the clock have been well studied in the model organism *Arabidopsis thaliana*; however, how this mechanism connects to clock-controlled outputs remains poorly understood. Here, we performed a genome-wide characterization of the direct targets of a key clock component in *Arabidopsis*. Our results emphasize the broad role of the plant clock in regulating multiple biological functions and provide direct links between the oscillator and clock-regulated outputs.

Author contributions: D.H.N., C.J.D., J.L.P.-P., and S.A.K. designed research; D.H.N., C.J.D., J.L.P.-P., R.J.S., and J.R.E. performed research; D.H.N., C.J.D., J.L.P.-P., and S.A.K. analyzed data; and D.H.N., C.J.D., and S.A.K. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus database and Sequence Read Archive series GSE70533.

¹D.H.N. and C.J.D. contributed equally to this work.

²Present address: Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695.

³Present address: Department of Genetics, University of Georgia, Athens, GA 30602.

⁴To whom correspondence should be addressed. Email: stevekay@dornsife.usc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1513609112/-DCSupplemental.

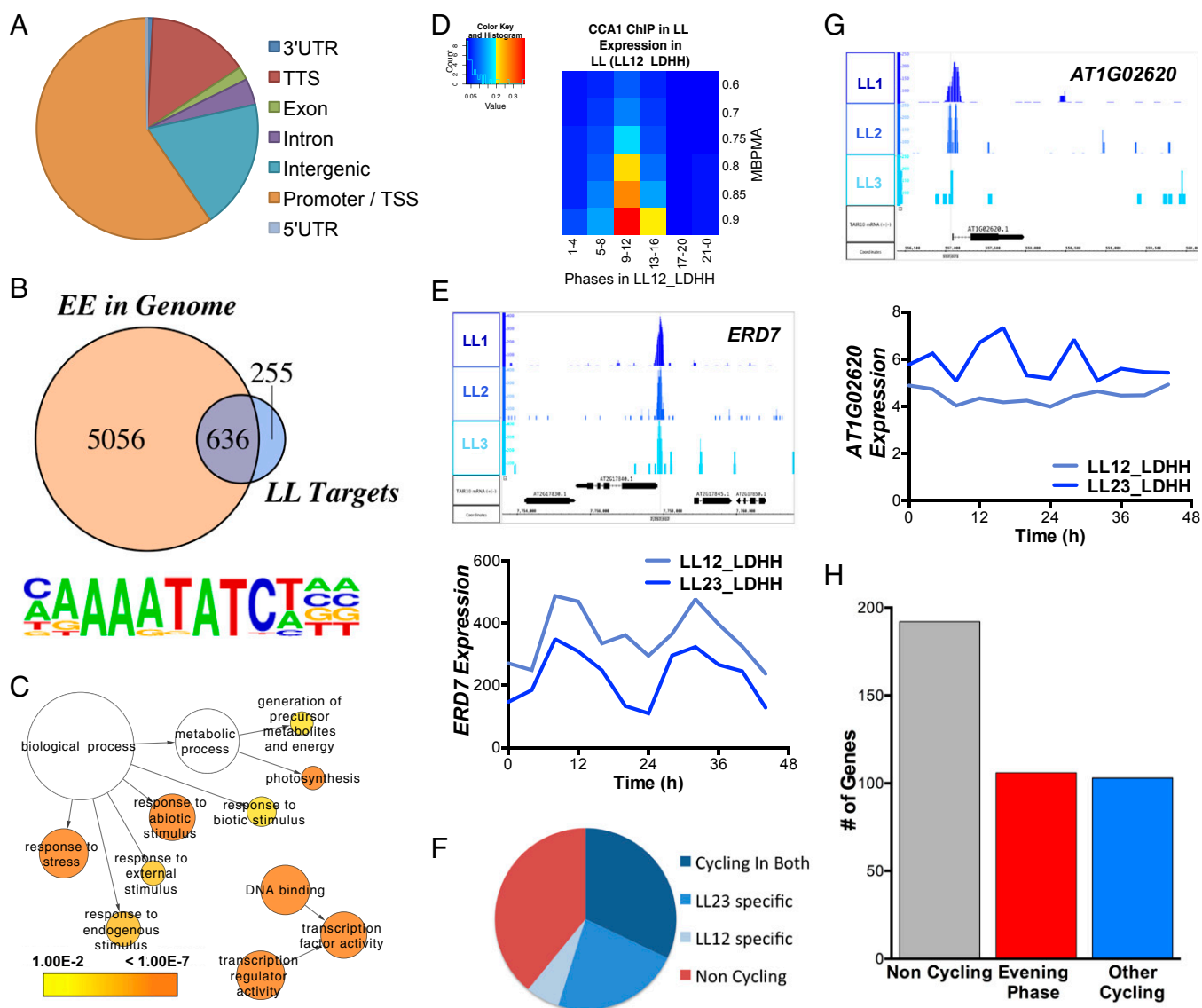








Fig. 1. (A) Genomic distribution of CCA1-occupied sites in LL. Promoter/TSS is defined as the region from $-1,000$ to $+1$ from the TSS. (B) The Venn diagram shows the overlap of the presence of an EE in a region 1-kb upstream of genes closest to CCA1-occupied peaks relative to the genome. (C) Functional enrichment analysis against GO Slim of the 1,100 putative CCA1 targets identified in LL with a peak within 1 kb of the TSS. The sizes of circles are based on the gene numbers. The color of each circle represents the enrichment P value for the GO term label on that circle, with orange representing highest enrichment and yellow the lowest enrichment above the cutoff [false-discovery rate (FDR) corrected 0.01]. The distance between nodes was arranged manually to optimize readability. The graph was generated using BINGO software (63). (D) Phase enrichment of cycling CCA1-occupied targets. The expression datasets include transcripts that display >0.5 MBPMA in constant light [LL12_LDHH, entrained in 12-h:12-h LD cycles at constant temperature (HH) and then subjected to LL; data from the Kay laboratory] (3). Genes were grouped by phase and MBPMA score (strength of cycling) into 36 groups. The ratio of CCA1 occupancy to the number of genes in each group is indicated by the intensity of the heat map (red and blue colors indicate the maximum and minimum ratios of putative targets in the bin, respectively). (E) Example of an EE-containing, evening-phased, and cycling CCA1 target, *EARLY-RESPONSIVE TO DEHYDRATION 7* (*ERD7*, *AT2G17840*). (Upper) Normalized tag counts by location in the genome for each of the three replicate ChIP-Seq experiments in LL. Genome annotation shows the *ERD7* gene visualized using IGB (55). (Lower) *ERD7* shows rhythmic expression in two of the previously published time series experiments in constant conditions, LL12_LDHH and LL23_LDHH [entrained in 12-h:12-h LD cycles at constant temperature (HH) and then subjected to LL; data are from the Millar laboratory (4)]. (F) The putative CCA1 targets were grouped based on their rhythmic expression in previous studies. Genes were classified as noncycling in either of the time series experiments performed in LL, as cycling in both experiments, or as cycling in only one experiment or the other. (G) Example of a CCA1 target, Ras-related small GTP-binding family protein (*AT1G02620*), which does not cycle in either published LL time series. (H) CCA1 targets, selected as having strong peaks of CCA1 occupancy within 1 kb of the TSS, were examined for their phase of peak expression in two published LL time series. Here we classified evening genes as those with peak phase of expression between ZT10 and ZT16.

comprehensively identify the regions of CCA1 occupancy in the *Arabidopsis* genome. Surprisingly, in addition to CCA1 occupancy at EE-containing, evening-expressed, clock-regulated genes, we also observed occupancy near many genes with peak expression in the morning and in proximity to genes that do not cycle in LL conditions. We further investigated the role of CCA1 in diel (light:

dark, LD) conditions and found additional CCA1-occupied targets. Together, these connections between CCA1 and a wide range of circadian output targets indicate a further role for the *Arabidopsis* circadian clock and provide previously unidentified, direct connections between the molecular oscillator and circadian-regulated outputs.

Table 1. Motifs enriched in LL and LD conditions

Condition	Motif	P value	% of targets	% of background
LL		1.00E-434	59.25	10.69
		1.00E-92	28.19	9.27
		1.00E-77	27.01	9.65
LD		1.00E-397	59.74	13.85
		1.00E-121	34.10	11.68
		1.00E-107	49.97	24.42

Results

CCA1-Occupied Regions Are Enriched for the EE and Evening-Expressed Genes in LL. To identify direct CCA1 targets, we performed ChIP-Seq analysis of plants expressing CCA1-GFP driven by the endogenous *CCA1* promoter (CCA1p::CCA1-GFP) (26). Plants were grown in 12-h:12-h LD cycles and then were shifted to LL and harvested at zeitgeber time (ZT) 26 and ZT38 (2 and 14 hrs after subjective dawn on the second day in LL, respectively). For the ZT26 time point, we performed three independent ChIP-Seq experiments, which were highly correlated (Fig. S1A). Few reads were obtained from the ZT38 time point, as expected, because the CCA1 protein driven by its own promoter should be present in very low levels in the evening (11). Therefore we did not perform any additional replicates for ZT38, which is shown here only as a comparison and which was not included in the quantitative analysis presented below.

To identify high-confidence sites of CCA1 occupancy, the three independent biological replicates were combined using the HOMER peak-calling software and ENCODE's irreproducible discovery rate (IDR) analysis (details are given in *Materials and Methods*) (27–30). We identified 1,433 peaks that were associated with the nearest transcription start site (TSS) of 1,305 unique nuclear-encoded putative CCA1 direct targets (Dataset S1). As expected for a transcriptional regulator, many of these peaks were located in the promoter region (Fig. 1A). Even for peaks not classified as predominantly in the promoter region, the majority (68%) were within 1,000 bp of a TSS (Fig. S1B). Additionally, many peaks were identified as associated with the transcription termination site (TTS); this association was unexpected, because previously CCA1 has been observed to interact in the promoter region (20, 24, 31). However, closer examination of these sites revealed that many were between two gene loci oriented in the same direction, and although the peak is within close proximity (<1 kb) to the TSS of the downstream gene, it is closer to the TTS of the upstream gene, reflecting this assignment.

Of the high-confidence peaks, 62% contained the sequence matching the canonical EE (Fig. 1B). Several other motifs were enriched in the CCA1-occupied peaks, including the previously identified protein box (PBX) element (GGGCCA) and G-box (CACGTG) and other motifs not previously reported to be enriched in circadian-regulated genes (Table 1 and Fig. S1D) (1, 21, 32). The 891 genes with a high-scoring CCA1 peak within 1 kb of the TSS were enriched for Gene Ontology (GO) terms describing responses to both external and internal stimuli and photosynthesis (Fig. 1C and Fig. S1E). Consistent with the known role of CCA1 in regulating responses to abiotic and biotic stresses, we observe enrichment in genes with GO terms associated with response to both biotic and abiotic stresses, including response to low temperature and light (Fig. S1E) (8, 9, 25, 33–35). Putative CCA1 targets also are enriched for genes involved in hormone response, including gibberellic acid and abscisic acid, as is consistent with the known role of the circadian clock in the regulation of daily cycles of hormone sensitivity (22, 36).

The majority of the transcripts associated with sites enriched for CCA1 occupancy showed strong cycling expression patterns with an evening phase (Fig. 1D). Specifically, the putative targets with a high-scoring peak of CCA1 occupancy within 1 kb of the TSS are enriched for a peak phase of expression between ZT9 and ZT12, based on published time series expression analyses performed in LL (Fig. 1D and Fig. S1C) (3, 21). These potential CCA1 targets also are strongly cycling, as indicated by enrichment with a model-based, pattern-matching algorithm (MBPMA) >0.8 in the existing processed data (Fig. 1D and E and Fig. S1C) (1). Surprisingly, when we analyzed the transcripts associated with CCA1-occupied sites in the LL time series datasets, we observed a large portion of putative CCA1 targets (37%) that did not show a rhythmic expression pattern in either LL time series (Fig. 1F and G). Although the two published datasets (LL12_LDHH and LL23_LDHH) are both in LL conditions, we observe that some transcripts cycle in one and not in the other, even when the cycling transcripts have different phases in the two datasets. This observation is consistent with previous reports that laboratory-specific conditions might contribute to noticeable differences in cycling calls between similarly derived datasets (1, 21). More than 50% of the peaks associated with these noncycling targets contained the AAATATCT sequence of the EE, suggesting that many clock-controlled genes probably are not rhythmic in LL. It is possible that the stress caused by the artificial conditions of light at night when plants are not expecting it may mask the cycling expression pattern of some genes (37). To determine if any of our putative CCA1 targets might be clock-regulated but masked for cycling expression in LL, we examined the expression of these noncycling CCA1 targets for their rhythmic expression in two LD time courses (LDHH_SM and LDHH_ST) (38, 39). Using the MBPMA cutoff of 0.8 as the threshold for cycling, about 50% of these potential targets were cycling in one of these two LD arrays, many with an evening phase of expression (Fig. 1H) (1). This finding suggests that many potential direct targets of the *Arabidopsis* circadian clock do not cycle in LL conditions and that the estimate of circadian-controlled genes obtained from experiments in LL conditions may underestimate the extensive role of the clock in plants.

Additional CCA1-Occupied Sites Are Observed in LD Conditions. To determine if additional CCA1 targets might be identified and to examine the role of CCA1 transcriptional regulation in LD conditions, we performed two independent ChIP-Seq experiments in LD. As in the LL analysis, we focused on high-confidence sites using IDR analysis and peak calls by HOMER (27–30). We identified 1,607 top-scoring peaks and selected a stringent list of 1,455 putative CCA1 target genes (Dataset S2). To examine the degree of overlap between CCA1-occupied sites in LL and LD, we systematically compared the CCA1 targets from LL and LD. Of the 1,455 occupied sites observed in LD conditions, 1,000 overlapped with the 1,305 sites observed in LL (Fig. 2A). Although the majority of peaks identified in LL also were observed in LD, the overlap between peaks identified in LL and LD was not complete (Fig. 2A), perhaps because of a lack of complete coverage

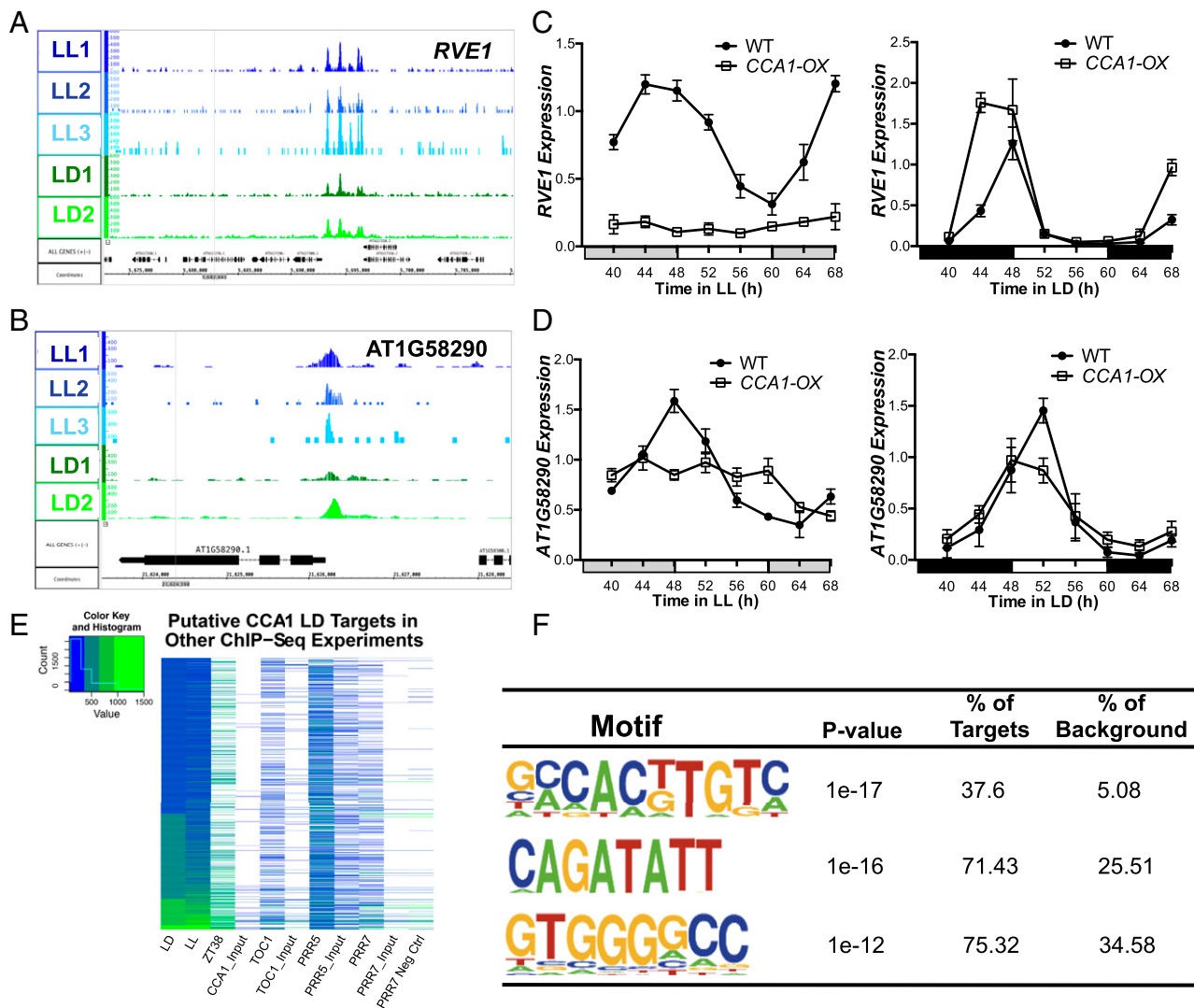


Fig. 3. (A and B) Examples of CCA1-occupied regions in the morning-expressed genes *RVE1* (*AT5G17300*) (A) and *HEMA1* (*AT1G58290*) (B) in LL and LD visualized using IGB (55). Normalized tag counts by location in the genome are shown for each of the three replicate LL ChIP experiments (blue) and for each of the two replicate LD ChIP experiments (green). Genome annotation for *RVE1* or *HEMA1* gene structures is shown below. (C and D) Real-time qRT-PCR of *RVE1* (C) and *HEMA1* (D) transcript levels in wild-type (Col-0) and *CCA1-OX* plants grown in LL after 10 d entrainment in 12-h:12-h LD cycles or only in LD. mRNA levels were normalized to *IPP2* and *PP2A* expression (mean values \pm SD, $n = 2$, two independent experiments). (E) Heatmap comparing the number of reads in this study and published ChIP-Seq experiments for other clock genes: *TOC1* (*AT5G61380*), *PRR5* (*AT5G24470*), and *PRR7* (*AT5G02810*) that map to the 500-bp region surrounding the peaks identified in LD (15, 43, 44). Read counts are normalized to 10 million reads. Normalization factors are LD, 7.48; LL, 3.44 ZT38, 439.75; *CCA1*_Input, 0.73; *TOC1*, 6.43; *TOC1*_Input, 2.24; *PRR5*, 1.38; *PRR5*_Input, 5.66; *PRR7*, 5.10; *PRR7*_Input, 0.64; and *PRR7*_Negative control, 2.04. Each row represents a peak in LD, and the number of reads in that peak area in the corresponding ChIP-Seq experiments for each column is indicated from low (blue) to high (green). Read totals less ≤ 80 are mapped as white. (F) Motifs significantly enriched in CCA1-occupied targets with a morning phase of expression.

with only two and three biological replicates or because of specific binding environments in LL versus LD (29). Again, the identified peaks were enriched for promoter regions, and most (>80%) occurred within 1 kb of a TSS (Fig. S2A and B). The top motif in both ChIP-Seq datasets was the EE, and, as in LL, the EE was highly enriched in the CCA1-occupied regions, along with the PBX element (GGGCCCA) and G-box (CACGTG) (Fig. 2B, Fig. S2C, and Table 1). In fact, for some promoters, more than one previously identified motif associated with circadian rhythms was identified, suggesting multiple regulatory inputs for the promoters of those genes. Although the EE was located at the peak center in both LL and LD, the position of the PBX element was more centered in the peaks identified in LD than in the peaks identified in LL (Fig. S2D). In addition to these

motifs, we also observed several other enriched motifs identified only in the CCA1 ChIP performed in LD (Fig. S2E).

Similar to the putative CCA1 targets in LL, functional analysis of the genes associated with LD peaks included categories for transcriptional regulation, photosynthesis, and responses to stresses (Fig. 2C and Fig. S2F). In addition, the GO analysis of the genes associated with CCA1 peaks in LD conditions also revealed enrichment for genes associated with response to osmotic stress and salinity and with response to metal ions, including cadmium. A link between cadmium and the circadian clock has been observed previously in *Arabidopsis* and other organisms (40–42).

Comparison of the CCA1-occupied targets in LD conditions with published expression datasets representing two LD time courses again showed enrichment for genes with high MBPMA

scores, highlighting groups with robust rhythms and evening-phased expression (Fig. 2D and Fig. S2G). CCA1-occupied sites in LD also are associated with many noncycling genes in LL (Fig. 2E and F). Many of these targets also were not considered as cycling in any of the four expression time series datasets in LD and LL (Fig. 2G). Specifically, 30% of these high-confidence CCA1 peaks did not appear to be cycling (MBPMA <0.8) in any of the 12-h:12-h conditions examined (1). However, the sequence in the region associated with the peaks of many (33, 40%) of these targets contains a perfectly matched EE (AAATATCT).

CCA1 Occupies Promoters of Morning-Phased Transcripts. Our analysis of the CCA1 ChIP-Seq data in LL and LD revealed a potentially larger role for CCA1 in the regulation of morning-expressed genes (Fig. 2G). CCA1 previously has been implicated in the control of specific morning-expressed genes, so perhaps the large number of targets with a morning peak of expression should not have been surprising (9, 23). Overall, these morning-expressed CCA1-occupied targets are enriched for a range of biological processes and functions (Fig. S3A and B). From our top-scoring targets from the LD CCA1 ChIP, we considered those that are morning expressed and have a peak of CCA1 binding within 1 kb of the TSS. Of the CCA1 targets identified in LD that showed strong rhythmic expression in at least one of the four LL or LD time courses (923 transcripts), about 40% showed a peak of expression in the morning (defined as ZT22–ZT4). A similar percent was observed for the CCA1-associated transcripts identified in LL. We selected targets that were in the top-scoring CCA1 peak list in both LL and LD, that were strongly cycling with a peak of expression around dawn in all four time courses, and that contained an EE in the 1-kb upstream region upstream of the TSS.

The genes associated with the top two peak scores that met these criteria were the CCA1 homolog *RVE1* and *HEMA1*, a gene involved in chlorophyll biosynthesis. For both genes, the CCA1 ChIP peak is centered around the EE in the 1-kb upstream region in all LL and LD ChIP experiments (Fig. 3A and B). In plants grown in LL, overexpression of *CCA1* (*CCA1-OX*) and/or loss of *CCA1* expression in combination with the close *CCA1* homolog *LHY* (*cca1-1/lhy-21*) disrupts the rhythmic expression of these two transcripts, indicating that they indeed are regulated by CCA1 (Fig. 3C and D and Fig. S3C and D). In LD conditions, *CCA1-OX* plants still show a sharp peak in *RVE1* expression in the morning, but the peak is slightly wider with an advanced phase (Fig. 3C) in the *CCA1-OX* plants compared to WT. For *HEMA1*, the peak of expression also occurs around dawn in *CCA1-OX* plants (Fig. 3D). These observations suggest that although the expression of both *RVE1* and *HEMA1* is severely affected by *CCA1* overexpression in LL, *CCA1* overexpression has a less significant effect in LD, reflecting the complicated nature of the circadian clock. Furthermore, although the effects of *CCA1* overexpression on *RVE1* and *HEMA1* are similar in LL, the differences in effects of *CCA1* overexpression on these two targets in LD conditions highlight possible feedback interactions between light signals and other clock components even on direct CCA1 targets. For example, *RVE1* is a direct target of CCA1 in our analysis but also has been identified as a direct target of PRR5 and PRR7 (43, 44). This potential for combinatorial regulation by CCA1 and other circadian regulators appears to be abundant. For example, comparison of the CCA1 targets in LD with *TOC1*, *PRR5*, and *PRR7* ChIP-Seq targets showed that the 500-bp regions surrounding the promoters of many of the CCA1 targets are occupied by these other clock genes as well (Fig. 3E) (43, 44).

Motif analysis of these putative CCA1-regulated morning transcripts indicated that the EE was not enriched above background in this set of targets. Instead, we observed three significant motifs, the first of which shows some similarity to the G-box and E-box motifs, suggesting that CCA1 might interact with basic

helix-loop-helix transcription factors to regulate some of these morning-expressed genes (Fig. 3F).

Expression of CCA1 Target Genes Is Disrupted by Overexpression of CCA1. To determine if CCA1 occupancy plays a functional role in the regulation of the predicted targets, we analyzed target gene mRNA levels by quantitative RT-PCR (qRT-PCR) in wild-type and *CCA1-OX* plants. As proof of concept, we first validated genes previously identified as direct targets of CCA1. For example, the promoters of the core clock gene *TOC1*, a recently identified clock component *JUMONJI DOMAIN PROTEIN 5* (*JMD5*), and three genes involved in reactive oxygen species homeostasis and oxidative stress responses, *COLD-REGULATED GENE 27* (*COR27*), *CATALASE 3* (*CAT3*), and the zinc-finger protein, *ZAT12*, were all occupied by CCA1 in both LD and LL (Fig. 4A and Fig. S4A–C) (13, 24, 31, 45). We further examined previously uncharacterized CCA1 targets that contain an EE and are evening expressed. Among others, we observe misregulation of *RING MEMBRANE-ANCHOR 1* (*RMA1*), an uncharacterized protein with possible E3 ubiquitin ligase activity, and *DREB2C*, a member of the DREB subfamily A-2 of the ERF/AP2 transcription factor family implicated in drought stress responses (Fig. 4B–E) (46–49). Interestingly, in addition to *DREB2C*, 39 other transcription factors identified as CCA1 targets in this study recently have been shown to interact with the *CCA1* promoter in a yeast one-hybrid assay (Dataset S3) (50). Of the 30,000 accessible genomic loci (or the actual 41,671 loci), our list of CCA1 targets is enriched for the transcription factors that show reciprocal regulation based in the yeast one-hybrid assay (hypergeometric *P* value <0.01). Finally, we found that the expression of noncycling CCA1 target genes, such as *AT1G02620* (a Ras-related small GTP-binding family protein) (shown in Figs. 1 and 2), and *ABCG2* (ATP-binding cassette G2, *AT2G37360*), is altered in *CCA1-OX* plants in LL (Fig. 4F and G and Fig. S4D). These results indicate that a functional relationship likely exists between CCA1 and noncycling transcripts and imply a broader role for the clock in the regulation of plant biological functions.

Our initial approach was to identify a set of high-confidence peaks that might be direct CCA1 targets to examine conservatively the potential role for this key circadian regulator. However, this stringent analysis misses some previously identified CCA1-regulated targets, such as *PRR9*, and therefore likely does not include many other real targets (Fig. S5A) (23). Our initial approach was to assign only the nearest TSS as a target of the peak, but we expanded our analysis to include all TSS within 1 kb of the peak, thus allowing more than one target per peak and the inclusion of bidirectional promoters (51, 52). This inclusive target list identified 2,327 TSS within 1 kb of the peaks in LL and 2,459 TSS in LD conditions (Dataset S4). These TSS correspond to 1,725 and 1,919 unique gene models that are within 1 kb of a high-confidence site of CCA1 occupancy in LL and LD, respectively. This inclusive target list identified *NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED3* (*LNK3*) as a CCA1 target, and the severely altered *LNK3* transcript level in *CCA1-OX* and *cca1-1/lhy-21* plants in LL confirms that it is indeed regulated by CCA1 (Fig. S5B–D) (53). We therefore prepared a more inclusive list of CCA1 targets and made the ChIP-Seq data available on the Integrated Genome Browser (IGB) Quickload site (bioviz.org/igb/index.html) (54) and through the AnnoJ browser (neomorph.salk.edu/Kay_CCA1.php) (55) to enable researchers to browse for potential CCA1 peaks near the promoters of their favorite gene (Dataset S4).

Discussion

Our analysis and results from these CCA1 ChIP-Seq experiments show that the circadian clock in *Arabidopsis* plays a much greater role than previously established through examination of experiments in LL and LD. Additionally, we have identified a set of

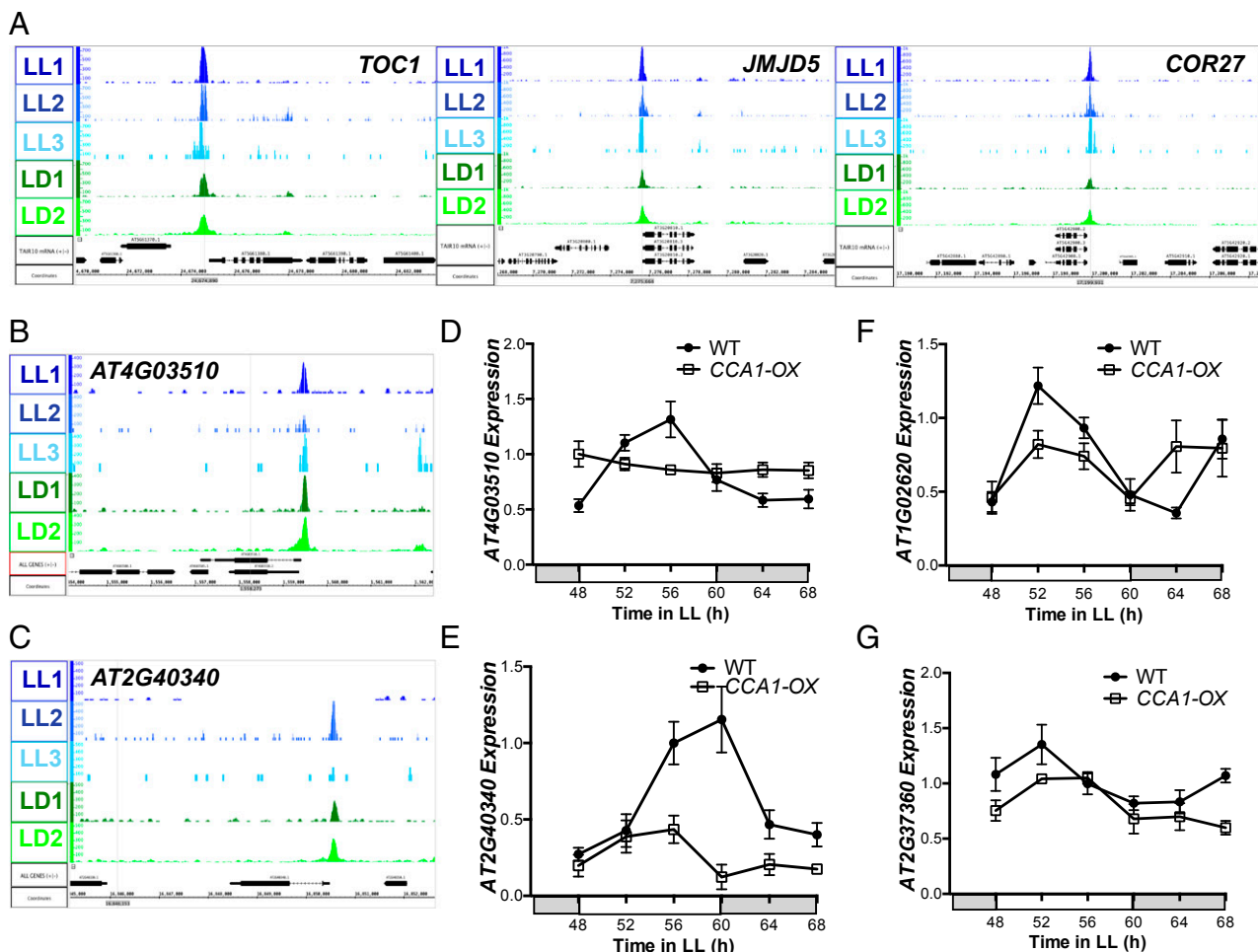


Fig. 4. (A–C) CCA1-occupied peaks for the previously identified CCA1 targets *TOC1* (*AT5G61380*), *JMJD5* (*AT3G20810*), and *COR27* (*AT5G42900*) (A) and uncharacterized CCA1 targets *RMA1* (*AT4G03510*) (B) and *DREB2C* (*AT2G40340*) (C). Normalized tag counts by location in the genome are shown for each of the three replicate LL ChIP-Seq experiments (blue) and for each of the two replicate LD ChIP-Seq experiments (green). Genome annotation for each gene structure, visualized using IGB (55), is shown below. (D–G) qRT-PCR of wild-type and *CCA1-OX* plants grown in LL after 10-d entrainment in 12-h:12-h LD cycles for *RMA1* (D), *DREB2C* (E), *AT1G02620* (F), and *ABCG2* (*AT2G37360*) (G). mRNA levels were normalized to *IPP2* and *PP2A* expression (mean values \pm SD, $n = 2$, two independent experiments).

high-confidence sites of CCA1 occupancy that now can be examined for direct links between the circadian clock and known and previously unidentified circadian-regulated outputs. Based on conservative criteria, we identified more than 1,000 potential direct CCA1 targets from both LD and LL experiments that were enriched for a broad range of biological processes. The majority of the genes associated with the top-scoring peaks contained the known CCA1-binding site, the EE, and showed a rhythmic pattern with a peak of expression in the evening in at least one of four LL and LD microarray time series. Interestingly some of the putative CCA1 targets do not pass the cycling criteria in either of the time courses in LL. Therefore, these are CCA1 targets that previously have not been considered as circadian clock-regulated genes in the classic sense.

The ChIP-Seq analysis performed in this study was aligned to the Columbia-0 (*Col-0*) genome; however the *CCA1::CCA1::GFP* plants are in the *Wassilewskija* (*Ws*) background. We performed a parallel analysis on a genome backbone constructed by replacing the Columbia genome with *WS* SNPs from the 1,001 genomes project and observed a large increase in the number of reads that mapped to the genome and a relative increase in the number of identified peaks (Dataset S5) (56, 57). Although we focused here on the peaks that were consistent between *Col-0*

and *Ws*, future sequencing of the *Ws* genome or development of a proper reference may identify additional ecotype-specific peaks (58).

We identified CCA1 occupancy at promoters of genes that were noncycling and therefore likely were never included as clock-regulated transcripts in any expression datasets. It is possible that some of these targets cycle in specific tissues but are diluted in the expression datasets, which are based on whole seedlings. Performing a tissue-specific analysis similar to that recently described (59) could help distinguish true noncycling transcripts. Although LL has been the standard condition for identifying clock-controlled genes in plants, this criterion might overlook a subset of circadian-regulated targets, because not all direct targets of the clock necessarily exhibit rhythmic expression in LL. It is also possible that an underlying rhythmic transcription of these genes is masked by changes in mRNA stability (60, 61). In addition, we observe occupied sites specific for the LD conditions, suggesting a possible condition-specific role for CCA1 regulation of select targets. CCA1 occupancy at promoters of morning genes is enriched for motifs other than the EE, suggesting that CCA1 might regulate a large subset of its targets through interaction with other transcription factors. A yeast two-hybrid assay of CCA1 against

the recently published *Arabidopsis* transcription factor library could help refine the mechanism of CCA1 regulation of target genes (50).

A comparison between our list of CCA1 targets and recently published CCA1 promoter interactors showed significant overlap. Interestingly, we observe peaks of CCA1 occupancy in the promoters of both transcription factors and their downstream targets as well as in the promoters of potential regulators of CCA1 expression, suggesting the possibility of both feedback and feed-forward loops between the clock and stress-response pathways. For example, DREB2C, a transcription factor up-regulated during heat stress, has a significant peak of CCA1 occupancy 88 bp upstream of the TSS centered directly above a canonical EE. In addition, CCA1 peaks also are observed at the DREB2C homologs C-repeat binding factor (CBF)1, -2, and -3 as previously reported, and, in fact, CBF1 also interacts with the CCA1 promoter (50). Many of the downstream targets of CBFs/DREB2C, including cold-responsive protein 6.6 (COR6.6/KIN2), COR27, and COR15B, are also occupied by CCA1, further highlighting the complexity and crosstalk between clock-controlled pathways (47).

An often overlooked challenge of ChIP-Seq analysis is the assignment of target genes to identified peaks. The most common approaches, optimized for the mammalian genome through the ENCODE project, have limitations when implemented in other species. One typical approach, applied to the data presented here, is to assign the peak of occupancy for the transcription factor to the nearest TSS. Although in a large genome, such as human or mouse, this approach often provides an unequivocal assignment to a single gene, the peak assignment may be ambiguous in the smaller genome of *Arabidopsis*. For example, *PRR9*, a known target of CCA1, did not appear in our high-confidence target list in either LD or LL (23). Examination of the *PRR9* genomic region revealed a significant peak centered less than 243 bp upstream of the *PRR9* TSS. However, the locus *AT2G46787* is located within this region upstream of *PRR9*, and the peak center is only 27 bp from the putative TSS of this region (Fig. S5A). Another example of a gene that would have been missed by this analysis is *LNK3*. A significant peak of CCA1 occupancy occurs 356 bp upstream of the *LNK3* TSS. However, this peak is also 222 bp upstream of the TSS of *AT3g12300* (Fig. S5B). In addition to relaxing the criteria for identifying transcripts associated with peaks of CCA1 occupancy, we also could relax our highly stringent peak-calling criteria to include additional peaks. Therefore, we have made the CCA1 ChIP-Seq data available through the Integrated Genome Browser's Quickload Site (bioviz.org/igb/index.html) (55) so that researchers can examine their regions of interest for peaks that did not meet the strict threshold used for considering peaks in our analysis (Dataset S4).

In summary, the data presented in this study can be useful for direct functional characterization of specific target genes that link

to key biological processes and can contribute significantly to our understanding of the underlying molecular mechanisms of clock-controlled pathways in *Arabidopsis* and possibly other plants.

Materials and Methods

Plant Materials. *Arabidopsis thaliana* Columbia-0 (Col-0) or Wassilewskija (Ws) ecotypes were used unless otherwise indicated. The CCA1::GFP-CCA1 (Ws), CCA1-OX (Col-0), and *cca1-1/hy-21* (Col-0) lines were described previously (26, 62). Plants were grown on plates containing Murashige and Skoog (MS) medium (Sigma) [1.5% agar (wt/vol) (Sigma)] under 12-h:12-h LD cycles at 22 °C.

RNA Preparation and qRT-PCR. Seeds were plated on MS plates supplemented with 3% (wt/vol) sucrose, stratified for 2–3 d at 4 °C, and grown in LD conditions at 22 °C for 10 d. For LL experiments, seedlings were transferred to LL for 2 or 3 d. Samples were collected every 4 h in both LD and LL conditions. Total RNA was isolated with the Qiagen RNeasy plant mini kit (Qiagen). cDNA was synthesized using 1 µg of total RNA and was reverse-transcribed with the iScript cDNA synthesis kit (Bio-Rad). Primers used in this study for mRNA expression were *TOC1* (5'-AATAGTAATCCAGCGCAATTTTCTTC-3'; 5'-CTTCACTACTTTTCTTCGGTGTCT-3'); *RVE1* (5'-AGAACATGTGGGCTCAAAGAC-3'; 5'-GGTATTACAATCGGCTCTACTGAG-3'); *AT1G58290* (5'-TCTCAGCATCTGGAGTTAAAG-3'; 5'-GTTGACAAATCTCTGAACTGGG-3'); *AT4G03510* (5'-TGTTGATGGT TGGAGAAATGG-3'; 5'-GTCCCTGCGAGATTGTAAGTG-3'); *AT1G02620* (5'-CACTGCTCCCAATCTCATCTTA-G3'; 5'-CACTTTCCTTTACAGATTGTG-3'); *AT2G40340* (5'-CCGTCGGAGATTGTTGACAG-3'; 5'-GGAGGAGGCTTTGGATTG-3'); *AT2G37360* (5'-AAGTCCTCAGAGAATCGCAC-3'; 5'-GAGGGAGATGTGTTGGTGAG-3'); and *LNK3* (5'-CAAAGGCTTTCTATGGTGCTC-3'; 5'-CCGACACTGTTATTGCTACTGG-3'). As a normalization control, we used isopentenyl-diphosphate delta-isomerase II (*IPP2*) (5'-GTATGAGTTGCTTGGAGCAAAG-3'; 5'-GAGGATGGCTGCAACAAGTGT-3') and *PP2A* (5'-TAACGTGGCCAAATGATGC-3'; 5'-GTTCTCCACAACCGATTGGT-3'). PCR conditions were 95 °C for 3 min followed by 40 cycles at 95 °C for 10 s, 55 °C for 15 s, and 72 °C for 15 s.

ChIP, Library Preparation, Deep Sequencing, and Peak Analysis. Detailed descriptions of the CCA1 ChIP-Seq procedure and data analysis are provided in *SI Materials and Methods*. Raw fastq files, normalized bed files, and peak calls for the LL and LD ChIP-Seq experiments have been uploaded to NCBI Gene Expression Omnibus (GEO) and Sequence Read Archive under GEO series GSE70533. Normalized read counts were uploaded to IGB Quickload (bioviz.org/igb/index.html) (55) and the AnnoJ Browser (neomorph.salk.edu/Kay_CCA1.php) (54).

ACKNOWLEDGMENTS. We thank Sabrina Sanchez for critical reading of the manuscript, members of the S.A.K. laboratory for helpful discussions, Ryan Sartor for assistance and guidance with final formatting, and Ann Loraine and Huaming Chen for assistance uploading the data to IGB and AnnoJ, respectively. The research reported in this publication was supported by Ruth L. Kirschstein National Research Service Award F32GM090375 (to D.H.N.), National Science Foundation Postdoctoral Research Fellowship in Biology Award 0906055 (to C.J.D.), National Institutes of General Medicine of the NIH Awards R01GM056006, R01GM067837, and RC2GM092412 (to S.A.K.), National Science Foundation Award MCB-1024999 (to J.R.E.), and Howard Hughes Medical Institute (HHMI) and Gordon and Betty Moore Foundation (GBMF) Award GBMF3034 (to J.R.E.). J.R.E. is an investigator of the HHMI-GBMF.

- Michael TP, et al. (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet* 4(2):e14.
- Filichkin SA, et al. (2011) Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS One* 6(6):e16907.
- Harmer SL, et al. (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290(5499):2110–2113.
- Edwards KD, et al. (2006) FLOWERING LOCUS C mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. *Plant Cell* 18(3):639–650.
- Carbonell-Bejerano P, et al. (2014) Circadian oscillatory transcriptional programs in grapevine ripening fruits. *BMC Plant Biol* 14(1):78.
- Covington MF, Harmer SL (2007) The circadian clock regulates auxin signaling and responses in *Arabidopsis*. *PLoS Biol* 5(8):e222.
- Fehér B, et al. (2011) Functional interaction of the circadian clock and UV RESISTANCE LOCUS 8-controlled UV-B signaling pathways in *Arabidopsis thaliana*. *Plant J* 67(1):37–48.
- Fowler SG, Cook D, Thomashow MF (2005) Low temperature induction of *Arabidopsis* CBF1, 2, and 3 is gated by the circadian clock. *Plant Physiol* 137(3):961–968.
- Dong MA, Farré EM, Thomashow MF (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(17):7241–7246.
- Schaffer R, et al. (1998) The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* 93(7):1219–1229.
- Wang Z-Y, Tobin EM (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* 93(7):1207–1217.
- Strayer C, et al. (2000) Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. *Science* 289(5480):768–771.
- Alabadi D, et al. (2001) Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* 293(5531):880–883.
- Gendron JM, et al. (2012) *Arabidopsis* circadian clock protein, *TOC1*, is a DNA-binding transcription factor. *Proc Natl Acad Sci USA* 109(8):3167–3172.
- Huang W, et al. (2012) Mapping the core of the *Arabidopsis* circadian clock defines the network structure of the oscillator. *Science* 336(6077):75–79.
- Pokhilko A, et al. (2012) The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol Syst Biol* 8:574.
- Pruneda-Paz JL, Kay SA (2010) An expanding universe of circadian networks in higher plants. *Trends Plant Sci* 15(5):259–265.
- Nagel DH, Kay SA (2012) Complexity in the wiring and regulation of plant circadian networks. *Curr Biol* 22(16):R648–R657.

19. McClung CR (2014) Wheels within wheels: New transcriptional feedback loops in the Arabidopsis circadian clock. *F1000Prime Rep* 6:2.
20. Harmer SL, Kay SA (2005) Positive and negative factors confer phase-specific circadian regulation of transcription in Arabidopsis. *Plant Cell* 17(7):1926–1940.
21. Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL (2008) Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol* 9(8):R130.
22. Harmer SL (2009) The circadian system in higher plants. *Annu Rev Plant Biol* 60(1):357–377.
23. Farré EM, Harmer SL, Harmon FG, Yanovsky MJ, Kay SA (2005) Overlapping and distinct roles of PRR7 and PRR9 in the Arabidopsis circadian clock. *Curr Biol* 15(1):47–54.
24. Lai AG, et al. (2012) CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proc Natl Acad Sci USA* 109(42):17129–17134.
25. Goodspeed D, Chehab EW, Min-Venditti A, Braam J, Covington MF (2012) Arabidopsis synchronizes jasmonate-mediated defense with insect circadian behavior. *Proc Natl Acad Sci USA* 109(12):4674–4677.
26. Pruneda-Paz JL, Breton G, Para A, Kay SA (2009) A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. *Science* 323(5920):1481–1485.
27. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576–589.
28. Li Q, Brown JB, Huang H, Bickel PJ (2011) Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* 5(3):1752–1779.
29. Landt SG, et al. (2012) ChIP-seq guidelines and practices of the ENCODE and mod-ENCODE consortia. *Genome Res* 22(9):1813–1831.
30. Karmel A (2014) homer-1d first-pass. Available at dx.doi.org/. Accessed June 27, 2015.
31. Jones MA, et al. (2010) Jumonji domain protein JMJD5 functions in both the plant and human circadian systems. *Proc Natl Acad Sci USA* 107(50):21623–21628.
32. Michael TP, et al. (2008) A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol* 6(9):e225.
33. Wang W, et al. (2011) Timing of plant immune responses by a central circadian regulator. *Nature* 470(7332):110–114.
34. Zhang C, et al. (2013) Crosstalk between the circadian clock and innate immunity in Arabidopsis. *PLoS Pathog* 9(6):e1003370.
35. Zhou M, et al. (2015) Redox rhythm reinforces the circadian clock to gate immune response. *Nature* 523:472–476.
36. Penfield S, Hall A (2009) A role for multiple circadian clock genes in the response to signals that break seed dormancy in Arabidopsis. *Plant Cell* 21(6):1722–1732.
37. Velez-Ramirez AI, van Ieperen W, Vreugdenhil D, Millenaar FF (2011) Plants under continuous light. *Trends Plant Sci* 16(6):310–318.
38. Smith SM, et al. (2004) Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in Arabidopsis leaves. *Plant Physiol* 136(1):2687–2699.
39. Bläsing OE, et al. (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. *Plant Cell* 17(12):3257–3281.
40. Hinrichsen RD, Tran JR (2010) A circadian clock regulates sensitivity to cadmium in *Paramecium tetraurelia*. *Cell Biol Toxicol* 26(4):379–389.
41. Maistri S, DalCorso G, Vicentini V, Furini A (2011) Cadmium affects the expression of ELF4, a circadian clock gene in Arabidopsis. *Environ Exp Bot* 72(2):115–122.
42. Jiménez-Ortega V, Cano Barquilla P, Fernández-Mateos P, Cardinali DP, Esquifino AI (2012) Cadmium as an endocrine disruptor: Correlation with anterior pituitary redox and circadian clock mechanisms and prevention by melatonin. *Free Radic Biol Med* 53(12):2287–2297.
43. Nakamichi N, et al. (2012) Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proc Natl Acad Sci USA* 109(42):17123–17128.
44. Liu T, Carlsson J, Takeuchi T, Newton L, Farré EM (2013) Direct regulation of abiotic responses by the Arabidopsis circadian clock component PRR7. *Plant J* 76(1):101–114.
45. Lu SX, et al. (2011) The Jumonji C domain-containing protein JMJ30 regulates period length in the Arabidopsis circadian clock. *Plant Physiol* 155(2):906–915.
46. Matsuda N, Suzuki T, Tanaka K, Nakano A (2001) Rma1, a novel type of RING finger protein conserved from Arabidopsis to human, is a membrane-bound ubiquitin ligase. *J Cell Sci* 114(Pt 10):1949–1957.
47. Lim CJ, et al. (2007) Over-expression of the Arabidopsis DRE/CRT-binding transcription factor DREB2C enhances thermotolerance. *Biochem Biophys Res Commun* 362(2):431–436.
48. Chen H, et al. (2010) Arabidopsis DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. *Biochem Biophys Res Commun* 401(2):238–244.
49. Hwang JE, et al. (2012) Overexpression of Arabidopsis dehydration-responsive element-binding protein 2C confers tolerance to oxidative stress. *Mol Cells* 33(2):135–140.
50. Pruneda-Paz JL, et al. (2014) A genome-scale resource for the functional characterization of Arabidopsis transcription factors. *Cell Reports* 8(2):622–632.
51. Wang Q, et al. (2009) Searching for bidirectional promoters in Arabidopsis thaliana. *BMC Bioinformatics* 10(Suppl 1):S29.
52. Kourmpetli S, et al. (2013) Bidirectional promoters in seed development and related hormone/stress responses. *BMC Plant Biol* 13:187.
53. Rugnone ML, et al. (2013) LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. *Proc Natl Acad Sci USA* 110(29):12120–12125.
54. Lister R, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133(3):523–536.
55. Nicol JW, Helt GA, Blanchard SG, Jr, Raja A, Loraine AE (2009) The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25(20):2730–2731.
56. Gan X, et al. (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature* 477(7365):419–423.
57. Schmitz RJ, et al. (2013) Patterns of population epigenomic diversity. *Nature* 495(7440):193–198.
58. Schneeberger K, et al. (2011) Reference-guided assembly of four diverse Arabidopsis thaliana genomes. *Proc Natl Acad Sci USA* 108(25):10249–10254.
59. Endo M, Shimizu H, Nohales MA, Araki T, Kay SA (2014) Tissue-specific clocks in Arabidopsis show asymmetric coupling. *Nature* 515(7527):419–422.
60. Koike N, et al. (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338(6105):349–354.
61. Menet JS, Rodriguez J, Abruzzi KC, Rosbash M (2012) Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife* 1:e00011.
62. Nagel DH, Pruneda-Paz JL, Kay SA (2014) FBH1 affects warm temperature responses in the Arabidopsis circadian clock. *Proc Natl Acad Sci USA* 111(40):14595–14600.
63. Maere S, Heymans K, Kuiper M (2005) BINGO: A Cytoscape plugin to assess over-representation of gene ontology categories in biological networks. *Bioinformatics* 21(16):3448–3449.
64. Andrews S Babraham Bioinformatics FastQ Screen. Available at www.bioinformatics.babraham.ac.uk/projects/fastq_screen/. Accessed June 27, 2015.
65. Andrews S (2013) FastQC: A quality control tool for high throughput sequence data. Available at www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed June 27, 2015.
66. Sawa M, Nusinow DA, Kay SA, Imaizumi T (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science* 318(5848):261–265.
67. Lassmann T, Hayashizaki Y, Daub CO (2009) TagDust—a program to eliminate artifacts from next generation sequencing data. *Bioinformatics* 25(21):2839–2840.
68. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864.
69. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
70. Lassmann T (2015) TagDust2: A generic method to extract reads from sequencing data. *BMC Bioinformatics* 16:24.
71. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAM tools. *Bioinformatics* 25(16):2078–2079.
72. Lawrence M, et al. (2013) Software for computing and annotating genomic ranges. *PLoS Comput Biol* 9(8):e1003118.
73. Carlson M TxDb.Athaliana.BioMart.plantsmart25: Annotation package for TxDb object(s). Available at www.bioconductor.org/packages/release/data/annotation/html/TxDb.Athaliana.BioMart.plantsmart25.html. Accessed June 20, 2015.
74. Droit A, Gottardo R, Roberston G, Li L (2014) rGADEM: De novo motif discovery. Available at www.bioconductor.org/packages/release/bioc/html/rGADEM.html. Accessed June 25, 2015.
75. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14(6):1188–1190.
76. Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.
77. RDC Team (2010) R: A Language and Environment for Statistical Computing (The R Project for Statistical Computing, Vienna) Available at www.R-project.org. Accessed June 25, 2015.
78. Warnes GR, et al. (2015) gplots: Various R Programming Tools for Plotting Data Available at cran.r-project.org/web/packages/gplots/index.html. Accessed June 25, 2015.