Circadian Clocks in the Real World: Effects of Dynamic Light Regimes on the Regulation of Circadian Gene Expression in Cyanobacteria

Circadian Clocks in the Real World: Effects of Dynamic Light Regimes on the Regulation of Circadian Gene Expression in Cyanobacteria

A dissertation presented

by

Joseph Robert Piechura

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

July, 2017

Dissertation advisor: Professor Erin K. O'Shea Joseph Robert Piechura

Circadian Clocks in the Real World: Effects of Dynamic Light Regimes on the Regulation of Circadian Gene Expression in Cyanobacteria

Abstract

Organisms use internal oscillators to control their physiology in conjunction with the predictable environmental changes of the day/night cycle. However, it is not always clear how internal timing information is used by molecular pathways called output pathways to change physiology. Further, these autonomous systems must operate in constantly fluctuating natural environments, and it is not clear how the functions of circadian clocks are affected by these conditions. We use the simple circadian clock in the model cyanobacterium *Synechococcus elongatus* PC7942 as a model system to explore these questions.

We asked how timing information encoded in the circadian clock in cyanobacteria is used to control a simple circadian output in the form of genome-wide changes in transcription (Chapter 2). We find that the gene encoding the transcription factor RpaA is required for genome-wide transcription rhythms. Further, we show that clock-controlled changes in the phosphorylation state of RpaA allow the protein to bind to promoters to activate expression of rhythmic genes. We demonstrate that phosphorylated RpaA drives dynamic expression patterns of hundreds of clock-controlled genes. Thus, the core circadian clock controls gene expression through phosphorylated RpaA, which acts as a master regulator to enact changes in expression of hundreds of genes.

Next, we asked how the transcriptional output of the circadian clock is affected by naturally-relevant changes in environmental conditions (Chapter 3). We grew cyanobacteria under dynamic light regimes mimicking natural conditions to demonstrate that the expression of clock-controlled genes is a function of changes in environmental light intensity.

Using genomics, we identify that these environmentally-responsive changes in gene expression are enacted by modulating the recruitment of RNA polymerase to promoters. Using a combination of genomics and mathematical modeling, we implicate light-induced changes in the phosphorylation of the transcription factor RpaB as an important mechanism by which environmental changes modulate the expression of clock-controlled genes. Further, we find that RpaA activity itself can be affected by environmental changes. Our work demonstrates the basic principles governing the integration of changes in light with the output of the circadian clock, and suggests several possible mechanisms underlying this behavior.

Our work demonstrates that relatively simple circuitry underlies the conversion of timing information from the circadian clock to genome-wide gene expression in cyanobacteria. Further, we demonstrate principles describing how circadian clock output will change in conjunction with environmental changes in nature.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Erin O'Shea, for her mentorship and support throughout my PhD. I benefited greatly from viewing Erin's clear, logical, and efficient approach to science and professional matters. I am thankful that Erin paired me with excellent mentors to help my transition into the lab, and then subsequently gave me considerable freedom to explore my own ideas. I am also thankful to my committee members Andrew Murray, Vlad Denic, and Mike Laub for mentorship and advice throughout my PhD. I would also like to thank Peter Arvidson, Jeff Offerman, Mike Lawrence, Patty Perez, Jack Conlin, and the MCB administration for their help in making my PhD possible.

I benefitted especially from the mentorship of Andrian Gutu, Joe Markson, and Kapil Amarnath. Andrian was my rotation advisor, and served a role as a second mentor throughout my PhD. He was always available for me to bounce hypotheses and experiments off of him, provide technical advice, and emotional support. I especially enjoyed getting to know Andrian personally and watching him in his journey of fatherhood as he raised his daughter Althea.

Once I decided to join the O'Shea lab, I was paired to be trained in various techniques by Joe Markson and work with him on his project studying the transcription factor RpaA. Joe was a very inspiring person to work with. I was incredibly impressed by the breadth of the science he had undertaken in the past, from in vitro biochemistry to mathematical modeling, and the fearlessness he showed as he moved into a new project in the totally different field of genomics. Joe's analytical, methodical, and thorough approach to science and his manner of treating no technical step as an obstacle left a significant impression on me. His congenial and empathetic nature made him a pleasure to work with, and I am proud of the work we completed together presented here in Chapter 2.

Kapil Amarnath joined the O'Shea lab in my third year of graduate school and became a great friend and intellectual mentor. Kapil was new to Biology, coming from a Physical Chemistry background, and I got to watch as he learned about the nature of Molecular Biology research, and hear his views on them from an outside perspective. In talking with Kapil I learned more of the drawbacks of traditional approaches to Biology, and mused on other approaches to problems that got more to the root of what we cared about. His friendship outside of science was just as influential on me, and I am proud of the work we did together presented here in Chapter 3.

Working in the O'Shea lab was one of the best parts of my PhD. I had the immense pleasure of working with brilliant people from all over the world with intense passion for different aspects of science, and different aspects of life. I am thankful for Shankar, Andrian, Kapil, Joe M., Lauren, Alicia, Bin, Chris Chidley, Anders, Luca, Ania, Siting, Xiao-yu, Rasi, Brian Zid, Xu, Elmar, Zhou, Matt, Kathleen, and Vikram for stimulating lunches, coffee clubs, dinners, Paintball battles, strenuous Hikes, Lab Half Marathons, 5K races we took too seriously, attempts at volleyball glory, trivia nights at Shankar's place using children's toys, ice luge shots, parties of varying levels of craziness, karaoke failures, lessons about life as a parent, lessons about the each other's homes, and incredibly stimulating scientific discourses. I smiled going to work many days knowing that I would get to spend time with these people.

I am grateful to have completed my PhD while in the stimulating community of the FAS

Center for Systems Biology. The Wednesday morning group meetings were a stimulating environment in which a broad array of scientists engaged in discussions about a broad array of science. I learned a lot from hearing the views of people trained in the fields I had little exposure to like physics and population genetics. The Friday Happy Hours provided a lot of great conversation and adventures where I got to get to know these bright minds in a different setting.

I am thankful to have shared my PhD journey Zain, Eddie, Ezgi, Phil, Sriram, Olivia, Ken, Christina, and Ezgi, who shared their passions and histories with me, accompanied me on adventures in Boston and around the world, and offered their comradery and support. I'm grateful for Alden and Tim and our adventures in Boston and Provincetown. I'm thankful to Efren for sharing his passions with me and standing by me. I'm thankful to Rick for sharing his love, support, passion for artistic pursuits, and absurd view of the world with me.

Lastly I am thankful for my family for their never-ending support. My parents never stop supporting me and sharing their pride and love. I'm happy to have been in Boston with my sister Laura and her wife Sarah, who were always there to share a fun time and give me some support. I'm thankful for my little sister Emma for being a source of joy and pride in my life.

Contributions

Chapter 1

Contributions: Joseph Robert Piechura (JRP) wrote the text.

Chapter 2

Contributions: Joseph S. Markson (JSM) and Erin K. O'Shea (EKO) conceived the project. JSM, JRP, and Ania M. Puszynska (AMP) designed and carried out the experiments. JSM, JRP, AMP and EKO wrote the manuscript.

Chapter 2 has previously been published with minor changes as the following paper:

Markson, J.S.*, Piechura, J.R.*, Puszynska, A.M. & O'Shea, E.K. (2013). Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* doi: 10.1016/j.cell.2013.11.005 *These authors contributed equally to this work

Chapter 3

Contributions: JRP conceived the project, designed and carried out experiments, analyzed data, contributed to mathematical modeling, and wrote the manuscript. Kapil Amarnath (KA) constructed the light apparatus, contributed to experimental design and overall direction, carried out mathematical modeling, and edited the manuscript. EKO contributed to experimental design and edited the manuscript.

Chapter 4

Contributions: JRP wrote the text.

Chapter 1

Introduction

1.0.1 Understanding the mechanisms behind outputs of the circadian clock

Organisms in nature must maintain homeostasis to survive in the face of constantly changing environmental conditions. Some of these environmental changes are caused by highly periodic events such as the rotation of the Earth. To maximize the ability to maintain homeostasis in the face of predictable environmental changes, organisms from all branches of life have evolved self-sustaining oscillators —known as circadian clocks —which provide an internal representation of the time-of-day [16]. Information in a *core oscillator* is interpreted by *output pathways* that alter physiology to prepare for the changes of the day/night cycle [16]. However, it is often not clear how information from the core oscillator is converted into a relevant physiological output.

Circadian clocks exert control over a wide range of organismal functions. For example, the circadian clock in fruit flies like *Drosophila melanogaster* regulates the sleep/wake cycle to synchronize these behaviors with the day night cycle [13], while the clock in plants like *Aradbiposis thaliana* controls the opening and closing of gas exchange organs to maximize photosynthesis and minimize water loss in the face of the light/dark cycle [25]. However,

the complex nature of circadian systems in higher organisms makes it difficult to dissect how core information on time-of-day is converted by output pathways. In organisms with a nervous system like *Drosophila melanogaster*, a large group of neurons with internal biochemical clocks interface with output neurons to communicate timing information to control behavior [13]. Further, core oscillators within cells of eukaryotic organisms like *Drosphila* and *Arabidopsis* involve complex interactions between multiple regulators which act multiple in cellular compartments [13, 25].

The cyanobacterium *Synechococcus elonagatus* PCC7942 has been an invaluable model organism for the study of circadian biology due to the relative simplicity of its circadian clock. First, timing information in this single-celled photoautroph is encoded in a simple core oscillator composed of only three proteins —KaiA, KaiB, and KaiC [47]. Interactions between these three proteins generate oscillations in the phosphorylation state of KaiC with a period of roughly 24 hours to provide information on the time of day to cyanobacteria [47]. This biochemical clock can be reconstituted in vitro using just the three Kai proteins, demonstrating that information required to generate timing information is encoded in the sequence of only a few proteins [47]. Further, the output of the cyanobacterial clock is quite simple —when synchronized to Light/Dark conditions and released into constant conditions, *S. elongatus* exhibits oscillations in the expression of a large portion of its genome [27, 90]. Also, this output has at least one known role for the organism —clock output is required for survival of cyanobacteria under light/dark conditions in order to prepare the bacteria for the metabolic constraints of night time [10, 67, 82]. Thus, in *S. elongatus* we can readily ask the question —how is information in the KaiABC core oscillator used by output pathways to control physiologically relevant genome-wide transcriptional rhythms?

Prior to the work presented in this dissertation, it was unclear how genome-wide tran-

scriptional rhythms are generated in cyanobacteria. Genetic screening had identified the gene *rpaA*, encoding an OmpR-type response regulator protein, as critical for oscillatory expression from some promoters under circadian-clock control [82], but the role of this gene in regulating genome-wide expression rhythms was unclear. Further, the mechanism of action of RpaA was unknown. OmpR-type response regulator proteins are typically phosphorylated on an aspartate residue by proteins known as histidine kinases [18]. Phosphorylation-dependent regulation of OmpR-type proteins allows them to directly regulate gene expression by binding promoters [18]. It was known that two histidine kinases —SasA and CikA —are regulated by the core KaiABC oscillator to generate oscillations in phosphorylation of RpaA (RpaA∼P) under constant conditions [21, 82]. However, the role of phosphorylation in regulating RpaA activity was unclear [22]. In Chapter 2, we explored the role of RpaA in regulating genome-wide transcriptional rhythms in *S. elongatus*. We find that levels of phosphorylated RpaA drive changes in almost all circadian genes, demonstrating that a single clock-controlled regulator can have pervasive control of gene expression. Further, we show that phosphorylation of RpaA regulates its activity by allowing it to bind directly to DNA to regulate the expression of some clock-controlled genes.

1.0.2 Exploring the interaction of circadian clock outputs with dynamic environments

The mechanisms of circadian clock function are typically studied under constant conditions, when changes to an organism are solely regulated by the circadian clock. However, constant conditions do not exist in nature. The rotation of the Earth imparts daily periodic variation in light intensity and temperature. Weather patterns cause changes in humidity, precipitation, and light intensity. Further, the orbit of the Earth around the sun causes

seasonal changes in temperature and day length. Thus, circadian clocks in nature function in highly dynamic environments.

Organisms integrate information about the environment with outputs from the circadian clock the regulate their physiology in response to the dynamic conditions of nature. The flowering plant *Arabidopsis thaliana* integrates output from the circadian clock with information about the length of day (photoperiod) to ensure that flowering only occurs when days have reached a specific length indicative of spring [76]. The *Arabidopsis* clock generates daily oscillations in the abundance of the mRNA encoding the flowering regulator CONSTANS, peaking at the end of the day [76]. Light-responsive pathways regulate CONSTANS to ensure that functional CONSTANS protein only accumulates in light [76]. Thus, CONSTANS only induces flowering if light remains until the end of the day, when the clock activates CONSTANS mRNA expression [76]. The combination of circadian and environmental regulation allows plants to time their flowering to seasonal changes in day length. Further, environmental responses can interact with clock-regulated factors in *Drosophila* to modulate activity like locomotion in response to dynamic environments [14].

It is unknown how the circadian clock in cyanobacteria will interact with changes in the environment to control physiology. Cyanobacteria are attractive models for bio-engineering to produce useful products from ambient sunlight [15], but optimal engineering of these organisms must incorporate knowledge of how different regulatory systems within the organism interact in natural environments [51, 93].

Light intensity is a critical variable for an organism like *S. elongatus* which produces all of its energy from photosynthesis. Further, light intensity varies in predictable and unpredictable ways in nature, and these changes cause changes to rates of photosynthesis [49]. As such, cyanobacteria respond to changes in light intensity by regulating photosynthetic

machinery and changing expression of genes to help maintain homeostasis [2, 56]. Among the best known light-responsive systems in cyanobacteria is the NblS-RpaB two component system, which enacts changes in gene expression in response to increases in light intensity (High light stress) [34]. Interestingly, several studies have suggested a link between the light-responsive NblS-RpaB two component system, and the circadian clock-regulated CikA/SasA-RpaA two component system [17, 22].

It is unclear how changes in an environmental variable such as light intensity will affect the transcriptional output of the circadian clock. It is possible that the clock promotes homeostasis in cyanobacteria by producing the same gene expression pattern every day, regardless of changes in the environment. However, given that many dusk genes directly regulated by the clock are involved in alternative metabolism [27, 48, 67, 90], it is likely that environmentally responsive systems also regulate the expression of these genes in natural environments. Thus, we decided to build upon our understanding of RpaA-based regulation of circadian gene expression (Chapter 2) and explore how this system interacts with environmental changes in light intensity to unveil principles governing the interaction of clock output with environmental changes in cyanobacteria (Chapter 3). By simulating changes in light intensity that occur in nature, we find that the transcriptional output of the circadian clock is modulated by dynamic light conditions, at least in part by modulating RpaA activity independent of the clock. Further, we find that light-responsive changes in RpaB activity are involved in controlling circadian gene expression, potentially by interacting with RpaA.

Chapter 2

Circadian control of global gene expression by the cyanobacterial master regulator RpaA

2.1 Abstract

The cyanobacterial circadian clock generates genome-wide transcriptional oscillations and regulates cell division, but the underlying mechanisms are not well understood. Here we show that the response regulator RpaA serves as the master regulator of these clock outputs. Deletion of *rpaA* abrogates gene expression rhythms globally and arrests cells in a dawn-like expression state. Although *rpaA* deletion causes core oscillator failure by perturbing clock gene expression, rescuing oscillator function does not restore global expression rhythms. We show that phosphorylated RpaA regulates the expression of not only clock components, generating feedback on the core oscillator, but also a small set of circadian effectors that in turn orchestrate genome-wide transcriptional rhythms. Expression of constitutively active

RpaA is sufficient to switch cells from a dawn-like to a dusk-like expression state as well as to block cell division. Hence, complex global circadian phenotypes can be generated by controlling the phosphorylation of a single transcription factor.

2.2 Introduction

The circadian clock of the cyanobacterium *Synechococcus elongatus* PCC7942 drives daily genome-wide oscillations in mRNA expression levels, controls genome compaction and supercoiling, and modulates cell division [29]. The clock contains a core oscillator consisting of the proteins KaiA, KaiB, and KaiC, which together generate circadian (i.e., approximately 24-hour) oscillations in KaiC phosphorylation [47]. Remarkably, the KaiC phosphorylation oscillations observed in vivo can be reconstituted in vitro simply by mixing the three Kai proteins and ATP [58]. In vivo, this proteinaceous post-translational oscillator (PTO) is embedded in a transcription-translation feedback loop (TTL) that regulates expression of the *kaiBC* operon, enhancing the precision of the clock by stabilizing its phase [29,68,86,97].

Although much is known about the mechanism by which the PTO keeps time, less is understood about how the clock uses the time information encoded in the PTO to generate outputs like global gene expression rhythms and modulation of cell division. Most genes show circadian expression oscillations in constant light, displaying a variety of amplitudes, phases, and waveforms [27, 29, 90]. The distribution of phases is biphasic, with expression of one population peaking around subjective dusk (class 1) and the other peaking around subjective dawn (class 2). ('Subjective dusk' and 'subjective dawn' refer to the times at which light-to-dark or dark-to-light transitions would occur in a 12 h light-12 h dark environmental cycle. The term 'subjective' is used when, as here, environmental conditions are held constant in order to isolate clock-driven from environmentally-driven processes.)

Genome compaction and DNA supercoiling also oscillate in a circadian manner [80, 94], and supercoiling oscillations contributes to the generation of gene expression oscillations [90]. Finally, the clock gates cell division, prohibiting it during the subjective night [12, 53, 95].

Genetic and biochemical approaches have revealed key players in the output pathway that connects the PTO to these output responses. The response regulator RpaA has been implicated genetically in circadian gene expression control [82, 84]. The phase of the PTO is transmitted to RpaA via the histidine kinases SasA and CikA [21,82,85]: in vivo, SasA phosphorylates and CikA dephosphorylates RpaA via their respective kinase and phosphatase activities. The PTO generates temporal separation of SasA and CikA activities, producing circadian oscillations in RpaA phosphorylation levels [21]. The response regulator RpaB, a paralog of RpaA, also may play a role in circadian clock output: it binds in vitro and in vivo to the promoters of several circadian genes, and this binding is antagonized by RpaA in vitro [22].

Deletion of *rpaA* eliminates oscillations in the activity of the ∼10 circadian promoters that were assayed by bioluminescence reporters [82], but the role of RpaA in regulating circadian expression genome-wide is not known. RpaA is predicted to be a transcription factor [82], but recent studies failed to detect binding of RpaA to candidate target promoters [22, 82]. Similarly, cell division gating by the clock requires RpaA [12], but the mechanism is not understood.

Here we employ a multifaceted approach to elucidate the molecular and functional roles of RpaA in circadian clock output. We find that circadian gene expression oscillations are absent genome-wide in an *rpaA* mutant, with cells being arrested in a subjective dawn-like transcriptional state. Through chromatin immunoprecipitation with high-throughput sequencing (ChIP-Seq) and in vitro DNase I footprinting, we show that phosphorylated

RpaA binds directly to over one hundred locations in the genome, including the promoter of *kaiBC*. Finally, we show that overexpression of a phosphomimetic mutant of RpaA is sufficient to drive cells from the subjective dawn to the subjective dusk gene expression state and also to close the cell division gate, demonstrating that RpaA is the global regulator of circadian output in this organism.

2.3 Results

2.3.1 *rpaA* is required for global circadian gene expression and its deletion arrests cells in a subjective dawn-like state

To determine whether RpaA is required for global gene expression rhythms, we measured gene expression over 48 hours by microarray in an *rpaA* mutant strain (Figure 2.1). We found that circadian oscillations were abolished genome-wide (Figure 2.1A), even for the genes that oscillate with the highest amplitude in the wild-type (Figure 2.2). Hence, RpaA is required for the generation of global circadian gene expression rhythms.

To gain insight into the role of RpaA in circadian gene regulation, we searched for genes which showed the greatest magnitude of transcript level difference between the wild-type and *rpaA* mutant strains. Consistent with previous results [82], *kaiBC* expression decreased approximately 3.5-fold, while *kaiA* expression was not affected substantially; this disparate effect on kai gene expression perturbs Kai protein stoichiometry, likely situating it in a regime that does support PTO function [58]. Overall, we found that the expression of 67 genes decreased more than two-fold, while the expression of 16 genes increased by at least that amount. Strongly downregulated genes included four sigma factors, two transcription factors, and several genes encoding proteins involved in energy production and metabolism

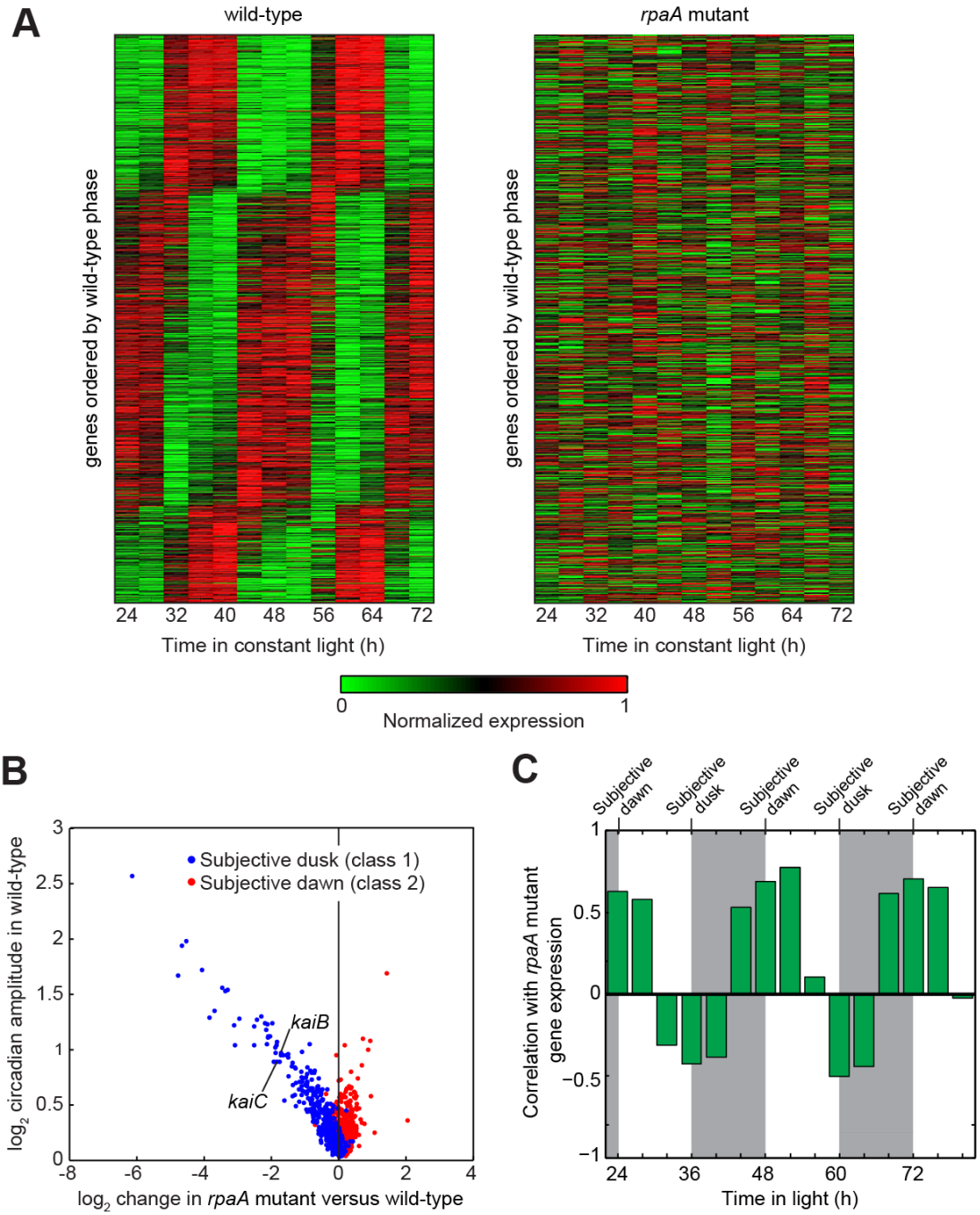

Figure 2.1: Gene expression is globally perturbed by deletion of *rpaA*. A. Global gene expression timecourse in the wild-type and *rpaA* mutant. (Left) Circadian gene expression in the wild-type strain (data from [90]. Expression timecourses of genes reproducibly oscillating with a circadian period (n = 856; see Methods) were normalized to the interval [0 1] and sorted by phase. (Right) Gene expression in the *rpaA* mutant. Expression timecourse of the same set of genes as in wild-type, displayed in the same order. B. Comparison of gene expression change in the *rpaA* mutant with circadian amplitude in the wild-type strain. We computed the expression change in the *rpaA* mutant by comparing the average *rpaA* mutant expression over one day with the average wild-type expression over one day (see Methods). Only genes that oscillate with circadian periodicity in the wild-type strain are shown (n = 856). *kaiB* and *kaiC* are indicated; *kaiA* is not classified as circadian and hence is not shown. C. Correlation of global gene expression in the *rpaA* mutant with each timepoint in the wild-type timecourse. Expression of all genes (both circadian and non-circadian) in the *rpaA* mutant was time-averaged as in B, and the correlation between this time-averaged expression and the expression in the wild-type strain at each timepoint over a 60-h timespan was computed. Wild-type data are from [90]. Subjective night is shaded. Subjective dawn occurs at 24, 48, and 72 h, while subjective dusk falls at 36 and 60 h.

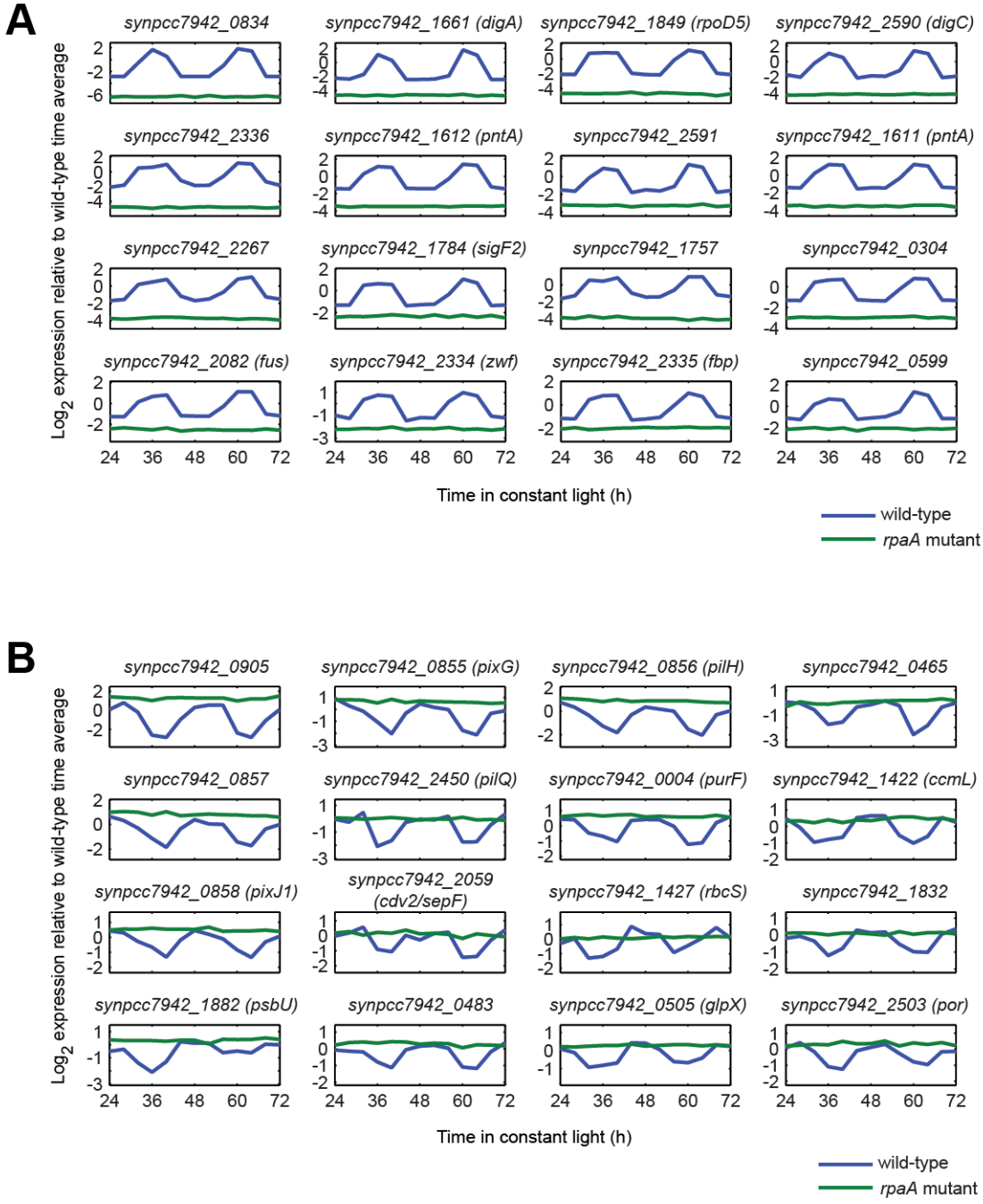

Figure 2.2: *rpaA* mutant gene expression timecourse. A. Wild-type and *rpaA* mutant gene expression timecourses for the highest-amplitude wild-type subjective dusk (class 1) genes. The expression timecourses of the sixteen highest amplitude subjective dusk genes in the wild-type strain are shown in blue; expression at each timepoint is reported relative to the wild-type time average ([90]). The timecourses of those same genes in the *rpaA* mutant strain are shown in green, also relative to the wild-type time average (see Methods). Note that the overall expression level is significantly lower in the *rpaA* mutant than in the wild-type strain for all 16 genes. Gene expression was measured by microarray. B. Wild-type and *rpaA* mutant gene expression timecourses for the highest-amplitude wild-type subjective dawn (class 2) genes. This plot is constructed in the same manner as in A but using the highest-amplitude wild-type subjective dawn genes. Note that the overall expression level generally is higher in the *rpaA* mutant than in the wild-type strain. Gene expression was measured by microarray.

(particularly carbohydrate metabolism). Circadian genes downregulated in the *rpaA* mutant were highly enriched for subjective dusk phasing, whereas upregulated genes were enriched for subjective dawn phasing (Figure 2.1B). Interestingly, the decrease in expression of subjective dusk genes in the *rpaA* mutant is directly proportional to the gene's circadian amplitude in the wild-type strain Figure 2.1B). These observations suggest that RpaA is responsible for promoting subjective dusk gene expression and repressing subjective dawn gene expression. Consistent with this scenario, global gene expression in the *rpaA* mutant is most positively correlated with wild-type subjective-dawn expression and most negatively correlated with wild-type subjective dusk expression Figure 2.1C). Hence, deletion of *rpaA* arrests cells in a subjective dawn-like state.

2.3.2 Rescue of PTO function in the *rpaA* mutant reveals that RpaA is directly responsible for the orchestration of circadian gene expression

Interpretation of the role of RpaA is confounded by the loss of PTO function in the *rpaA* mutant [82]: the absence of gene expression oscillations in the *rpaA* mutant could merely be a secondary effect of the loss of PTO function rather than an indication of a master regulator role for RpaA. Indeed, it is possible that the primary role of RpaA is simply to sustain Kai post-translational oscillations by, for example, modulating *kaiBC* expression to maintain permissive Kai protein stoichiometry.

To distinguish between direct and indirect contributions of RpaA to circadian gene expression, we asked whether rescuing the PTO in an *rpaA* mutant background would restore global gene expression oscillations. We rescued the PTO by ectopically expressing *kaiBC* from the IPTG-inducible Ptrc promoter [57] in a *rpaA*- *kaiBC*- background (*rpaA*- *kaiBC*-

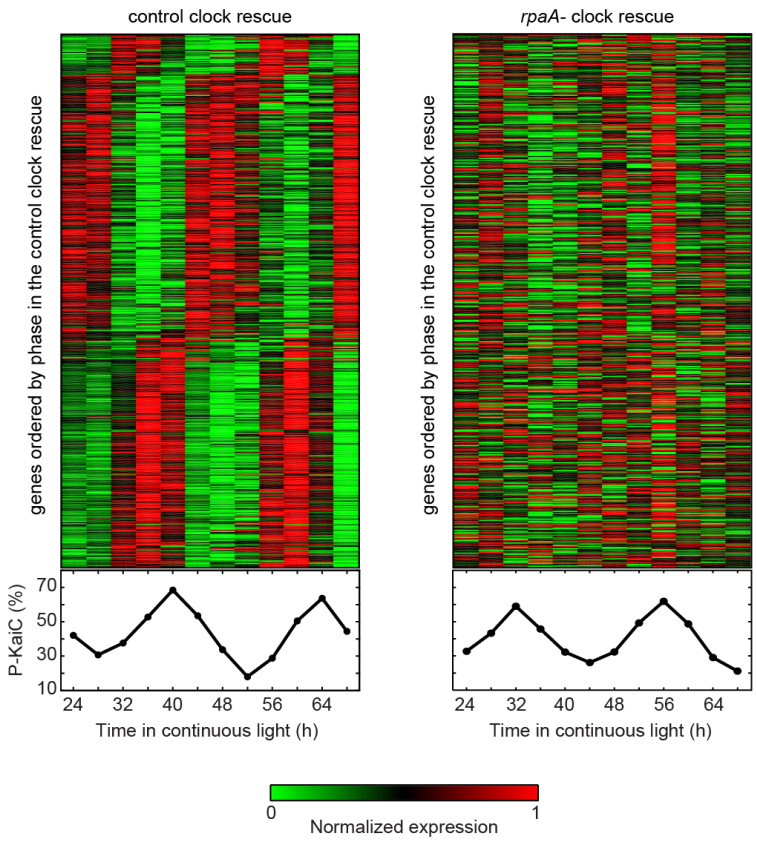

Figure 2.3: RpaA is required for control of global gene expression by the PTO. (Left) Gene expression timecourse upon rescue of Kai oscillator function in a *kaiBC*- background via ectopic expression of *kaiBC* (*kaiBC*- Ptrc::*kaiBC*, termed the 'control clock rescue'). The heatmap shows the expression timecourse of the 471 genes that oscillate with a circadian period in the control clock rescue in each of two biological replicates, with genes normalized individually and ordered by phase. KaiC phosphorylation levels during the timecourse are shown below the heatmap. (Right) Gene expression timecourse upon rescue of Kai oscillator function in the absence of rpaA via ectopic expression of kaiBC (*rpaA*- *kaiBC*- Ptrc::*kaiBC*, termed the '*rpaA*- clock rescue'). The heatmap shows the expression timecourse of the same set of genes as in (A), displayed in the same order. KaiC phosphorylation levels during the timecourse are shown below the heatmap.

Ptrc::*kaiBC*), and, as a control, in a *kaiBC*- background (*kaiBC*- Ptrc::*kaiBC*). We refer to these strains as the *rpaA*- clock rescue and the control clock rescue, respectively. In both strains, we obtained strong KaiC phosphorylation rhythms of similar amplitude in the presence of 6 μM IPTG (Figures 2.3A and 2.4A). Note that ectopic expression of KaiA was not necessary because its levels are not substantially perturbed by deletion of *rpaA* [82].

We used microarrays to examine global gene expression dynamics in the two rescue strains (Figure 2.3). In the control clock rescue, circadian gene expression oscillations were restored robustly for both subjective dusk and subjective dawn genes (Figure 2.3, 2.4B, and 2.4C). In contrast, no strong oscillations were observed in either class of genes in the *rpaA*- clock

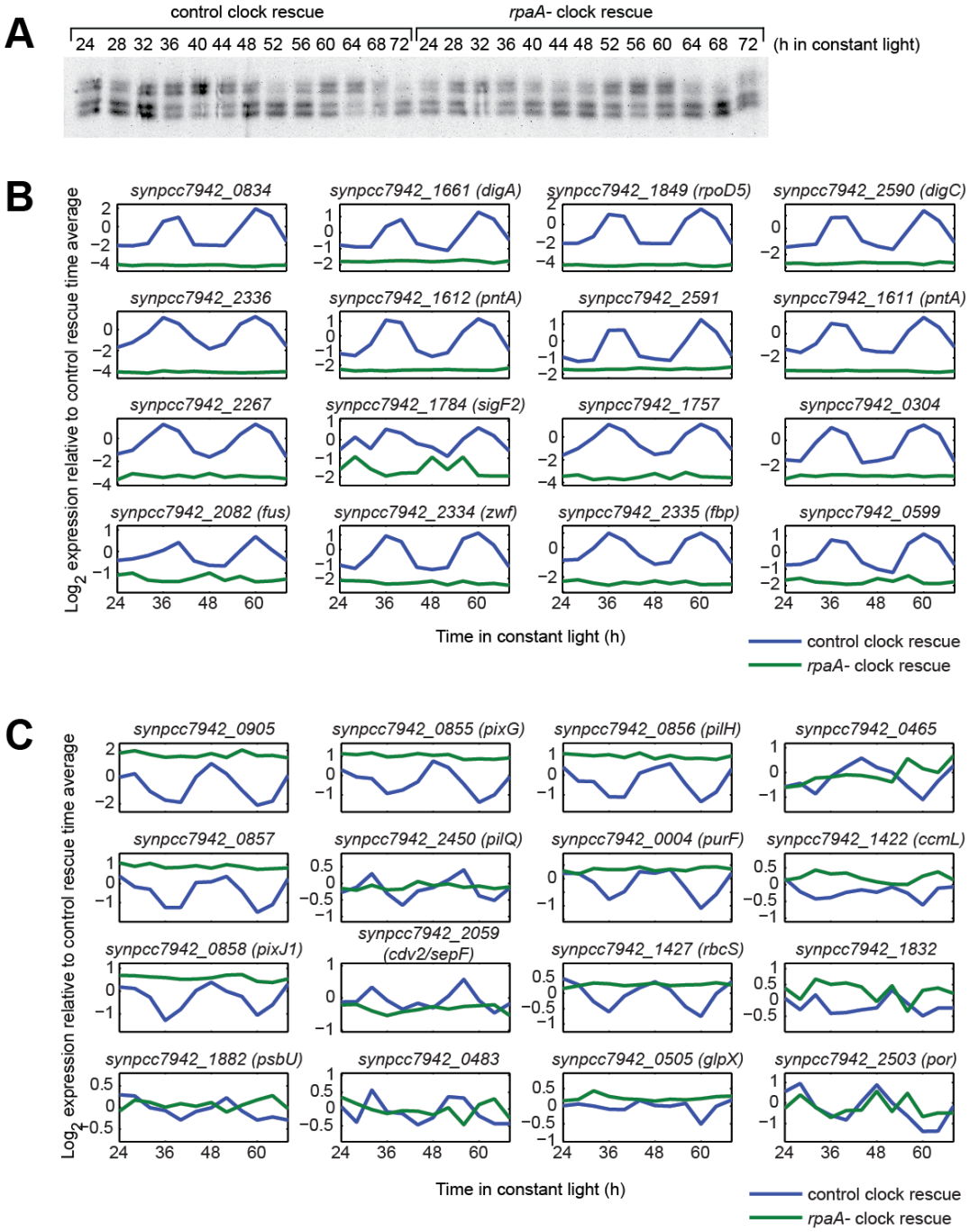

Figure 2.4: Rescue experiments. A. Western blot of KaiC phosphorylation for the control and *rpaA*- clock rescue timecourses shown in Figure 2.3. Quantification of these blots is shown below the heatmaps in Figure 2.3. Samples for Western blotting were acquired contemporaneously with those used for microarray analysis. Samples were lysed in urea lysis buffer, and equal amounts of total protein from each lysate were loaded onto a 4-20% SDS-PAGE gradient gel for Western blot analysis. See Methods for details. B. Microarray-based gene expression timecourses in the rescue experiments (Figure 2.3) for the subjective dusk (class 1) genes with the highest amplitude in the wild-type strain (same set as in Figure 2.2A). Expression at each timepoint in both strains is reported relative to the control clock rescue time average. The control clock rescue time average was prepared in the same manner as the wild-type time average in Figure 2.2 (see Methods), except that the dye-swap experiment was not performed. C. Microarray-based gene expression timecourses in the rescue experiments (Figure 2.3) for the subjective dawn (class 2) genes with the highest amplitude in the wild-type strain (same set as in Figure 2.2B). This plot is constructed in the same manner as in A but using the highest-amplitude subjective dawn genes in the wild-type strain.

rescue (Figure 2.3, 2.4B, and 2.4C). We conclude that RpaA is required for the orchestration of robust circadian gene expression by the PTO, ruling out the scenario in which RpaA's role is limited to supporting oscillator function.

2.3.3 Phosphorylated RpaA binds to the *kaiBC* promoter and upregulates *kaiBC* expression

While the rescue experiments demonstrate that RpaA plays a critical and central role in producing global gene expression rhythms, they do not indicate the means by which it does so. As previous studies have failed to identify direct RpaA binding to DNA, it has been proposed that RpaA acts indirectly via displacement of RpaB from circadian promoters [22]. However, these studies have assayed only a few regions for RpaA binding, and it is possible that RpaA acts via rhythmic binding in conditions or at locations that have not been examined.

To determine whether and where RpaA associates with genome, we generated an anti-RpaA antibody (Figure 2.6A) and used it to perform chromatin immunoprecipitation (ChIP) over a 36 h circadian timecourse. We found that RpaA localized to the kaiBC promoter in an oscillatory manner, in phase with both RpaA phosphorylation (RpaA∼P) and expression of the kaiBC transcript (Figures 2.5A and 2.6B). These results are not an artifact of off-target binding of the anti-RpaA antibody, as ChIP-qPCR analysis of a strain expressing epitope-tagged RpaA (HA-RpaA) using an anti-HA antibody demonstrated that it localized to the kaiBC promoter with a circadian period and in phase with HA-RpaA phosphorylation (Figure 2.6C).

To determine whether localization of RpaA to the *kaiBC* promoter results from direct association with DNA, we assayed for physical interaction of RpaA with promoter DNA in vitro using DNase I footprinting. We observed a clear footprint, strictly dependent on RpaA

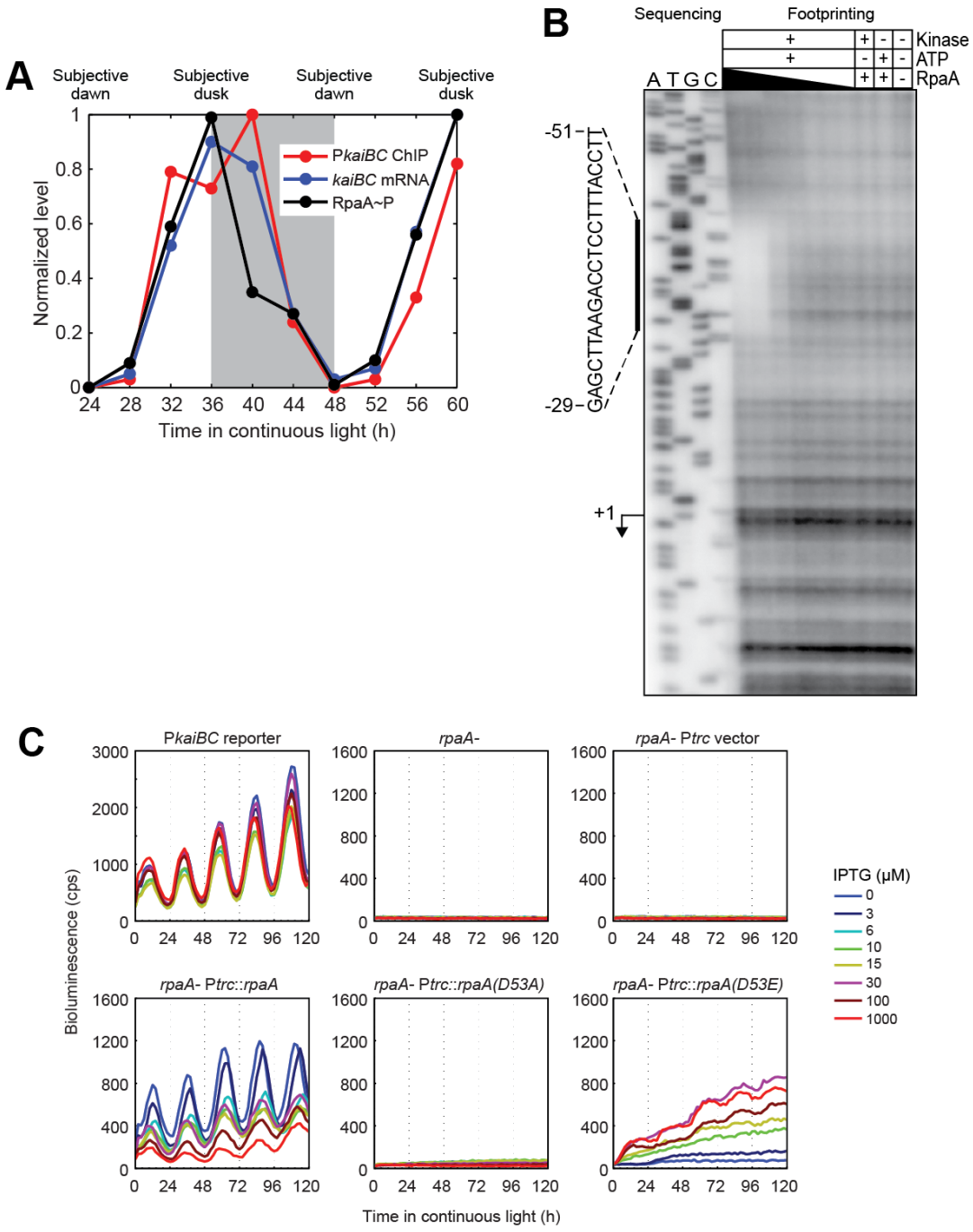

Figure 2.5: RpaA binds to the *kaiBC* promoter in vivo and in vitro and promotes *kaiBC* expression in a phosphorylation-dependent manner. A. Correlation between RpaA phosphorylation, RpaA enrichment at the *kaiBC* promoter, and abundance of the *kaiBC* transcript. Subjective night is shaded. RpaA phosphorylation was measured by Phos-tag Western blot (see Figure S3B for gel image), association with P*kaiBC* by ChIP-qPCR, and *kaiBC* expression by RT-qPCR. Note that while the apex of the ChIP-qPCR enrichment is at 40 h (four hours after subjective dusk) in this experiment, the precise phase of RpaA (as well as that of global gene expression oscillations) varies between experiments, with RpaA binding typically peaking at or a few hours before subjective dusk. B. In vitro DNase I footprinting of RpaA on the *kaiBC* promoter as a function of recombinant RpaA phosphorylation and concentration. Sanger sequencing reactions used to identify the location of the footprint are shown on the left; footprinting reactions are shown on the right. The region protected from digestion by RpaA∼P is indicated by the vertical bar. The *kaiBC* transcription start site [38] is indicated with an arrow. RpaA pre-treatment and concentration are indicated above each footprinting lane. RpaA was at least 50% phosphorylated in the presence of both CikA and ATP but was unphosphorylated otherwise (Figure 2.6D). RpaA was added to a final concentration of 6.0, 3.0, 0.6, 0.3, 0.06, or 0.03 μM as indicated by the thickness of the wedge. C. Activity of the *kaiBC* promoter was assayed using a P*kaiBC*::luxAB luciferase reporter in various genetic backgrounds: wild-type, *rpaA*-, and *rpaA*- expressing wild-type RpaA, unphosphorylatable RpaA (D53A), or an RpaA phosphomimetic (D53E) from the IPTG-inducible Ptrc promoter. IPTG was added at the indicated concentration prior to entrainment with two 12-h dark pulses.

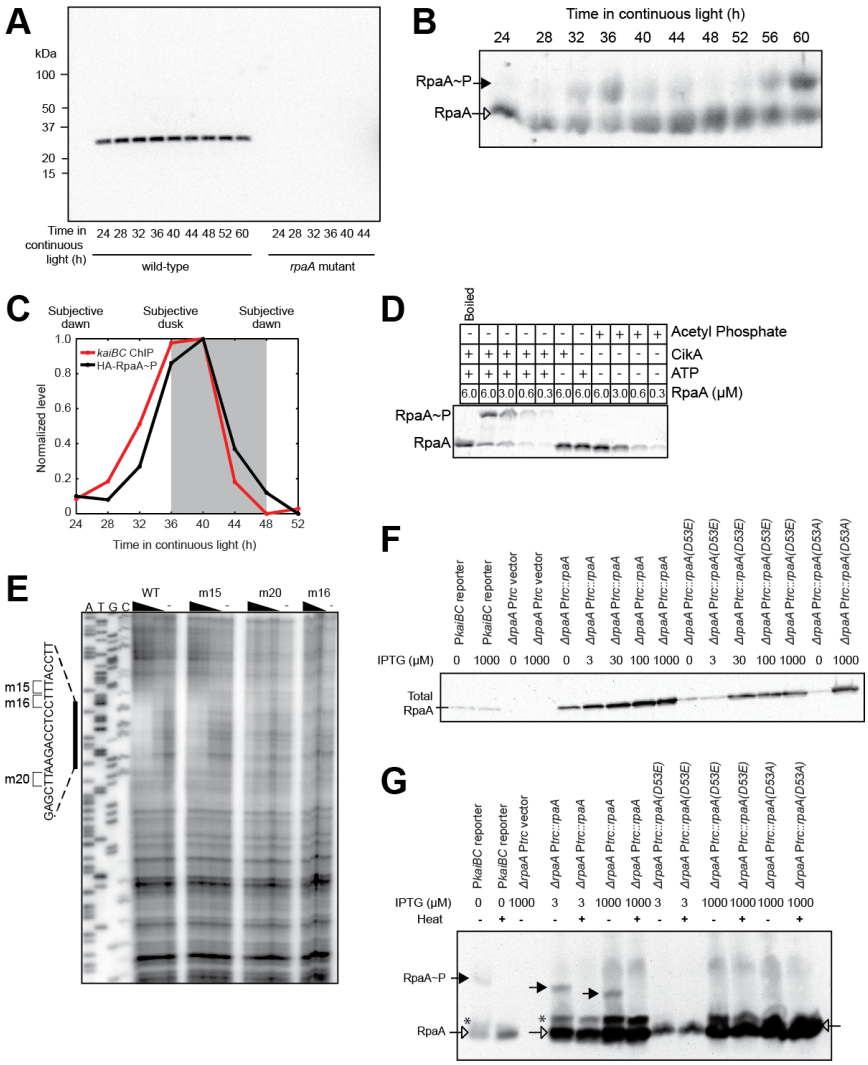

Figure 2.6: RpaA binds to the *kaiBC* promoter in vivo and in vitro and promotes *kaiBC* expression in a phosphorylation-dependent manner (continued). A. Analysis of affinity-purified anti-RpaA antibody specificity by Western blot. Samples from wild-type and *rpaA* mutant circadian timecourses were blotted with our anti-RpaA antibody. B. Western blot analysis of the phosphorylation state of RpaA for Figure 2.5A. Closed, solid black arrows, phosphorylated RpaA (RpaA∼P); open arrows, unphosphorylated RpaA. Equal masses of total protein from each lysate were separated on a Phos-tag gel and analyzed by Western blotting [21]. C. Correlation between HA-RpaA phosphorylation and HA-RpaA enrichment at the *kaiBC* promoter. RpaA phosphorylation was measured by Phos-tag Western blot [21], while association with PkaiBC was measured by ChIP-qPCR performed with anti-HA antibody (see Methods). Subjective night is shaded in gray. D. Phosphorylation of RpaA used in DNase I footprinting reactions. RpaA was incubated DNase I footprinting reaction buffer for 1 hour at 30 C with kaiBC footprinting probe (end-labeled with non-radioactive ATP) in the presence of 1.5 μM recombinant CikA, 1 mM ATP, and/or 10 mM lithium potassium acetyl phosphate (Sigma Aldrich), as indicated. After incubation, 2.5 μL of each of these reactions was resolved on a Phos-tag SDS-PAGE gel and visualized by staining with Sypro Ruby as described in the Methods. The identity of the band corresponding to phosphorylated RpaA (RpaA∼P) was revealed by boiling a reaction(reaction is marked with 'Boiled'). RpaA phosphorylation was observed only when CikA and ATP were present. E. DNase I footprinting of RpaA on *kaiBC* promoter mutants that produce diminished expression compared to the wild-type promoter [38]. *kaiBC* promoter mutants m15, m16, and m20 contain T to A or G to C mutations at the positions indicated in the sequence next to the vertical bar, which marks the region of the wild-type promoter protected from digestion by high levels of RpaA∼P. Sanger sequencing reactions used to identify the location of the footprint are shown on the left; footprinting reactions are shown on the right. All reactions contained 1.5 μM CikA and 1 mM ATP. Reactions marked with a '-' did not contain RpaA, while reactions marked with a wedge contained 0.6, 3, or 6 μM RpaA as indicated by the thickness of the wedge. For m16, the reaction containing 0.6 μM was not run on the gel.

phosphorylation, between 29 and 51 bp upstream of the transcription start site (Figure 2.5B). The strict phosphorylation dependence may explain the inability to detect RpaA binding to the kaiBC promoter in previous in vitro studies [22, 82]: these studies used either unphosphorylated RpaA [82] or RpaA putatively phosphorylated by treatment with acetyl phosphate [22], which we found does not actually result in RpaA phosphorylation (Figure 2.6D). Intriguingly, the footprint of RpaA∼P in the *kaiBC* promoter coincides with a region in which mutations substantially reduce promoter activity in vivo [38]. We tested three of those mutations in the footprinting assay, finding that all of them reduced or eliminated RpaA∼P binding (Figure 2.6E).

To test whether RpaA∼P drives expression of the *kaiBC* transcript in vivo, we expressed phosphorylation site mutants mimicking phosphorylated or unphosphorylated RpaA in an *rpaA* mutant background and used a P*kaiBC*-driven luciferase reporter to assess the effect of these mimetics on P*kaiBC* activity (Figure 2.5C). Aspartate 53 (D53) of RpaA is predicted by sequence homology with the well-studied E. coli OmpR protein to be the site of phosphorylation by SasA, and we therefore used a glutamate mutation at residue 53 (D53E) to mimic RpaA∼P and an alanine mutation (D53A) to mimic unphosphorylated RpaA. All of the RpaA variants were expressed in the presence of IPTG (Figure 2.6F). We observed no expression of the luciferase reporter in a control strain containing the Ptrc promoter without RpaA. Leaky expression of wild-type RpaA from the Ptrc promoter (Figure 2.6F) rescued circadian activity of the kaiBC reporter (Figure 2.5C). Consistent with a previous report [84] increasing levels of wild-type RpaA with IPTG induction progressively repressed promoter activity, likely due to an increase in the ratio of unphosphorylated RpaA to RpaA∼P (Figure 2.6G). In contrast, *kaiBC* promoter activity was absent at all doses of RpaA(D53A). Expression of RpaA(D53E), however, restored activity from the promoter in a dose-dependent

manner (Figure 2.5C), consistent with positive regulation of promoter activity by RpaA∼P. No aspartate phosphorylation was present in RpaA(D53A) or RpaA(D53E) (Figure 2.6G). Collectively, our observations suggest that rhythmic association of RpaA∼P with the *kaiBC* promoter drives circadian expression of this operon. Therefore, the circadian TTL is directly mediated by RpaA.

2.3.4 ChIP-Seq reveals the landscape of RpaA binding

To identify RpaA binding sites genome-wide, we used our anti-RpaA antibody to perform circadian timecourse ChIP analyzed by high throughput sequencing (ChIP-Seq). We identified 110 binding sites (peaks), all located on the main chromosome (see Methods) (Figure 2.7A). A well-defined binding site was identified upstream of the kaiBC locus, the occupancy of which varied with circadian time (Figure 2.7B). RpaA binds in a circadian manner and with a similar phase at all 110 binding sites (Figure 2.7C).

To ensure that the enrichment we observed did not reflect off-target binding of the anti-RpaA antibody, we performed a ChIP-Seq experiment using the HA-RpaA strain and the anti-HA antibody. The genome-wide HA-RpaA binding profile (Figure 2.8A) resembled that of wild-type RpaA, although enrichments were generally substantially lower (Figure 2.8B), consistent with the lower overall phosphorylation of HA-RpaA (Figure 2.8C). We found that 66 out of the 110 wild-type binding sites (60%) were also present in the HA-RpaA ChIP experiment at a minimum of 2-fold enrichment (Figure 2.8D).

To determine whether RpaA can bind directly to DNA at locations other than the *kaiBC* promoter, we performed in vitro DNase I footprinting on a strongly-enriched region upstream of the gene encoding the sigma factor RpoD6. RpaA bound to the *rpoD6* promoter in a phosphorylation- and concentration-dependent manner (Figure 2.8E).

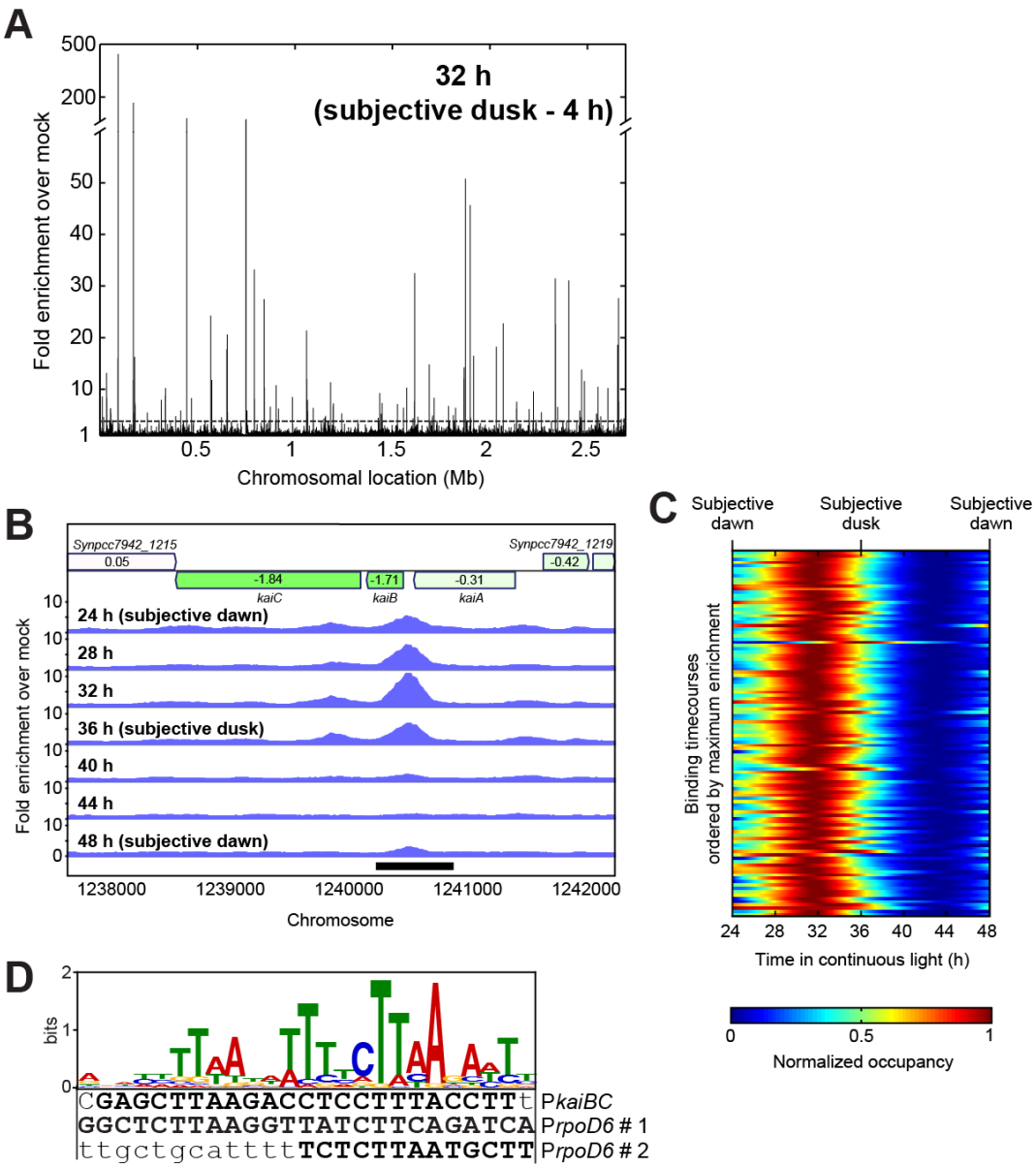

Figure 2.7: Identification of RpaA binding sites by ChIP-Seq. A. Genome-wide binding profile of RpaA by ChIP-Seq. The enrichment of read density in the RpaA ChIP-Seq (anti-RpaA antibody on the wild-type strain) at four hours prior to subjective dusk (32 h), relative to the mock ChIP-Seq (anti-RpaA antibody on the rpaA mutant), is plotted as a function of position on the chromosome. The dotted line indicates the 3-fold enrichment cutoff for identification of RpaA binding sites. B. Genome browser view of ChIP-Seq enrichment profiles in the vicinity of the *kaiBC* locus over one day in the wild-type strain. Genes are shown at the top, with the log2 of their expression change in the *rpaA* mutant represented by shading (green is decreased expression and red is increased expression) and indicated by the text inside the gene. C. Timecourse of RpaA enrichment at the 110 RpaA binding sites. ChIP-Seq was performed every four hours for 24 h, and enrichment relative to the mock IP was calculated at the location of maximum wild-type ChIP-Seq read density within each binding site. Enrichment at intermediate timepoints was computed by interpolation with cubic splines. Each row in the heatmap represents the binding timecourse for one binding site; the rows are sorted by the maximum enrichment observed during the timecourse, which ranged from 411-fold (top) to 3.1-fold (bottom). The dynamic range (maximum enrichment divided by minimum enrichment for each binding site) varied from 38-fold to 1.4-fold. D. A 25-basepair motif is overrepresented near RpaA binding sites (E-value, 1.7 x 10-36) and is found within the RpaA∼P footprint in the *kaiBC* promoter (Figure 2.5B) and in both footprints in the textitrpoD6 promoter (Figure 2.8E). Bases protected by RpaA∼P binding in the DNase I footprinting assays are capitalized and boldfaced.

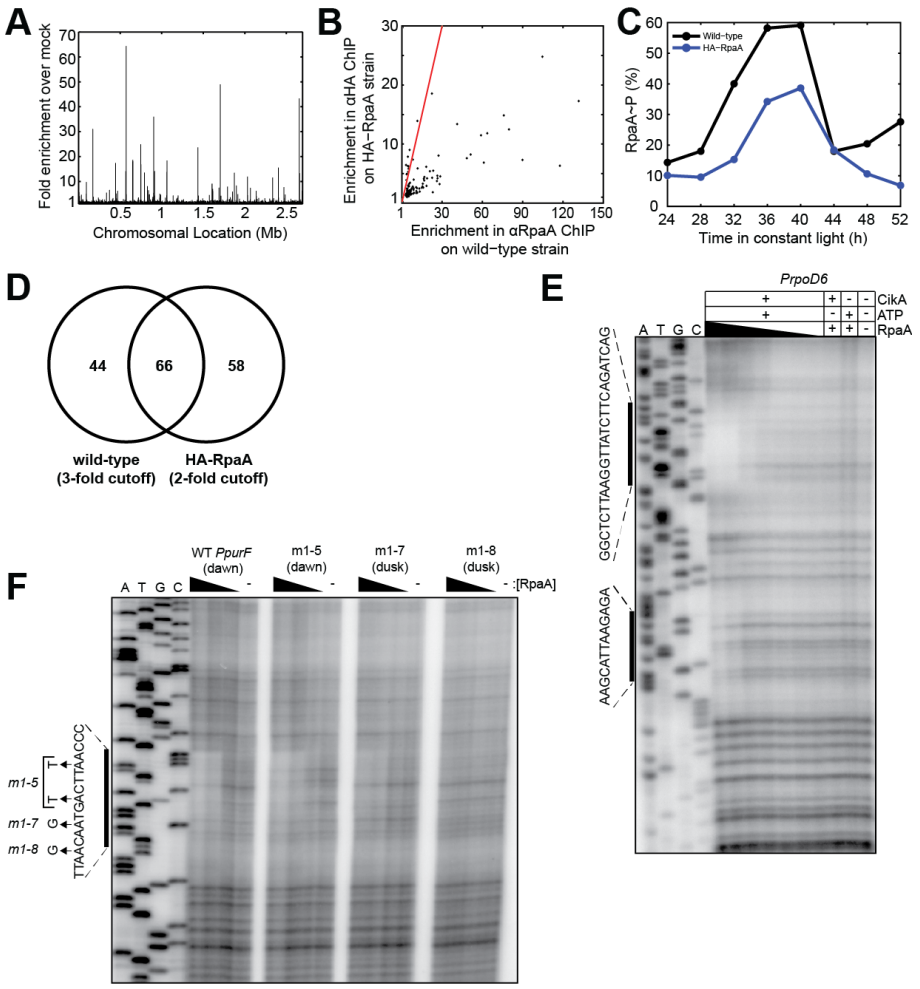

Figure 2.8: Identification of RpaA binding sites by ChIP-Seq (continued). A. Genome-wide binding profile of HA-RpaA measured by ChIP-Seq. The enrichment of read density in the HA-RpaA ChIP-Seq (anti-HA antibody with the HA-RpaA strain) at peak binding, relative to the mock ChIP-Seq (anti-HA antibody with the wild-type strain), is plotted as a function of position on the chromosome. B. Comparison of ChIP-Seq enrichments in the wild-type and HA-RpaA strains at the timepoints of maximum RpaA binding in each strain. For each RpaA binding site identified in the wild-type strain (Figure 2.7; see Methods), we computed the maximal enrichment of read density in the RpaA ChIP-Seq (anti-RpaA antibody with the wild-type strain or anti-HA antibody with the HA-RpaA strain) relative to its respective mock ChIP-Seq. One data point, corresponding to the most strongly enriched binding site in both the anti-RpaA (411-fold) and anti-HA (31-fold) ChIP-Seqs, falls outside the bounds of the plot and is not shown. The red line denotes y = x, i.e., equal enrichments for wild-type and HA-RpaA. C. Comparison of phosphorylation levels of wild-type and HA-RpaA over 1.5 days. Wild-type and HA-RpaA strains were grown side-by-side, entrained with two 12-h dark pulses, and released to continuous light (∼100 μE m-2 s-1) as described in Methods. RpaA phosphorylation was measured by Phos-tag Western blotting [21]. D. Venn diagram showing the overlap between RpaA binding sites identified in the wild-type strain using the anti-RpaA antibody (3-fold enrichment cutoff) and those identified in the HA-RpaA strain using the anti-HA antibody (2-fold enrichment cutoff). D. In vitro DNase I footprinting of RpaA on the *rpoD6* promoter as a function of recombinant RpaA phosphorylation and concentration. Two regions protected from digestion by high levels of RpaA∼P are indicated by the vertical bars on the left. Sanger sequencing reactions used to identify the location of the footprint are shown on the left; footprinting reactions are shown on the right. RpaA pre-treatment and concentration are indicated above each footprinting lane. RpaA was added to a final concentration of 6.0, 3.0, 0.6, 0.3, 0.06, or 0.03 μM as indicated by the thickness of the wedge. F. Footprinting of RpaA∼P on the *purF* promoter. Footprinting was performed on the wild-type (WT) promoter and on mutants displaying altered phases of gene expression [89]. Labeled as in Figure 2.6E with the phase of each promoter (subjective dawn or subjective dusk) indicated.

We identified an A/T-rich motif overrepresented within the RpaA binding sites (Figure 2.7D). This motif is present in the footprints of RpaA∼P on the *kaiBC* and *rpoD6* promoters (Figures 2.5B, 2.7D, and 2.8E), and over half (55%) of the RpaA binding sites contain one or more instances of this motif with p-values less than 0.001. We conclude that the peaks observed in the ChIP-Seq data primarily result from direct binding of RpaA to DNA.

2.3.5 The RpaA regulon contains genes mediating a variety of cellular processes

We systematically identified targets of RpaA genome-wide by searching for genic (annotated mRNA, tRNA, or rRNA) transcripts and high-confidence non-coding transcripts [88] with 5' ends near the RpaA binding sites (see Methods). We found 134 such target transcripts; together, these comprise the RpaA regulon. Ninety-three of these transcripts collectively encode 170 genes (many being co-expressed in operons) (Figure 2.9A), while 41 of the target transcripts are high-confidence non-coding RNAs. Three peaks were located too far from any transcripts to be associated with a target.

Expression of most of the genic RpaA ChIP targets oscillates with circadian periodicity, and these targets are strongly enriched for subjective dusk expression (Figure 2.9B). Expression of a small number of targets peaks at subjective dawn, including, quite interestingly, the canonical subjective dawn gene *purF* [63]. In the *rpaA* mutant, subjective dusk RpaA ChIP targets decrease in expression while subjective dawn targets increase in expression (Figure 2.9C), suggesting that RpaA functions as an activator of subjective dusk targets and a repressor of subjective dawn targets. This helps to explain the dawn phase arrest of the *rpaA* mutant (Figures 2.1B and 2.1C). We also found that most strongly downregulated genes in the *rpaA* mutant are RpaA ChIP targets, while more weakly downregulated genes typically

are not targets (Figure 2.9D). This implies that RpaA acts as a master regulator by directly modulating the expression of a subset of high-amplitude circadian genes, whose products in turn effect the fine global pattern of circadian gene expression.

The RpaA regulon (Figure 2.9A) contains targets involved in transcription regulation, including four sigma factors. Interestingly, RpaA also targets the *himA* gene (*synpcc7942_2248*) encoding the nucleoid protein HU, which is downregulated 1.7-fold in the *rpaA* mutant. Because both DNA compaction [80] and *himA* expression are regulated as a function of circadian time ([90]), *himA* could link oscillatory RpaA activity to circadian genome compaction, providing another route for RpaA to influence gene expression globally.

Several RpaA targets are enzymes of the glycolysis, glycogen, and pentose phosphate metabolic pathways, suggesting a direct link between the circadian clock and energy production and storage. RpaA also directly targets the translation initiation factor IF-3 (*infC*), the protein chaperone trigger factor (*tig*), and a ClpXP protease system (*clpX* and *clpP2*), implying that the clock may modulate translation and protein homeostasis directly, connections that have not been reported previously in this organism. In addition, the *rpaA* promoter itself is an RpaA ChIP target, suggesting the presence of autoregulatory feedback. We note that some 37% of the RpaA ChIP targets have no known function; among these targets there may be unforeseen control nodes with roles in global gene expression regulation.

Cells mutated for *rpaA* display increased efficiency of energy transfer from the light-harvesting phycobilisomes to photosystem II relative to photosystem I, and for this reason the gene was named regulator of phycobilisome association A [5]. We find that RpaA directly targets the *rpaC* gene, which encodes an integral membrane protein implicated in controlling the stability of the photosystem II-phycobilisome interaction [30]. *rpaC* is circadianly expressed and is downregulated approximately two-fold in the *rpaA* mutant, consistent with

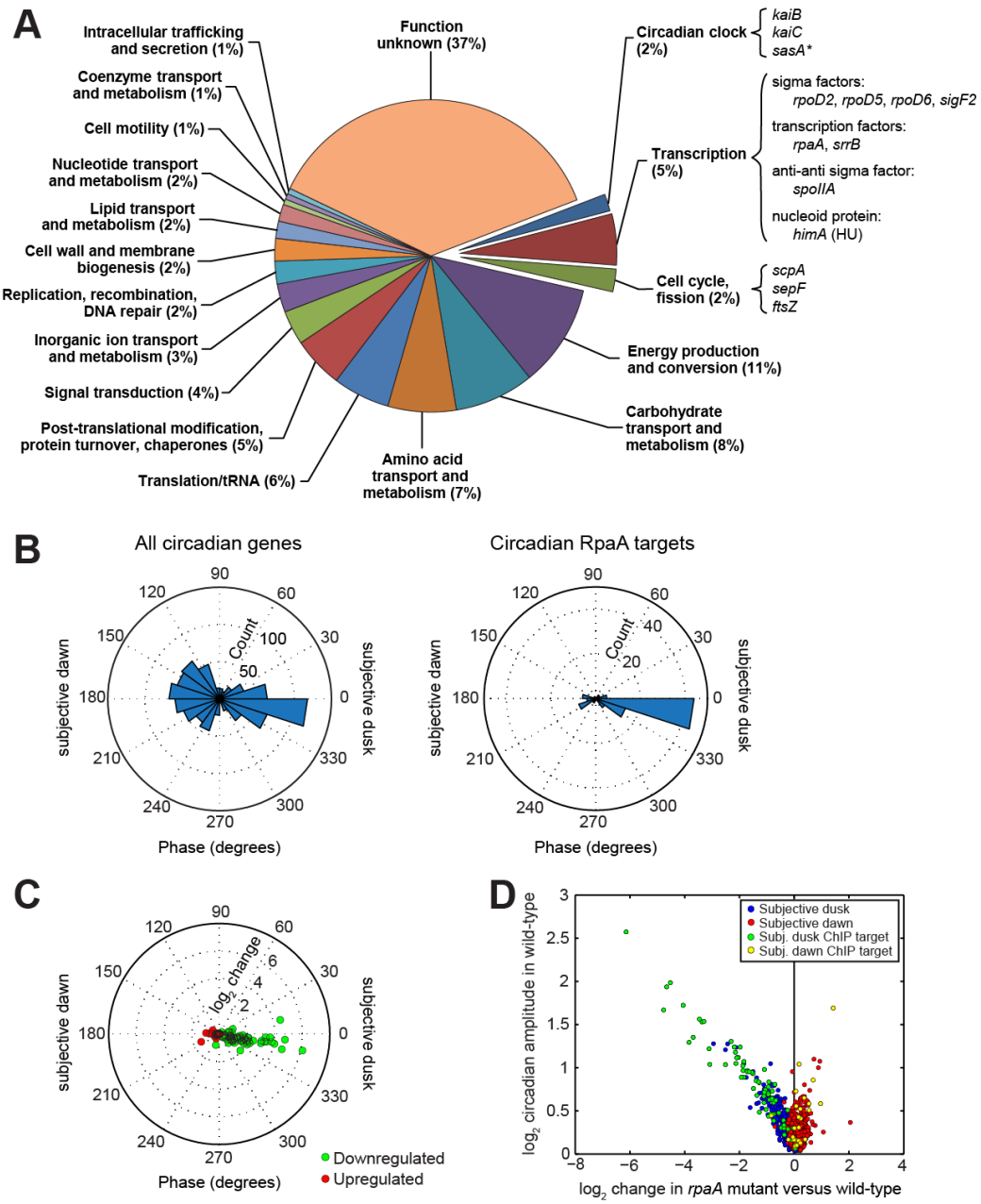

Figure 2.9: The RpaA regulon. A. Functional characterization of protein and tRNA ChIP targets of RpaA. We identified 134 transcripts closest to the 110 binding sites (see Methods). Of those transcripts, 93 encode proteins or tRNAs (corresponding to 170 genes, some of which are co-expressed in operons), while the other 41 are classified as non-coding RNAs [88]. Because the function of the non-coding RNAs is not known, we restrict our functional analysis to the 170 protein-coding or tRNA genes ('RpaA ChIP target genes'). RpaA ChIP target genes were categorized as described in the Methods. Some genes of particular interest are highlighted. The asterisk (*) indicates that the gene's classification as an RpaA ChIP target is artifactual because of assignment to an incorrectly-demarcated operon containing a bona fide target [88]. B. Comparison of the distribution of phases of all circadian genes (left, n = 856, from Figure 2.1A) and of the ChIP target genes whose expression oscillates with circadian periodicity (right, n = 95). C. Change in expression of circadian ChIP target genes downregulated (green, n = 72) or upregulated (red, n = 23) in the *rpaA* mutant plotted as a function of their phase in the wild-type strain. D. Comparison of gene expression change in the *rpaA* mutant with circadian amplitude in the wild-type strain (from Figure 2.1B). Only genes that oscillate with circadian periodicity in the wild-type strain are shown (n = 856). Circadianly-expressed RpaA ChIP target genes (n = 95) are highlighted.

positive regulation by RpaA.

Finally, we found that the RpaA regulon is enriched for genes whose expression increases in darkness in a *kaiABC*-dependent manner [24] (p = 4.4 x 10-6, Fisher's exact test), suggesting that RpaA is involved in clock modulation of the gene expression response to darkness.

2.3.6 Active RpaA is sufficient to switch cells between the two major gene expression states produced by the circadian clock

To test directly whether RpaA serves as the master regulator governing circadian gene expression, we asked whether the RpaA∼P phosphomimetic RpaA(D53E) is sufficient to induce global changes in expression similar to those that occur over the course of a circadian cycle. We hypothesized that overexpression of RpaA(D53E) in an *rpaA* mutant would switch cells from the subjective dawn state in which the mutant resides to the subjective dusk state that coincides with the time of maximal RpaA activity in the wild-type strain. To isolate transcriptional changes resulting directly from activity of RpaA(D53E) from potentially confounding Kai oscillator-dependent processes, we introduced the Ptrc::*rpaA*(D53E) construct into a strain lacking *kaiBC* as well as *rpaA*, producing a *rpaA*- *kaiBC*- Ptrc::*rpaA*(D53E) strain that we refer to as 'OX-D53E.'

We used high-throughput RNA sequencing (RNA-Seq) to compare gene expression changes caused by induction of RpaA(D53E) expression with IPTG (Figure 2.11A) to those experienced during the course of a day in constant light in the wild-type strain. First we calculated the correlation between expression of circadian genes during a timecourse of OX-D53E induction and during a wild-type circadian timecourse (Figure 2.10A). Consistent with the gene expression state of the rpaA mutant (Figure 2.1C), pre-induction OX-D53E is most correlated with wild-type at 24 h (subjective dawn) and most anti-correlated with wild-type

at 36 h (subjective dusk). Over 12 h of RpaA(D53E) induction with IPTG, the correlations reverse (Figure 2.10A): OX-D53E becomes most similar to the wild-type 36 h timepoint and most anti-correlated with the wild-type 24 h timepoint. Importantly, these shifts are not observed when IPTG is added to a control strain (OX-mock) in which no gene is inserted downstream of the Ptrc promoter or when IPTG is not added to the OX-D53E strain (Figures 2.11B and 2.11C).

The flip of the dawn-to-dusk gene expression switch is illustrated by plotting each circadian gene's expression change upon IPTG induction in OX-D53E against its change between subjective dawn and dusk in the wild-type (Figure 2.10B). After induction, subjective dusk gene expression increases in proportion to its change from subjective dawn to dusk in the wild-type, while subjective dawn gene expression decreases in proportion to its change in the wild-type (correlation = 0.8). Furthermore, expression of 85% of circadian genes (725 of 856) differed from baseline by greater than 1.5-fold in at least one timepoint after IPTG induction. The strong correlation between the dawn-to-dusk expression change in the wild-type and the expression change upon RpaA(D53E) induction in OX-D53E shows that active RpaA suffices to switch cells between dawn and dusk expression states.

Consistent with our analysis of the ChIP-Seq data (Figures 2.9B-2.9D), constitutively active RpaA strongly induced expression of ChIP dusk targets, while it had weaker repressive activity toward a minority of its subjective ChIP dawn targets (Figure 2.10B). Interestingly, RpaA ChIP targets comprise only a subset of significantly affected genes; the remaining genes must be activated or repressed by one of RpaA's direct targets. Collective these results demonstrate that RpaA is the master regulator of circadian gene expression, acting as the most upstream node in a network of circadian regulators that together orchestrate global gene expression rhythms.

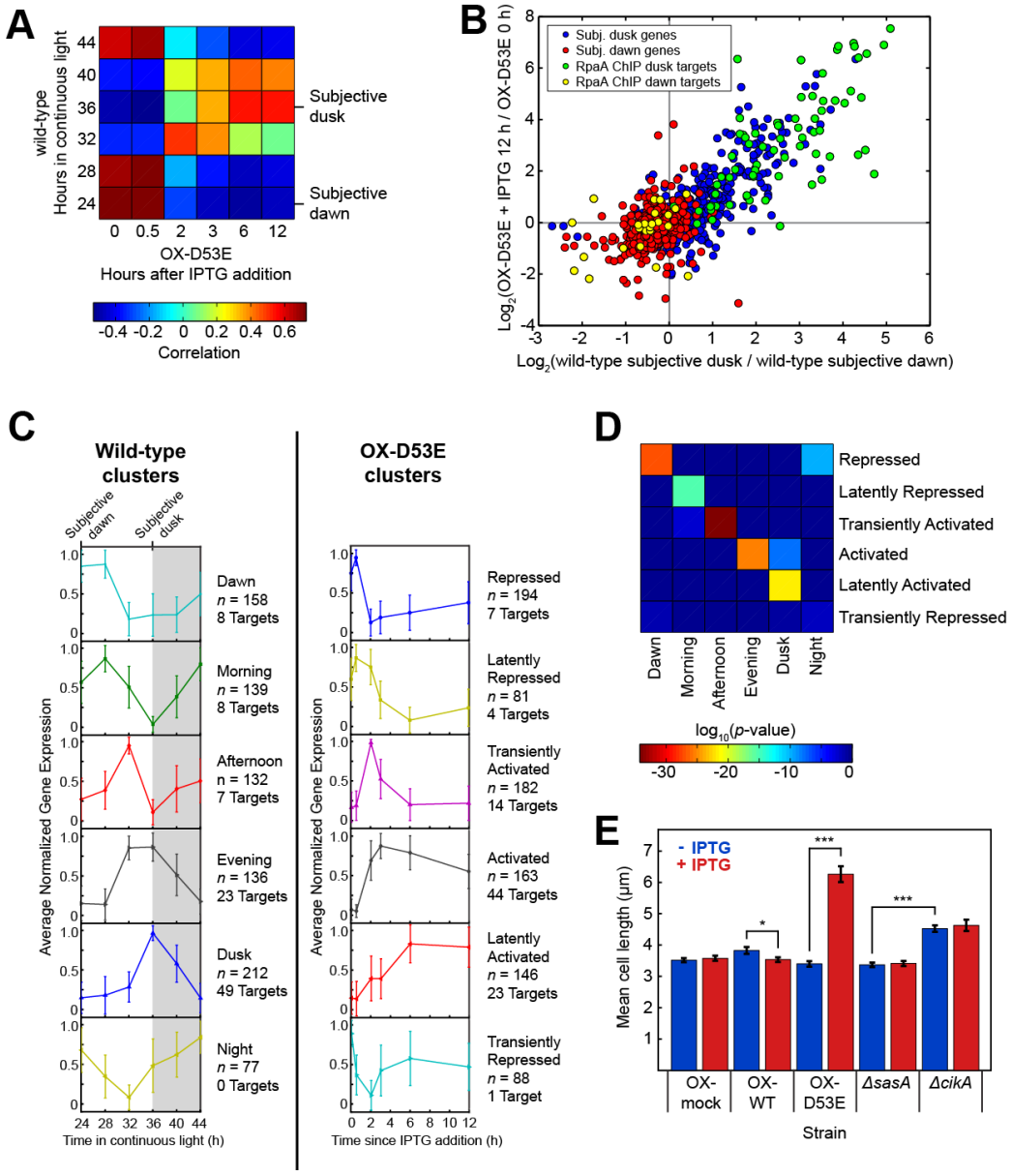

Figure 2.10: RpaA orchestrates global circadian gene expression and controls the cell division gate. A. Correlation between the expression of circadian genes (n = 856) in the wild-type strain over the course of one day and the expression of those genes in the *rpaA*- *kaiBC*- Ptrc::*rpaA*(D53E) (OX-D53E) strain before (t = 0 h) and after induction with IPTG. Gene expression was measured by RNA-Seq. B. Correlation between the change in expression of circadian genes (n = 856) caused by induction of RpaA(D53E) in the OX-D53E strain (y-axis) with the change in expression between subjective dusk and dawn in the wild-type strain (x-axis). RpaA ChIP target genes are highlighted (n = 95; 71 subjective dusk and 24 subjective dawn). Gene expression was measured by RNA-Seq. C. K-means identification of gene expression clusters in the wild-type and OX-D53E strains. Gene expression was measured by RNA-Seq. With K = 6, wild-type circadian genes (n = 856) were separated into six clusters with distinct expression phases (left), consistent with previous microarray observations [90]. Timecourses of the same set of genes in the OX-D53E strain were also clustered using K = 6 (right). The traces show the average normalized timecourse of genes within each cluster; error bars show standard deviation. The numbers of all genes (n) and RpaA ChIP target genes in each cluster are indicated. Non-coding RNAs were not included in this analysis. D. Mapping between clusters in the wild-type (x-axis) and OX-D53E (y-axis) strains. Each element of the heatmap shows the log10 of the statistical significance (Fisher's exact test) of the overlap between the corresponding clusters on each axis. E. Mean cell lengths in *rpaA*- *kaiBC*- strains containing a Ptrc promoter driving expression of wild-type RpaA (OX-WT), RpaA(D53E) (OX-D53E), or an empty multi-cloning site (OX-mock) grown in the presence (red) or absence (blue) of the inducer IPTG (100 μM). At least 80 cells were analyzed for each strain. Error bars show standard error of the mean (SEM). *, p ¡ 0.05; ***, p ¡ 10-13 (one-way ANOVA).

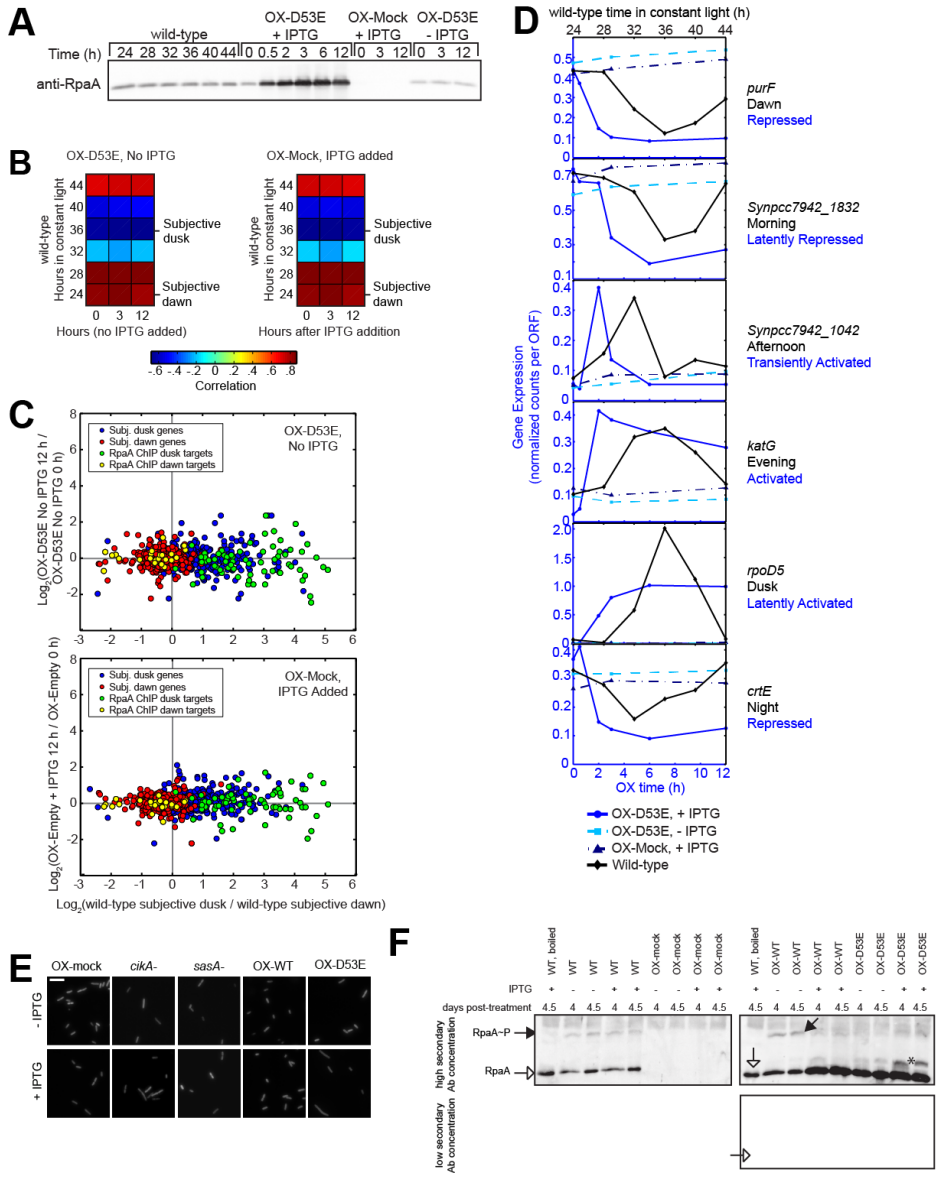

Figure 2.11: RpaA orchestrates global circadian gene expression and controls the cell division gate (continued). A. RpaA protein abundance in samples used for RNA sequencing for Figure 2.10 and related experiments. We prepared whole protein lysates from samples acquired contemporaneously with those used for RNA-seq (Figure 2.10A-2.10E and 2.11B-D) and probed for RpaA by Western blotting using the anti-RpaA antibody used for ChIP-seq. As a control, we added IPTG to a *rpaA- kaiBC-* strain in which the Ptrc construct was inserted into the genome, but without *rpaA* inserted downstream of Ptrc (OX-Mock). We also carried out the same sampling procedure on OX-D53E with no IPTG added (OX-D53E - IPTG). B. Correlation between the expression of circadian genes (n = 856) in the wild-type strain over the course of one day and the expression of those genes in OX-D53E over time in the absence of IPTG addition (left), or those genes in OX-Mock before (time 0 h) and after IPTG addition (right). Gene expression was measured by RNA sequencing. C. Correlation between the change in expression of circadian genes (n = 856) caused by growing OX-D53E in the absence of IPTG for 12 hours (top, y-axis) or incubating OX-Mock with IPTG for 12 hours (bottom, y-axis) with the change in expression between subjective dusk and dawn in the wild-type strain (x-axis). D. Gene expression timecourses of representative genes assigned to corresponding clusters in the wild-type and OX-D53E experiments (Figure 2.10D). Normalized expression is shown on the y-axis as described in the Methods. Timecourses are shown in shades of blue for OX-D53E with IPTG (+ IPTG), OX-D53E without IPTG (- IPTG), or OX-Mock with IPTG (+ IPTG); blue traces correspond to the blue (bottom) x-axis. Timecourses of the same genes in entrained wild-type cells are shown in black; these traces correspond to the black (top) x-axis. Beneath each gene name is shown the name of the wild-type cluster (black) and the OX-D53E cluster (blue) to which the gene was assigned by K-means clustering. E. Representative micrographs of cells from the cell length experiment analyzed in Figure 2.10E. Cells were imaged using their intrinsic red autofluorescence. Scale bar = 10 m. F. Western blot analysis of the phosphorylation state of RpaA in the strains employed in Figure 2.10E. Indicated strains were grown for four days in constant light (74 μE m-2 s-1) with the 100 μM IPTG. Closed, solid black arrows, phosphorylated RpaA (RpaA~P); open arrows, unphosphorylated RpaA. Asterisk (*), heat-stable band of unknown identity, observed in all RpaA-expressing strains. Note that several of the unphosphorylated RpaA bands in the overexpression strains are saturated in the top right image.

2.3.7 The RpaA regulon produces complex gene expression dynamics that parallel those observed during a circadian period

We examined whether the dynamics of the subjective dawn to dusk transition induced by active RpaA mirror those observed in the wild-type strain. We used K-means clustering to identify patterns in expression of circadian genes in the wild-type strain and of those same genes in the OX-D53E strain (Figures 2.10C and 2.11D). In the wild-type strain, clustering separated genes according to their time of maximum expression: Dawn, Morning, Afternoon, Evening, Dusk, and Night [90]. We named the clusters obtained for the OX-D53E strain according to their dynamics following induction: Repressed, Latently Repressed, Transiently Activated, Activated, Latently Activated, and Transiently Repressed. Some genes in the Repressed and Latently Repressed category were slightly induced at 30 minutes after IPTG addition but were subsequently repressed as RpaA(D53E) levels further increased (Figure 2.11A). This suggests that low levels of RpaA promote the expression of these genes, while higher levels have a much stronger repressive effect on their expression. Notably, the timescale of response to induction varies amongst the clusters, with some clusters responding in concert with RpaA(D53E) accumulation (Activated and Repressed) and others more slowly (Latently Activated and Latently Repressed). The Transiently Activated and Transiently Repressed clusters display markedly non-monotonic responses to RpaA(D53E) induction. Hence, the network of regulators downstream of RpaA encodes a variety of responses to continuous accumulation of active RpaA. Note that RpaA ChIP targets never comprise more than one-quarter of the genes in a cluster. Therefore, the majority of genes in each cluster are controlled by the RpaA regulon rather than by RpaA itself.

If the gene expression dynamics observed in the OX-D53E induction timecourse resemble

those in the wild-type circadian timecourse, there should be an overlap between the clusters in each strain. Indeed, we find a nearly one-to-one mapping between the two sets of clusters (Figure 2.10D). The Dawn cluster in the wild-type strain maps to the Repressed cluster in OX-D53E, Morning maps to Latently Repressed, Afternoon maps to Transiently Activated, Evening maps to Activated, Dusk maps to Latently Activated, and Night maps to Repressed ($p \leq 0.001$ for each pair). Remarkably, this mapping arises despite the difference in the timing of RpaA activation in the two strains. In the wild-type strain, RpaA$\sim$P abundance varies sinusoidally, increasing over a $\sim$12 hour timespan and then decreasing over the same timespan (Figure 2.5A). In contrast, RpaA(D53E) accumulates rapidly, plateauing after 3 h of induction and remaining high for the remainder of the timecourse (Figure 2.11A). Nonetheless, genes respond to active RpaA in a stereotyped manner in both strains, suggesting that the dynamics of global circadian gene expression are hard-wired into the RpaA regulon.

2.3.8 Active RpaA closes the cell division gate

During the subjective night cells elongate but do not divide, a phenomenon referred to as cell division gating [12, 53, 95]. When the gate is closed, cells form elongating rods, allowing the status of the gate in a given strain to be inferred from the cell length distribution [12]. The observation that deletions of *sasA* and *cikA* have opposite effects on cell cycle gating [12] suggests that RpaA$\sim$P could be responsible for closing the gate, as SasA and CikA have opposite effects on RpaA phosphorylation [21].

To test this hypothesis, we examined the effect of overexpression of wild-type and phosphomimetic RpaA on cell length. We characterized cell length in strains ectopically expressing wild-type RpaA (OX-WT), RpaA(D53E) (OX-D53E), or an empty multi-cloning

site (OX-mock) from the Ptrc promoter with and without IPTG treatment (Figures 2.10E, 2.11E, and 2.11F). Consistent with our hypothesis, cells showed pronounced elongation in the OX-D53E strain after induction. Conversely, induction of OX-WT shortens the median cell length, consistent with the reduction in RpaA∼P levels (Figure 2.11F) and the repression of *kaiBC* gene expression in this condition (Figure 2.5C). Induction of OX-mock had no effect on cell length. These data are consistent with a causative role for RpaA∼P in clock-mediated cell division gating.

Several RpaA ChIP targets are involved in cell division and so might mediate gating. The most prominent of these are the bacterial tubulin homolog *ftsZ* and the FtsZ regulator *sepF* [45]. Intriguingly, FtsZ is mislocalized in *cikA*- strains [12]. However, mean *ftsZ* and *sepF* expression is unchanged by deletion of *rpaA*, so the functional relevance of these targets is unclear. Other ChIP targets of interest are genes involved in the peptidoglycan biosynthetic pathway (*synpcc7042_0482*, *murB*, *murC*), the last of which has been shown to interact with several Fts cell division proteins [55].

2.4 Discussion

2.4.1 RpaA is the hub through which the circadian clock controls cellular physiology

RpaA phosphorylation links the core Kai oscillator to two of the most striking physiological outputs of the clock: global transcriptome oscillations and gating of cell division (Figure 2.12). Time information encoded in the PTO is read out through the histidine kinases SasA and CikA, which antagonistically regulate RpaA phosphorylation to generate oscillations of RpaA∼P that peak at or immediately preceding subjective dusk [21] (Figures

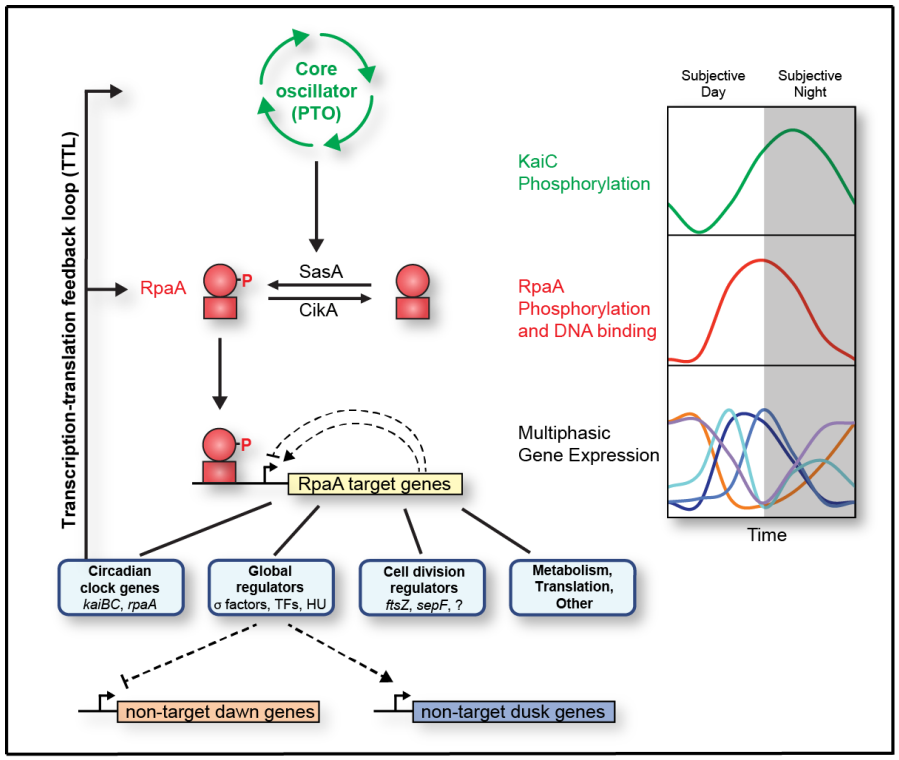

Figure 2.12: Model for RpaA and the cyanobacterial circadian program. Time encoded in the PTO is transduced into RpaA phosphorylation via SasA and CikA, producing oscillations in RpaA∼P (red) that phase-lead those of phosphorylated KaiC (green) by approximately 4 hours [21]. RpaA∼P binds to DNA and controls the expression of the RpaA regulon, which consists of global regulators, cell division regulators, certain clock genes (*kaiBC* and *rpaA*), and genes involved in metabolism and translation. The global regulators are at the top of a transcriptional cascade that orchestrates multiphasic circadian gene expression, repressing subjective dawn genes while activating subjective dusk genes. Fine patterns within the dusk and dawn categories could be generated by a network of interactions amongst RpaA ChIP targets (hypothetical positive and negative feedbacks are shown as dotted lines). RpaA control of *kaiBC* and *rpaA* expression forms the clock TTL.

2.5A and 2.12). The accumulation of RpaA∼P switches the cell's gene expression program from subjective dawn to subjective dusk (Figures 2.10A and 2.10B) via transcriptional programs hard-wired into the RpaA regulon (Figures 2.10C and 2.10D). RpaA∼P accumulation also initiates closure of the cell division gate (Figure 2.10E). RpaA∼P decreases during the subjective night [21] (Figure 2.5A), allowing gene expression to revert to its default dawn-like state (Figures2.1B, 2.1C) and also opening the cell division gate.

Previous reports have implicated circadianly-regulated DNA supercoiling in driving gene expression oscillations [90, 94]. We sought to investigate the relationship between RpaA binding and supercoiling by measuring supercoiling in the *rpaA* mutant, *rpaA*- clock rescue, and OX-D53E strains, but had difficulty obtaining reproducible results and so do not report any here. Nonetheless, we can gain insight into to the relative importance of supercoiling and RpaA binding in generating global gene expression oscillations by comparing the magnitudes of expression changes in response to perturbations in supercoiling and RpaA activity. When supercoiling is rapidly relaxed by treatment with a pharmacological inhibitor of DNA gyrase [90], global gene expression changes in the same manner as it does upon induction of RpaA(D53E) (Figure 2.10B). However, the magnitude of change is substantially larger for induction of RpaA(D53E) than for relaxation of supercoiling (compare Figure 2.10B here to Figure 4 in [90]). Moreover, induction of RpaA(D53E) almost quantitatively reproduces the magnitude of gene expression change observed over circadian time in the wild-type strain (Figure 2.10B). On this basis, we suggest that circadian oscillations in RpaA activity play a dominant role in driving global circadian gene expression oscillations. Perhaps one or more members of the RpaA regulon induces oscillation in supercoiling, which acts in concert with RpaA via a feedforward loop to actuate circadian gene expression. Alternatively, RpaA binding affinity could be positively influenced by supercoiling, in which case relaxation of

supercoiling by pharmacological inhibition of gyrase would release RpaA from the chromosome, leading to observed effects on global gene expression state [90]. Future studies will be required to establish the molecular connections between RpaA activity and supercoiling. Our observation that RpaA targets the *himA* gene encoding the nucleoid protein HU could provide a starting point for such studies.

Also meriting future investigation are possible roles for RpaA in the integration of environmental and time information. Environmental cues may influence RpaA phosphorylation directly, as the activity of its phosphatase CikA is regulated not only by the PTO but also by the cellular redox state, which in turn reflects environmental conditions such as light availability [28, 35]. Hence, information about the environment and circadian time may be integrated at the level of RpaA phosphorylation. RpaB also may play a role: RpaB phosphorylation is regulated by a number of environmental cues [54], and RpaB binds to the promoters of kaiBC and rpoD6 [22], both of which are RpaA targets. In fact, RpaB binding to the kaiBC promoter is antagonized by RpaA [22], consistent with the overlap of the RpaA footprint (Figure 2.5B) with the HLR1 motif bound by RpaB. A systematic exploration of the interaction of RpaA and RpaB at promoters and its effect on gene expression will be required. A recent study found that activity of the canonical subjective dawn promoter P*purF* is affected by the presence of the *kaiC* gene in an *rpaA* mutant background, suggesting the presence of an RpaA-independent output pathway [63]. The authors proposed a model for P*purF* control in which PTO-modulated RpaA∼P levels repress the activity of a 'predominant' RpaA-independent output pathway, which in turn links serine-phosphorylated KaiC to activation of the promoter. This model predicts that *purF* transcript abundance will oscillate in our *rpaA*- mutant clock rescue experiment, in which KaiC phosphorylation oscillates with high amplitude (Figures 2.3 and 2.4). Instead, we found *purF* levels to be constant,

locked at the level of maximum expression in the control clock rescue (Figure 2.4C).

While our data do not rule out the existence of an RpaA-independent clock output pathway, they do suggest that any such pathway is weak relative to the RpaA-dependent output pathway we describe here. Moreover, our ability to switch the genome-wide transcriptional program from subjective dawn to dusk by expression of an RpaA phosphomimetic in a strain lacking both *kaiB* and textitkaiC (Figures 2.10 and 2.11) shows that RpaA phosphorylation is sufficient to control the circadian gene expression program. Notably, *purF* expression is switched from a high to a low state upon phosphomimetic induction (Figure 2.11D); this is straightforwardly explained by the fact that *purF* is a direct ChIP target of RpaA. We found that RpaA∼P binds to P*purF* in vitro and that point mutations affecting its expression phase [89] fall within the RpaA∼P footprint; at least two of these mutations impair RpaA∼P binding (Figure 2.8F).

2.4.2 Generation of complex gene expression patterns with a one-dimensional signal

The complex gene expression dynamics observed upon induction of OX-D53E (Figure 2.10C) demonstrates that a smooth, univariate signal like RpaA phosphorylation (Figure 2.5A and [21]) can generate a diverse array of dynamic responses. We suggest that the smooth RpaA∼P signal is converted into a mosaic of dynamic patterns through network motifs composed of the gene expression regulator targets of RpaA (e.g., sigma factors) and their own downstream targets.

Interestingly, the RpaA∼P level does not uniquely specify the time of day, as intermediate RpaA∼P levels are experienced during both the subjective day and subjective night (Figure 2.5A). The single-bit encoding of time in RpaA phosphorylation contrasts sharply with its

encoding in the PTO, in which differential phosphorylation at two residues of KaiC specifies two bits of information [61,71]. The four phosphoforms of KaiC appear in an ordered pattern during each circadian cycle in a manner that uniquely maps the time of day to a particular phosphoform distribution [61,71]. It seems paradoxical that time encoded in the PTO would be read out through a single-bit channel that cannot uniquely encode it, causing a loss of information. However, the network motifs that likely generate the complex circadian gene expression dynamics (Figure 2.10C and 2.10D) could make the functional effect of a given RpaA∼P level history-dependent (hysteretic). The current level of RpaA∼P and its history together would suffice to fully specify time.

2.4.3 Relevance to eukaryotic circadian clock output pathways

In directly controlling a large number of transcriptional regulators, RpaA resembles eukaryotic circadian effectors like White Collar Complex (WCC) in Neurospora crassa [79], CLOCK and CYCLE in Drosophila melanogaster [1], CLOCK and BMAL1 in mice [36,69], and PRR5 in Arabidopsis thaliana [59]. Transcription factors in these eukaryotic clock output pathways have been proposed to initiate hierarchical transcriptional cascades that effect widespread circadian rhythms in gene expression (Edery, 2011), similar to what we propose occurs with RpaA. However, the output pathways in these eukaryotic clocks are inseparable from the core circadian oscillators: the core oscillators are built from the output pathways themselves, which form time-delayed TTLs that drive oscillations. Multiple interlocking positive and negative transcriptional feedback loops exist in these clocks, with transcription factors serving simultaneously as clock outputs and TTL components. In contrast, the cyanobacterial clock is fundamentally a PTO with a subsidiary and dispensable TTL [68,86], with the output pathway serving mainly as a conduit for transmitting time information from

the PTO to the genome in order to control circadian gene expression and cell division gating. More work is required to determine if these are indeed bona fide differences in the topology of these systems, or if the eukaryotic and prokaryotic clocks share more similarities than currently appreciated.

2.4.4 Synthetic biology applications

Our work has implications for synthetic biology. Identification of the binding motif and binding dynamics of RpaA (Figures 2.5A and 2.7) provide the molecular details required to rationally design connections between the PTO and synthetic transcriptional outputs, enabling reconstitution of the PTO with a transcription-based reporter in orthogonal organisms. Industrial applications arise, as cyanobacteria present an attractive platform for the engineered production of chemicals directly from the abundant fuel source of sunlight [15]. Synthetic pathways could be coupled to the circadian clock via RpaA; as the clock enhances the fitness of the wild-type organism in oscillating environmental conditions [62], so too might it enhance the productivity of synthetic pathways.

2.5 Materials and methods

2.5.1 Cyanobacterial strains

Strains were constructed using standard procedures for genomic integration by homologous recombination [9] and are described in the figure 2.13.

Strains AMC395 and AMC408 were gifts from Susan Golden (University of California, San Diego). Plasmids pAM1573, pAM1580, pAM2055, and pAM2991 were gifts from Susan Golden. Plasmids pDRpaA(Kmr) and pNS2KmPtrc-kaiBC were gifts from Takao Kondo

(Nagoya University).

2.5.2 Cell culture

For the experiments described in Figures 2.1, 2.5A, 2.7, 5, 2.2, 2.6A, 2.8A, 2.8B, and 2.8D, cells were grown in a turbidostat as described previously [90]. Cultures were entrained by exposure to 12 hours of darkness, followed by 12 h of light, followed by another 12-h of darkness, after which they were grown in constant, continuous light.

For Figures 2.3, 2.10, 2.4, 2.6B, 2.8C, and 2.11A-2.11D, cultures were grown in tissue culture flasks illuminated with 100 μE (μmoles photons m^{-2} s^{-1}) of cool fluorescent light and bubbled continuously with 1% CO_2 in air, with the OD750 maintained near 0.3 by diluting the cultures every four hours with fresh medium. Medium was supplemented with 10 mM HEPES-KOH pH 8.0 maintain the pH. For the rescue and RpaA overexpression experiments (Figures 2.3 and 2.4, and 2.10 and 2.11A-D, respectively), cultures were grown initially in the absence of IPTG, treated with two 12-h dark pulses separated by 12 h of light, and released to constant light (100 μE) concomitant with addition of IPTG to a final concentration of 6 μM (Figures 2.3 and 2.4) or 100 μM IPTG (Figure 2.10 and 2.11A-D). Culture conditions for cell length measurements are described in the Cell Length Measurements section below.

2.5.3 Identification of reproducible circadian genes

In order to focus our analysis on genes that show reproducible circadian oscillations, we considered as circadian only those genes identified as circadian in both the 60 h timecourse presented in [90], and in a previously unpublished biological replicate timecourse 48 hours in length (V. Vijayan, personal communication). We also required that genes be expressed in the same class (class 1 or class 2) in both timecourses. Specifically, we considered a gene

Name in this manuscript	EOC number	Genotype*	Background	Reference
wild-type	-	P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*purF::luxAB* (NS 2.1, Cm^r)	AMC408	(Min et al., 2004)
"*rpaA-*" or "*rpaA* mutant"	EOC66	*rpaA*-(Km^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*purF::luxAB* (NS 2.1, Cm^r)	AMC408	This work
control rescue	EOC72	*kaiBC*-(Cm^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*trc::kaiBC* (NS 2.1, Km^r)	AMC408	This work
rpaA- rescue	EOC101	*rpaA*-(Cm^r) *kaiBC*-(Gm^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*trc::kaiBC* (NS 2.1, Km^r)	AMC408	This work
HA-tagged RpaA	EOC67	*3xHA-rpaA*(Km^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*purF::luxAB* (NS 2.1, Cm^r)	AMC408	This work
P*kaiBC* reporter	EOC113	P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r)	AMC395 (Min et al., 2004)	This work
rpaA- P*trc* vector (Figure 3)	EOC339	*rpaA*-(Km^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r) P*trc::*- (NS 2.2, Gm^r)	EOC113	This work
rpaA- P*trc::rpaA* (Figure 3)	EOC341	*rpaA*-(Km^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r) P*trc::rpaA* (NS 2.2, Gm^r)	EOC113	This work
rpaA- P*trc::rpaA(D53A)* (Figure 3)	EOC345	*rpaA*-(Km^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r) P*trc::rpaA(D53A)* (NS 2.2, Gm^r)	EOC113	This work
rpaA- P*trc::rpaA(D53E)* (Figure 3)	EOC346	*rpaA*-(Km^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r) P*trc:: rpaA(D53E)* (NS 2.2, Gm^r)	EOC113	This work
OX-mock	EOC370	*rpaA*-(Km^r) *kaiBC*-(Cm^r) P*trc::*- (NS 2.2, Gm^r)	ATCC PCC7942	This work
OX-WT	EOC374	*rpaA*-(Km^r) *kaiBC*-(Cm^r) P*trc:: rpaA* (N S2.2, Gm^r)	ATCC PCC7942	This work
OX-D53E	EOC371	*rpaA*-(Km^r) *kaiBC*-(Cm^r) P*trc:: rpaA(D53E)* (NS 2.2, Gm^r)	ATCC PCC7942	This work
sasA-	EOC116	*sasA*-(Gm^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r)	EOC113	(Gutu and O'Shea, 2013)
cikA-	EOC118	*cikA*-(Gm^r) P*psbAI::luxCDE* (NS 1, Sp/Sm^r) P*kaiBC::luxAB* (NS 2.1, Cm^r)	EOC113	(Gutu and O'Shea, 2013)

Figure 2.13: Cyanobacterial strains used in this study. Km^r, kanamycin resistance; Cm^r, chloramphenicol resistance; Gm^r, gentamycin resistance; Sp/Sm^r, spectinomycin/spectromycin resistance; NS 1, neutral site 1 (Genbank U30252); NS 2.1, neutral site 2.1 (BstEII site in Genbank SPU44761); NS 2.2, neutral site 2.2 (BglII site in Genbank SPU44761).

to be reproducibly circadian if it had: (a) a Cosiner period [37] of 22 - 26 h in the 60-h timecourse described in [90]; (b) a Cosiner period of 20 - 28 h in the 48-h timecourse; and (c) had the same phase-based class (1 or 2) in both timecourses. The period window is wider for the 48 h timecourse because it was shorter and therefore the period was less precisely determined. 856 genes satisfy these criteria.

2.5.4 Anti-RpaA antibody production and affinity purification

Antibody against full-length recombinant RpaA protein was produced in rabbits and affinity purified prior to use. Recombinant RpaA was produced as described [82], and was used to inoculate two rabbits. Cocalico Biologicals performed the live animal work. To purify the antibody, an affinity column was prepared using Affi-Gel matrix (1:1 mixture of Affi-Gel 10 and Affi-Gel 15, Bio-Rad) to which recombinant RpaA was immobilized. Affinity purification was carried out at room temperature. Antisera from the two rabbits was combined in a 1:1 ratio by volume and applied to a column. The flow-through was passed back through the column. The column was then washed with 20 column volumes (CVs) of 10 mM TrisCl pH 7.5 followed by 20 CVs of 10 mM TrisCl pH 7.5 with 500 mM NaCl. Immobilized antibody was eluted with 10 CVs of 100 mM glycine pH 2.5 in 1 ml fractions and immediately neutralized with TrisCl pH 8.0. A base elution was then performed using 100 mM triethanolamine pH 11.5, neutralized with TrisCl pH 8.0. Protein abundance in the eluates was estimated by BCA assay. Protein was found predominantly in the acid eluate, with little to none present in the base eluate. When the base eluate showed no evidence of protein, it was discarded. Fractions containing protein were combined and dialyzed in 10 KDa MWCO Slide-A-Lyzer units (Pierce) against PBS containing 0.02% NaN_3. Antibody was then concentrated using a 10 KDa MWCO Amicon spin filter. Antibody concentration

was determined by absorbance at 280 nm or by BCA assay.

The specificity of the antibody was assayed by comparing reactivity against wild-type and *rpaA* mutant (*rpaA*-) lysate by Western blot (Figure 2.6A). Only a single band was observed for the wild-type lysate, while no reactivity was observed against the *rpaA* mutant lysate.

2.5.5 Western blot analysis

Cells were collected on nitrocellulose or cellulose acetate filters using vacuum filtration. Filters were flash-frozen in liquid nitrogen and stored at -80° C until lysis. To prepare lysates, cells were eluted from the filters using ice-cold lysis buffer (7.5 M urea, 20 mM HEPES pH 8.0, 1 mM DTT, and 1x Roche Complete protease inhibitor tablet, with or without 1 mM EDTA). Resuspensions were transferred to 500 μl screw-cap tubes containing 0.1 mm glass beads. The cells were then lysed by bead-beating at 4° C for a total of 5 min with periodic cooling on ice. The lysate was centrifuged for 5 min at 16,000 x g at 4° C, after which the supernatant was transferred to a clean microcentrifuge tube. The protein concentration of each sample was then measured by BCA assay (Pierce) using bovine serum albumin (BSA, Bio-Rad) diluted into lysis buffer as the standard. For each Western blot, an equal mass quantity of each lysate was loaded onto an SDS-PAGE gel. SDS-PAGE gel compositions, electrophoresis conditions, and Western blotting procedures were performed as described in [21], except that affinity-purified anti-RpaA antibody (0.17 μg/ml) was used in place of anti-RpaA serum.

2.5.6 Chromatin immunoprecipitation (ChIP)

For each ChIP reaction, approximately 18 OD750 nm units of culture were crosslinked for 15 min with 1% formaldehyde followed by quenching for 5 min with 125 mM glycine. Cells were collected by centrifugation for 10 min at 6000 x g at 4° C and then washed twice with 30 ml of ice-cold phosphate-buffered saline (PBS), centrifuging for 10 min at 3000 x g at 4° C after each wash. Samples were then resuspended in 1 ml ice-cold PBS and pelleted in a microcentrifuge tube for 3 min at 3,200 x g at 4° C. The supernatant was discarded and the pellet was flash frozen in liquid nitrogen and stored at -80° C.

Samples were thawed on ice and resuspended in 500-600 μl of ice-cold lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 1x Roche Complete EDTA-free Protease Inhibitor Cocktail). Cells were lysed by beadbeating at 4° C in 2 ml screw-top tubes with 0.1 mm glass beads for 10 cycles of 30 seconds each separated by at least 30 sec of cooling on ice. Lysate was separated from the beads by piercing the bottom of each tube with a small-diameter needle, placing the tube into a clean 1.5 ml microcentrifuge tube, and centrifuging for several minutes at 500 x g at 4° C to transfer the lysate to the 1.5-ml tube. Chromatin was then sheared to a length of 300 bp by sonication on ice in a Misonix sonicator 3000 for 9-11 cycles of 15 sec separated by at least 90 sec of cooling. Cell debris was removed by centrifuging twice at 14,000 x g for 15 min each at 4° C. Protein concentration in the lysates was determined by BCA assay using BSA as a standard.

For a given ChIP timecourse, equal mass quantities (typically 1-1.5 mg) of lysate from each timepoint were prepared in 500 μl of lysis buffer each. For anti-RpaA ChIP, equal amounts (typically 10-15 μg) of affinity-purified antibody were added to each tube; for anti-HA ChIP, 40 μl (bed volume) of anti-HA agarose beads (Pierce) equilibrated in lysis

buffer were added to each tube. Samples were incubated overnight in the dark at 4° C with continuous rotation. The following morning, for anti-RpaA ChIP only, 50-70 μl of rProtein A Sepharose Fast Flow beads (GE Healthcare; 33% slurry in lysis buffer) was added to each sample, and the samples were then incubated for 2 h at 4° C with rotation. For both anti-RpaA and anti-HA ChIP, beads were isolated by centrifugation for 1 min at 1000 x g at room temperature. Beads were then washed twice with 1 ml lysis buffer, once with 1 ml of buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate), once with 1 ml of wash buffer (10 mM TrisCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.1% sodium deoxycholate) and finally with 1 ml of TE pH 7.5 (10 mM TrisCl pH 7.5, 1 mM EDTA); each wash was conducted for 5 min at room temperature on a tube rotator followed by isolation of beads by centrifugation for 1 min at 1000 x g at room temperature. Protein-DNA complexes were then eluted with 250 μl elution buffer (50 mM TrisCl pH 8.0, 10 mM EDTA, 1% SDS) for 1 h at 65 C.

For both the eluate and a matched sample of lysate not subjected to immunoprecipitation ('input DNA'), crosslinks were reversed for 6-18 h at 65° C. Next, 250 μl of TE was added to each sample to dilute the SDS, followed by addition of 100 μg of proteinase K and 80 μg of glycogen. Samples were incubated for 2 h at 37° C to digest proteins. Samples were then supplemented with 55 μl of 4 M LiCl and extracted with 1 ml of phenol/chloroform/isoamyl alcohol followed by extraction with 1 ml chloroform. DNA in the aqueous phase was precipitated with ethanol and then washed once with 80% ethanol. Pellets were air-dried and then resuspended in 50 μl of TE containing 20 ng/μl of DNase-free RNase (Fermentas) and incubated for 1 h at 37° C to digest RNA. Samples were then supplemented with 150 μl of TE and 22.2 μl of 3 M sodium acetate pH 5.2, extracted with phenol/chloroform/isoamyl alcohol, and precipitated with ethanol; alternatively, DNA was purified using a Qiagen PCR purifi-

cation kit. DNA concentration in ChIP samples was estimated either by PicoGreen assay (Invitrogen) or by qPCR based on the abundance of the coding region of *synpcc7942_0612*, a region showing no enrichment for RpaA binding in ChIP-Seq experiments; for both methods, input DNA quantified by absorption spectrophotometry was used as a standard.

Typical immunoprecipitation efficiencies (fraction of RpaA depleted from the lysate) were greater than 50%.

2.5.7 qPCR for ChIP and gene expression

qPCR was carried out using Taq polymerase, SybrGreen dye, and the promoter-specific primers described below.

RT-qPCR was performed as described previously [89], with the abundance of the *kaiBC* transcript normalized to that of *hslO* transcript, whose abundance is constant in time [89,90].

For ChIP-qPCR, DNA abundances were calculated using a standard curve constructed from input DNA, which was prepared in the same manner as ChIP DNA but using sonicated lysate not subjected to immunoprecipitation.

Assay	Target	Primer sequences	Reference
ChIP-qPCR	P*kaiBC*	F:!TTTTACGAGGGCTCATACGC! R:!CCCACGAGAAACCTGAAAAG	This work
ChIP-qPCR	*synpcc7942_0612* CDS	F:!AGAAGCAACGACAGCGTAGG! R:!GTCGCAGCTCGGATTTTTG!	This work
RT-qPCR	*kaiBC*	F: TACATTCTCAAGCTCTACG R: CGTCGCTAGGATTTTATCC	(Vijayan and O'Shea, 2013)
RT-qPCR	*hslO*	F: CAGACCAACTGATTCGAGCG R: GGAGGCCAGGAGCAGTC	(Vijayan and O'Shea, 2013)

2.5.8 ChIP-Seq library preparation and sequencing

Libraries for Illumina sequencing of ChIP DNA were prepared following a protocol developed by Ethan Ford (*http://ethanomics.files.wordpress.com/2012/09/chip_truseq.pdf*), with modifications. Specifically, 0.15-3 ng of ChIP DNA was used for each sample. DNA ends were blunted by treatment with 1.4 units of T4 DNA polymerase (NEB), 0.45 units of Klenow fragment (NEB), 4.5 units of T4 polynucleotide kinase (NEB), and 0.4 mM dNTPs (NEB) in 1x T4 DNA ligase buffer (NEB) in a total of 50 μl volume for 30 min at 20° C. DNA was purified using 50 μl of AMPure XP beads (Beckman) and 50 μl of a solution containing 20% PEG8000 (Sigma) and 1.25 M NaCl. DNA was eluted in 16.5 μl of TE/10 (10 mM TrisCl pH 8.0, 0.1 mM EDTA). DNA was then A-tailed at the 3' ends by treating the eluate with 2.5 units of Klenow fragment lacking 3'-5' exonuclease activity (Klenow fragment 3'–¿5' exo-, NEB) and 0.2 mM dATP (GE Healthcare) in 1x NEB Buffer 2 in a total volume of 20 μl for 30 min at 37° C. TruSeq adapters (Illumina or Eurofins MWG Operon) were ligated onto the A-tailed DNA by addition of 25 μl of 2X Quick Ligase Buffer (NEB), 0.5 μl of 250 nM TruSeq adapter (1:30 dilution of Illumina stock), 3 μl of nuclease-free H2O, and 1.5 μl of Quick Ligase (NEB) followed by incubation for 20 min at 21° C. The ligation reaction was stopped by addition of 5 μL of 0.5 M EDTA pH 8.0 (Ambion). Next, DNA was purified using 55 μl of AMPure XP beads without additional PEG or salt. DNA was eluted in 15.5 μl of TE/10. The Y-shaped adapters were then linearized with 5 cycles of PCR (initial denaturation of 30 sec at 98° C followed by 5 cycles of [10 sec at 98° C, 30 sec at 60° C, 30 sec at 72° C] followed by 5 min at 72° C) using Phusion polymerase (Thermo) in HF Buffer and 1 μl of TruSeq primers (25 μM) in a total volume of 31 μL. Linearized DNA was purified using 30 μl of AMPure XP beads without additional PEG or salt. DNA was eluted in 30 μl of TE/10. Fragments between 300 and 500 bp were size-selected using agarose gel

purification and the QIAquick gel extraction kit (Qiagen) or using a Pippin Prep (SAGE Science). Purified DNA was further amplified with 13-14 cycles of PCR, as described above, in a total volume of 62.5 μl. Following PCR, DNA was purified using 51 μl of AMPure XP beads without additional PEG or salt. DNA was eluted in 12 μl of TE/10. Libraries were assessed using a DNA High Sensitivity chip on an Agilent Bioanalyzer 2100. Samples were sequenced on a HiSeq instrument or Genome Analyzer II (Illumina) by the core facility at the Harvard FAS Center for Systems Biology. Reads were aligned to the S. elongatus genome using Bowtie [40], counting only those aligning uniquely to one location with up to three mismatches. Samples averaged 2 million aligned reads.

2.5.9 RpaA binding motif

Enriched motifs were identified with MEME [6] using sequences within 125 bp of each RpaA peak and a background consisting of a fourth-order Markov model of the entire genome. We searched for motifs between 6 bp and 26 bp in width, the latter value being the width of the largest DNase I footprint we observed (P*rpoD6* no. 1, Figures 2.7D and 2.8E). Only one statistically significant motif (Figure 2.7D, E-value 1.7 x 10^{-36}) was located. We identified instances of this motif in the query sequences using FIMO [19].

2.5.10 DNase I Footprinting

The P*kaiBC*, P*rpoD6*, and P*purF* promoter sequences were amplified from *S. elongatus* genomic DNA and mutated where indicated with the QuikChange II XL site-directed mutagenesis kit (Agilent).

DNA probes for DNase I footprinting were prepared as follows. First, a forward primer was 5' end-labeled with ^{32}P by incubation with T4 purine nucleotide kinase (Promega) and

γ-32P ATP (Perkin Elmer). For the P*kaiBC* probes, this primer was TTTTACGAGGGCTCATACGC, for the P*rpoD6* probe this primer was ATCTCTTGTTCGTCGCTGAG, and for the P*purF* probes this primer was CATGCCCTTGCAGAGCTC. This primer was subsequently used to prepare the footprinting probe via PCR amplification from the appropriate plasmid template from above with a reverse primer. For the P*kaiBC* probes, the reverse primer was GTGAGATGTATCGACGGTCTATCC, for the P*rpoD6* probes, the reverse primer was TCCCTCTTACATTTTCGACACA, and for the P*purF* probes, the reverse primer was GCGATCGAACGTCGTTTG.

Binding reactions were carried out in a buffer containing 150 mM KCl, 5 mM MgCl2, 20 mM HEPES-KOH pH 8.0, 10% glycerol (w/v), 1 mM DTT, and 30 ng/μL poly dI-dC (Affymetrix). Binding reactions contained the appropriate DNA probe, varying concentrations of RpaA (purified as described in [82]), 1.5 μM CikA (purified as described previously [21], and 1 μM ATP, as indicated. Binding reactions were incubated for 1 hour at 30° C.

To initiate DNase I digestion, an equal volume of DNase I (0.025 U/μL, New England Biolabs), in a buffer containing 25 mM TrisCl pH 8.0, 10 mM MgCl2, 5 mM CaCl2, 1 mM EDTA, and 10% glycerol was added to the reaction and incubated at room temperature for 2 minutes. Digestion was stopped by adding an equal volume of stop solution containing 200 mM NaCl, 30 mM EDTA, 1 % SDS (w/v), and 100 μg/mL yeast tRNA (Sigma).

DNA was then purified using phenol-chloroform extraction and ethanol precipitation and resuspended in a buffer containing 30 mM NaOH, 60% v/v formamide, and 0.1% w/v bromophenol blue. Sanger sequencing reactions were prepared with the Sequenase DNA sequencing kit (USB) using the appropriate plasmid as template. The reactions were run on a 6% acrylamide-urea gel prepared with the Ureagel system (National Diagnostics). Radio-

labeled DNA was visualized with phosphorimaging.

For Phos-tag electrophoresis, reactions were carried out as described above but with kaiBC DNA probe prepared with unlabeled ATP. After one hour of incubation at 37° C, 2.5 μL of reaction mixture was diluted into 1x Phos-tag loading buffer and resolved on a 7% acrylamide gel containing 75 μM MnCl2 and 50 μM Phos-tag AAL-107 reagent (Wako Chemicals) at 4° C. After electrophoresis, the gel was stained with Sypro Ruby (Invitrogen) according to manufacturer's instructions and imaged on a Typhoon Scanner (GE Healthcare).

2.5.11 ChIP-seq data analysis

Data were analyzed using a custom-coded, modified form of the PeakSeq algorithm [70] that narrows the regions identified as peaks by requiring that each 50-bp window within a putative peak be enriched ($p \leq 0.05$) relative to both the mock ChIP and the input DNA. The fold enrichment for each trimmed peak was calculated by finding the maximum ChIP-to-mock ratio within 50 bp of the location of the peak maximum in the raw ChIP-Seq signal (not in the enrichment ratio). For ChIP using the anti-RpaA antibody, we required peaks to be present in each of two biological replicates with enrichment $\geq$ 3-fold in one replicate and $\geq$ 2.22-fold in the other replicate, which had 26% lower enrichment overall. Multiple hypothesis-corrected q-values were less than 10^{-78} (smallest q-value between the two replicates).

To call targets of a given RpaA peak, we identified all transcripts with 5' ends lying within 500 bp of the location of maximum ChIP-Seq signal. (For this analysis, we considered only annotated mRNA, tRNA, or rRNA transcripts and high-confidence non-coding transcripts [88]). Within this set of nearby transcripts, we called as the target that transcript with its

5' end nearest the ChIP-Seq signal maximum. To account for possible inaccuracies in the 5' end identifications, we also assigned as targets any other transcripts with 5' ends residing within 50 bp of the nominally closest transcript.

RpaA ChIP target genes were categorized by their Cyanobase-assigned functional category [60] using manually-updated assignments.

2.5.12 RNA-seq

Ribosomal RNA was depleted from 6 μg of total RNA (purified as described in [90]) using the MICROBExpress Bacterial mRNA enrichment kit (Applied Biosystems) according to manufacturer's instructions. Strand-specific RNA-sequencing libraries were prepared from 375 ng of rRNA-depleted RNA using the TruSeq Stranded mRNA Sample Prep Kit (Illumina). Samples were multiplexed and sequenced on an Illumina HiSeq machine by the core facility at the Harvard FAS Center for Systems Biology. Sequencing reads were aligned to the *S. elongatus* chromosome as described above for ChIP-Seq. Samples averaged 1 million aligned reads.

To quantify gene expression, we counted the number of coding-strand sequencing reads with 5' ends between the start and stop positions of the coding region of each gene. To normalize gene expression values between samples, we utilized median normalization as described elsewhere [4]. First, we calculated a pseudo-reference for each gene by determining the geometric mean of the expression counts for that gene across all samples, and calculated the ratio of the expression values with the appropriate pseudo-reference value. We determined the median value of these ratios within each sample and took this as a size factor that estimates the sequencing depth of each sample. To normalize, we divided all gene expression values within a sample by the appropriate size factor. For Figure 2.11D, gene expression

values were further normalized by dividing the gene expression values by the length of the appropriate open reading frame.

To calculate correlations for Figures 2.10A and 2.11B, we determined the average expression of each gene in wild-type cells over the course of a day. Then, we expressed gene expression at each timepoint as the log of the ratio of the expression at that timepoint to the average wild-type expression. If a circadian gene showed an expression value of 0 (meaning that no reads were present) in any timepoint from this experiment, we did not consider it when calculating correlations (n = 24 genes). We also ignored expression of *kaiB* and *kaiC* in this analysis, as these genes were disrupted in OX-D53E. We calculated Pearson's correlation coefficient between samples for the remaining high-confidence circadian genes with these log-transformed values.

For K-means clustering, we first normalized all gene expression values for reproducibly circadian genes in the OX-D53E timecourse or the wild-type timecourse using z-score normalization, and then clustered these genes into 6 groups using K-means clustering (MATLAB) with squared Euclidean distance as the distance metric.

2.5.13 Bioluminescence timecourses

Cultures were inoculated into 96-well plates containing 250 μl BG-11M and the specified concentration IPTG in each well. After incubation in light (60 μE m^{-2} s^{-1}) for a minimum of 12 h, the cultures were entrained with two 12-h dark pulses and then released to constant light. Bioluminescence was measured every 2 h in constant light in a TopCount luminometer (PerkinElmer) as described previously [89]. Replicate experiments produced qualitatively similar results.

2.5.14 Cell Length Measurements

Cultures were grown in medium containing appropriate antibiotics but no IPTG under constant light (74 μE m^{-2} s^{-1}) at 30° C. After reaching OD750 nm of 0.4, cell cultures were diluted 1:10 into medium containing 100 μM IPTG and then grown for 4 days. Control strains received no IPTG. Cells were imaged by capturing their red autofluorescence using an AxioObserver Z1 inverted microscope (Zeiss) equipped with a Plan-Apochromat 100X/1.40 Oil Ph3 objective (Zeiss) and an Evolve EMCCD Camera (Photometric). Cell length analysis was performed in ImageJ (National Institutes of Health).

Chapter 3

Responses to natural changes in light intensity interact with the circadian clock to control gene expression in cyanobacteria

3.1 Abstract

The circadian clock in the cyanobacteria *Synechococcus elongatus* PCC7942 generates dynamic changes in the expression of a large group of genes under constant light conditions, but it is unclear how the expression of these genes is affected by the dynamic environmental changes in light intensity that occur in nature. We employ genome-wide approaches to demonstrate that natural changes in light intensity cause substantial changes in the expression of clock-regulated genes through changes in RNA polymerase recruitment, at least in part by controlling the activity of the transcriptional regulators RpaA and RpaB. Using

mathematical modeling, we show that the environmentally-responsive expression patterns of large groups of co-expressed circadian genes can be explained by measured values of phosphorylated RpaA and RpaB. Our work shows that changes in the environment are integrated with circadian output to control the expression of large groups of genes in response to time-of-day and environmental changes.

3.2 Introduction

Photosynthetic organisms must maintain homeostasis in an environment in which sunlight intensity constantly changes. Cyanobacteria, such as the model species *Synechococcus elongatus* PCC7942, are prokaryotic photosynthetic organisms that harvest light energy to grow and divide in the face of excesses or shortages of sunlight. These photoautotrophic organisms employ regulatory systems that adjust physiology in response to both predictable and unpredictable changes in light availability, but it is unclear how these systems interact when acting at the same time in the dynamic conditions of nature [93].

The day/night cycle caused by the rotation of the Earth imposes a predictable challenge to cyanobacteria. Sustained darkness is a significant metabolic challenge for photoautotrophs and leads to a cessation of cell division and downregulation of both transcription and translation [8, 27, 78, 83]. To cope with the predictable onset of night, *S. elongatus* uses a circadian clock to generate circadian (∼24 hour) oscillations in the mRNA abundance (expression) of large portion of its genes [27, 90]. The circadian clock promotes survival of cyanobacteria under light/dark regimes by controlling the expression of genes important for utilizing alternative energy sources at night [10, 67].

The circadian clock is a self-sustaining oscillatory system that produces oscillations in expression of genes even when cyanobacteria are continuously illuminated with constant

low amounts of light (*Constant light conditions*). In Constant light, genes with oscillating expression levels (circadian genes) either show maximal expression when cells expect dusk (dusk genes), or maximal expression when cells expect dawn (dawn genes) ([27, 90]; Figure 3.1A, Figure 3.2). The oscillatory expression of these circadian genes under Constant light conditions is generated by the master OmpR-type transcription factor RpaA [48, 82]. The circadian clock generates circadian oscillations in the amount of phosphorylated RpaA (RpaA∼P), with RpaA∼P levels peaking at subjective dusk [21, 82], Figure 3.1A). RpaA binds DNA when phosphorylated to activate the expression of a subset of dusk genes, and indirectly activating and repressing the expression of the rest of the dusk and dawn genes, respectively ([48], Figure 3.1A, Figure 3.2).

Constant light conditions do not exist in nature, and in the real world circadian clocks must function in cells that are exposed to environmental changes. For example, light intensity on the surface of the Earth changes on both short and long time scales. On cloudless days, light intensity varies in a parabolic manner due to the rotation of the Earth (Figure 3.1B, [66]). Shading events, such as when clouds obscure the sun, can cause random changes in the amount of light that reaches the Earth's surface which last for varying lengths of time (Figure 3.1C, [66]). A complete understanding of circadian clock function in cyanobacteria must incorporate knowledge of how such natural changes in light intensity will affect the gene expression output of the clock.

Changes in light intensity like those seen in nature cause variations in photosynthetic output [49], and cyanobacteria must respond to these changes to maintain homeostasis. Among the best understood light-responsive pathways in cyanobacteria is the NblS-RpaB OmpR-type two component system which adjusts the expression of genes involved in photosynthesis in response to changes in light intensity and other stimuli [23, 33, 34, 54, 72, 73, 81]. Rapid in-

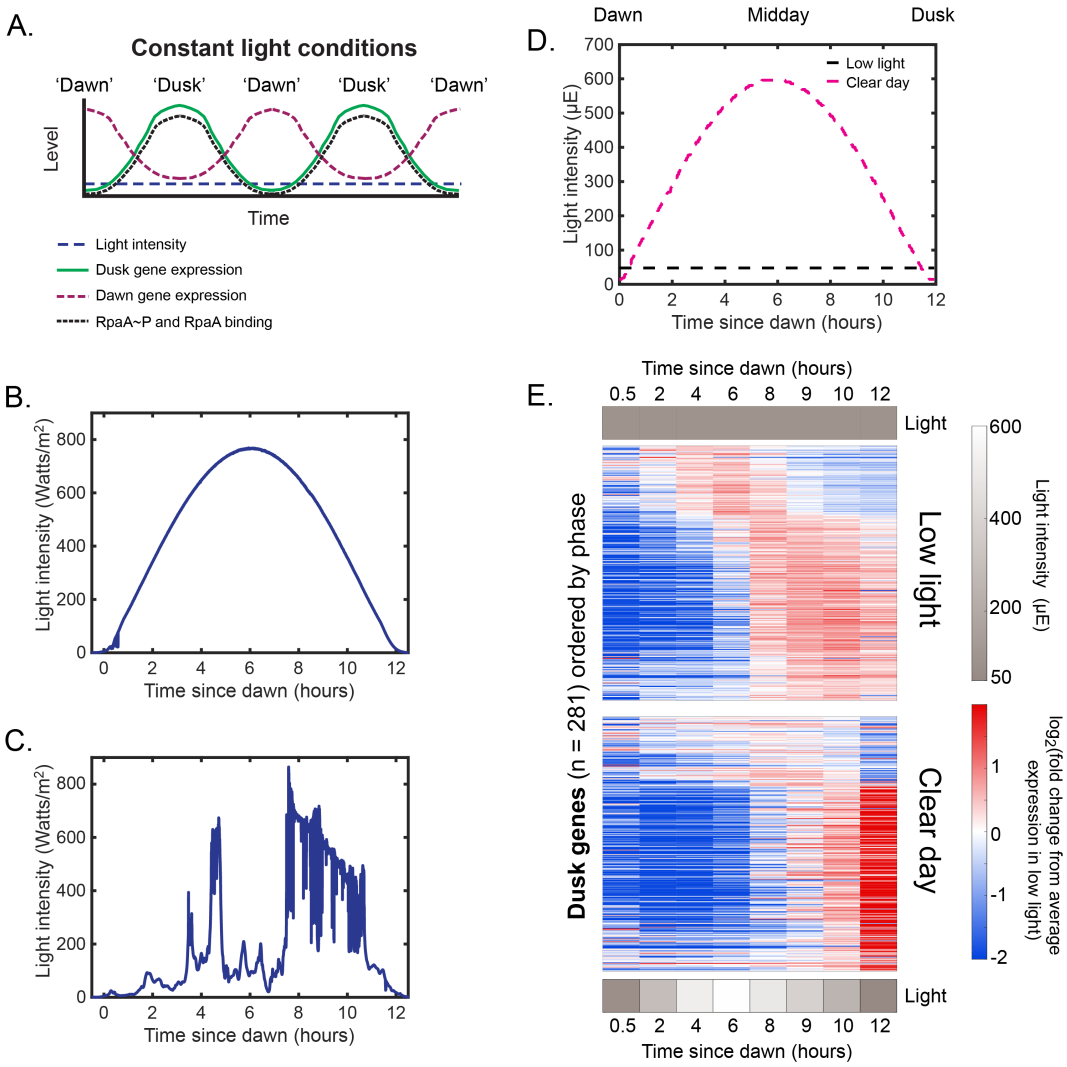

Figure 3.1: Clear day conditions modulate the expression of dusk genes. A. Output of the circadian clock under Constant light conditions. B. Solar irradiance measurements in units of Watts m^{-2} at 342.5 meters above sea level in Madison, WI on March 23rd, 2013 [66]. 'Dawn' was defined as the first time when solar irradiance reached a level of 5 watts m^{-2}. Solar irradiance measurements on April 12th, 2014 in Madison, WI, plotted as in B [66]. D. Light intensity profiles of Low light (black) and Clear day (magenta) conditions, in units of E (μmol photons m^{-2} s^{-1}). E. Gene expression dynamics of all dusk genes (n = 281) under Low light (top) and Clear day (bottom) conditions. Gene expressed was measured at each time point using RNA sequencing. Gene expression is expressed as the log2 fold change from the average expression of the gene over all time points in the Low light condition. Genes are plotted in the same order in both heat maps and were sorted by phase under constant light conditions [90]. Light intensity at each time point is indicated in a heat map adjacent to the corresponding condition.

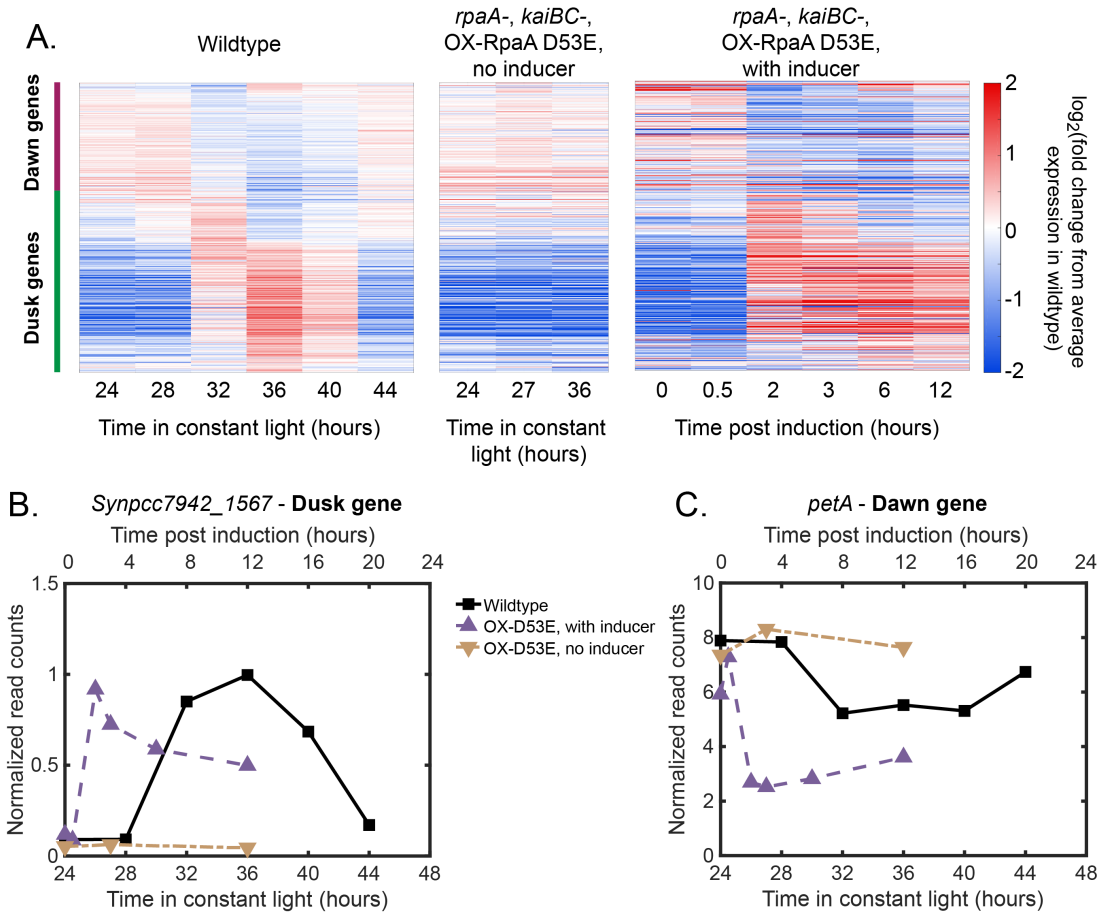

Figure 3.2: Gene expression dynamics of dusk and dawn genes under Constant light conditions (data from [48]. A. Gene expression of circadian genes over one day in Constant light conditions in wildtype cells (left heat map) and OX-D53E cells (*rpaA*-, *kaiBC*-, Ptrc::*rpaA*(D53E)) (center and right heat maps). Gene expression was measured in various strains using RNA sequencing. Gene expression is expressed as the log2 fold change from the average expression of the gene over all time points in the wildtype cells under the Constant light condition. Genes are ordered based on their phase in constant light [90], with the group of dusk and dawn genes indicated with colored bars next to the left heat map. Dusk genes increase in expression under constant light in wildtype cells and are maximally expressed when cells expect it to be dusk. Dusk genes are constantly lowly expressed in cells lacking *rpaA* (OX-D53E without inducer, center) and increase in expression when an RpaA phosphomimetic is induced in an *rpaA*- background (OX-D53E with inducer, right). B. Data from (A) plotted for the representative dusk gene *Synpcc7942_1567*. C. Data from (A) plotted for the representative dawn gene —textitpetA.

creases in light intensity cause decreases in the amount of phosphorylated RpaB (RpaB∼P), which leads to a decrease in RpaB binding to DNA in vivo and changes in gene expression [23, 54]. It has been shown that RpaB binds to the promoters of some RpaA-regulated circadian genes and can affect the expression of circadian genes when overexpressed, suggesting that RpaB could play a role in regulating the expression of circadian genes in response to environmental changes [17, 22].

We sought to understand how naturally-relevant changes in light intensity affect the transcriptional output of the circadian clock in cyanobacteria. We find that growth in light conditions that mimic natural conditions causes significant changes in the expression pattern of dusk genes. By exposing cells to systematic changes in light intensity, we find that expression of dusk genes changes rapidly through changes in RNA polymerase recruitment in a manner anti-correlated with the change in environmental light intensity. We explore how RpaA and RpaB activity is affected by light intensity using genome-wide techniques, and find that the recruitment of both proteins to the promoters of dusk genes is affected by changes in light intensity. Finally, we use mathematical models to demonstrate that changes in expression of dusk genes in response to dynamic light regimes can be explained by changes in levels of RpaA∼P and RpaB∼P, and suggest mechanisms of regulation downstream of these transcription factors which can explain the observed responses. Our work illustrates that the expression of circadian genes in cyanobacteria is intimately tied to environmental changes and suggests mechanisms behind the integration of circadian and environmental information to control physiology.

3.3 Results

3.3.1 Changes in light intensity modulate the transcription of dusk genes

We first explored how parabolic variation in light intensity that mimics the change in light intensity on a cloudless day would affect the expression of circadian genes. To this end, we developed a culture setup in which cells were illuminated via programmable white LED arrays. As a control, we grew cells in a *Low light* condition in which they were illuminated with the constant low light intensity of 50 μmol photons m^{-2} s^{-1} (μE) for 12 hours, which is comparable to the intensity used in most studies of free running behavior of the cyanobacterial circadian clock (Figure 3.1D, black dashed line). As a comparison, we grew cells under a *Clear day* condition in which light intensity parabolically varied over a 12 hour period, gradually increasing from light onset (dawn) to a maximum light intensity of 600 μE at 6 hours after light onset (midday), and gradually decreasing until 12 hours after light onset (dusk) (Figures 3.1D, dashed magenta line; 3.3, 3.4).

We used RNA sequencing to measure the mRNA levels of circadian genes in Low light and Clear day conditions. Almost all dusk genes show a delay in their activation under Clear day conditions, with most genes showing highest expression at dusk under Clear day conditions (Figure 3.1E). Compared to Low light conditions, dusk gene expression tends to be lower at midday in Clear day conditions when light intensity is high (Figure 3.1E, 6 hours). When light intensity decreases at the end of the Clear day condition ('sunset'), most dusk genes sharply increase in expression to levels much higher than under Low light conditions (Figure 3.1E, 12 hours). Dawn genes show the opposite behavior —they have higher expression under Clear day conditions, although this trend is less pronounced (Figure

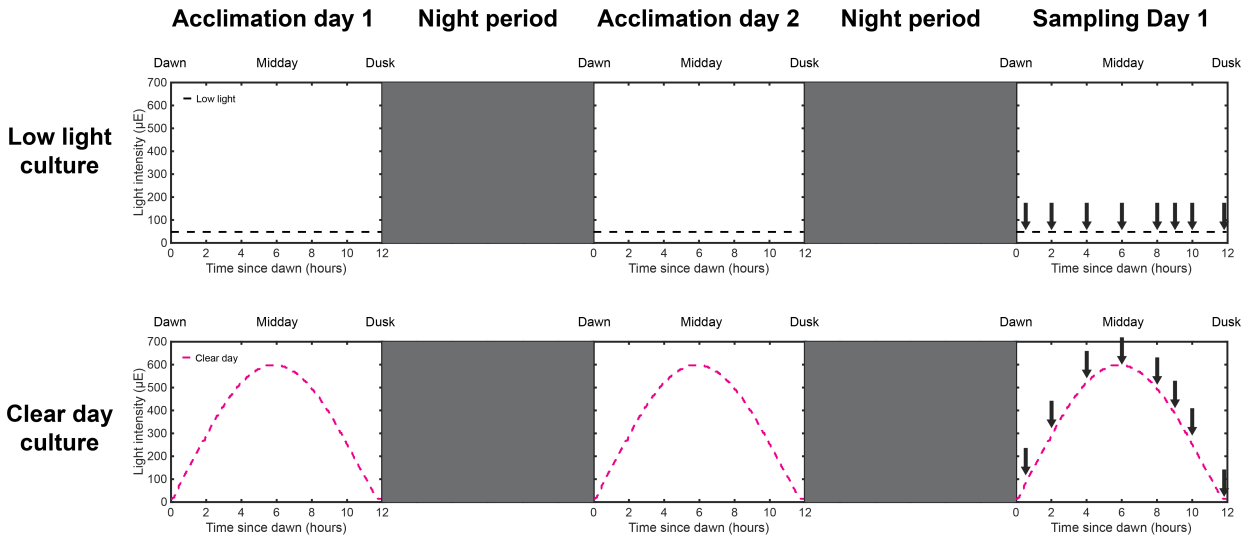

Figure 3.3: Culture set up to compare the effects of Low light and Clear day conditions. Cultures were maintained in exponential growth with periodic dilution while receiving constant bubbling of 1% CO_2 in air. Cultures were grown in either Low light (black, top) or Clear day (magenta, bottom) conditions for 12 hours, followed by 12 hours of darkness, for 2 cycles (days) to acclimatize the cultures to the light conditions. Cultures were grown in their respective condition for a third day and sampled at the times indicated with arrows (Sampling day 1) for measurement of RNA levels (Figure 3.1) and levels of phosphorylated RpaA and RpaB (Figure 3.10).

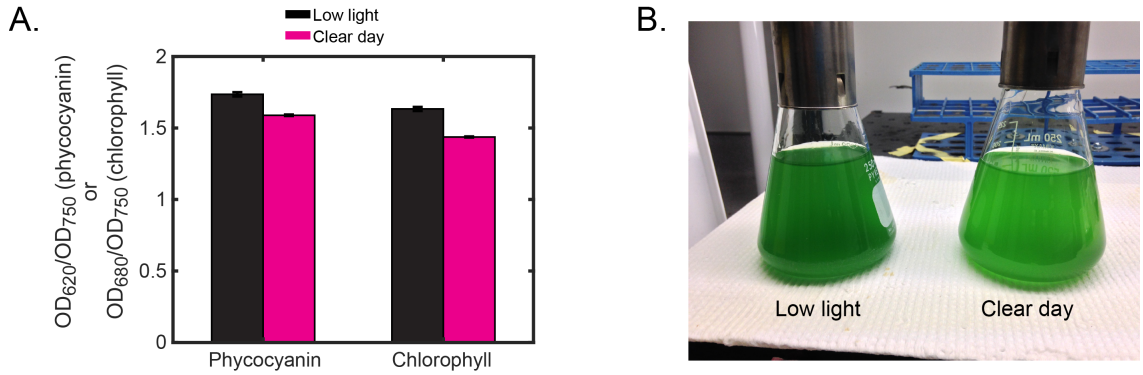

Figure 3.4: Pigment levels of cyanobacteria grown under Low light or Clear day conditions. A. Estimation of phycocyanin and chlorophyll levels in cells grown under Low light (black) or Clear day (magenta) conditions for two days, measured at midday on the third day. Phycocyanin and chlorophyll levels were estimated by measuring optical density of the culture at 620 nm or 680 nm, respectively, and normalized to optical density at 750 nm to account for differences in cell density. Error bars show the standard deviation of three independent measurements. Cells grown under Clear day conditions show lower levels of both phycocyanin and chlorophyll. B. Image of cells harvested from cultures grown under Low light (left) and Clear day (right) for two days, at midday on the third day (OD750 = 0.3). The more yellow-green color of the cells from Clear day conditions compared to the blue-green color of the Low light grown cells is indicative of diminished levels of phycocyanin. Cells grown in Clear day roughly twice as fast at midday compared to Low light cells (∼6 hour doubling in Clear day compared to ∼12 hour doubling time in Low light). This suggests that the cells adjusted the levels of photosynthetic components to cope with the increased photon flux of Clear day, while using this extra energy to allow for faster growth.

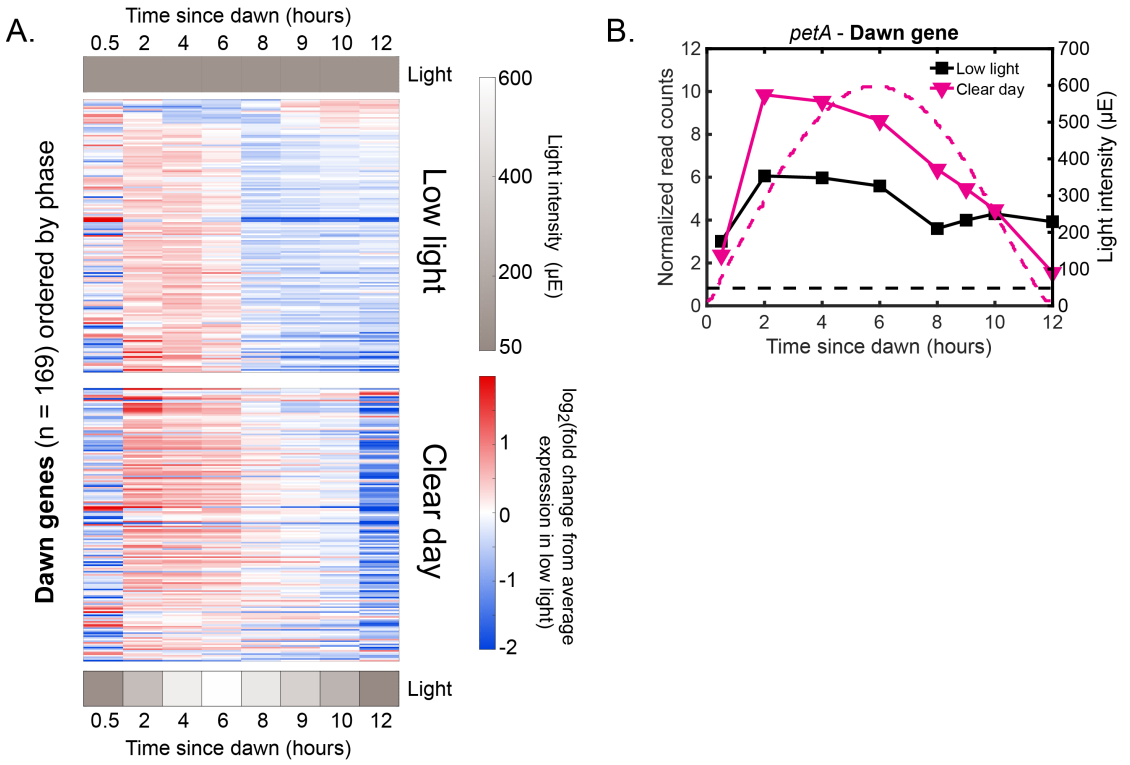

Figure 3.5: Clear day conditions modulate the expression of dawn genes. A. Gene expression dynamics of all dawn genes (n = 169) under Low light (top) and Clear day (bottom) conditions. Gene expressed was measured at each time point using RNA sequencing and expressed as the Log2 fold change from the average expression of the gene over all time points in the Low light condition. Genes are plotted in the same order in both heat maps and were sorted by phase under constant light conditions [90]. Dawn genes tend to be expressed more highly under Clear day conditions compared to Low light conditions. B. Gene expression dynamics of the representative dawn gene *petA* under Low light (black) and Clear day (magenta) conditions (data from A, left y-axis). The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis.

3.5). Thus, the changes in light intensity of the Clear day condition modulate the expression dynamics of circadian genes.

To more systematically determine how changes in light affect the expression of circadian genes, we exposed cells to rapid increases or decreases in light intensity. We chose conditions that simulate the abrupt and transient changes light intensity that occur in nature from events such as changes in cloud cover (Figure 3.1D). We focused our analysis on dusk genes, as these genes showed more significant changes in expression under Clear day conditions. Cells were grown in Low light or Clear day conditions for 3 days and then exposed to a rapid increase (*High light pulse*, Low light culture —Figure 3.6A) or decrease in light intensity (*Shade pulse*, Clear day culture —Figure 3.6B) for 1 hour prior to returning the cultures to

their original condition (Figure 3.7). We conducted the rapid increases in light intensity at 8 hours after dawn, when most dusk genes were highly expressed in Low light cells.

The expression of dusk genes was rapidly modulated in response to abrupt changes in light intensity. Dusk gene expression precipitously decreased after the rapid increase in light intensity of the High light pulse (Figure 3.6C). In contrast, dusk gene expression increased after the rapid decrease in light intensity of the Shade pulse (Figure 3.6D). These results suggest that the expression of dusk genes responds inversely to the change in light intensity experienced by cells. When cells were transitioned back to their original condition (High light to Low light —Figure 3.6C, Shade to Clear day —Figure 3.6D), dusk gene expression rapidly changed again in a manner inversely correlated to the change in environmental light. Interestingly, dawn genes tended to show the opposite behavior of dusk genes, albeit with less dramatic expression changes (Figure 3.8). These results demonstrate that the mRNA levels of circadian genes will change in response to rapid and repeated changes in light intensity as occur in nature.

Next, we asked whether these changes in gene expression resulted from changes in recruitment of RNA polymerase (RNAP) to dusk genes. We carried out chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) of RNAP in cells immediately before and after 15 or 60 minutes of exposure to rapid changes in light intensity. We find that RNAP enrichment increased upstream of dusk genes after exposure to Shade, while RNAP enrichment decreased upstream of dusk genes after exposure to High light, correlating strongly with the change in expression of the downstream dusk gene (Figure 3.6E, Figure 3.9). These results strongly suggest that RNAP recruitment to dusk genes is inhibited by rapid increases in light intensity (Low light to High light, Shade to Clear day), while RNAP recruitment to dusk genes is enhanced by exposure to decreases in light intensity (High light

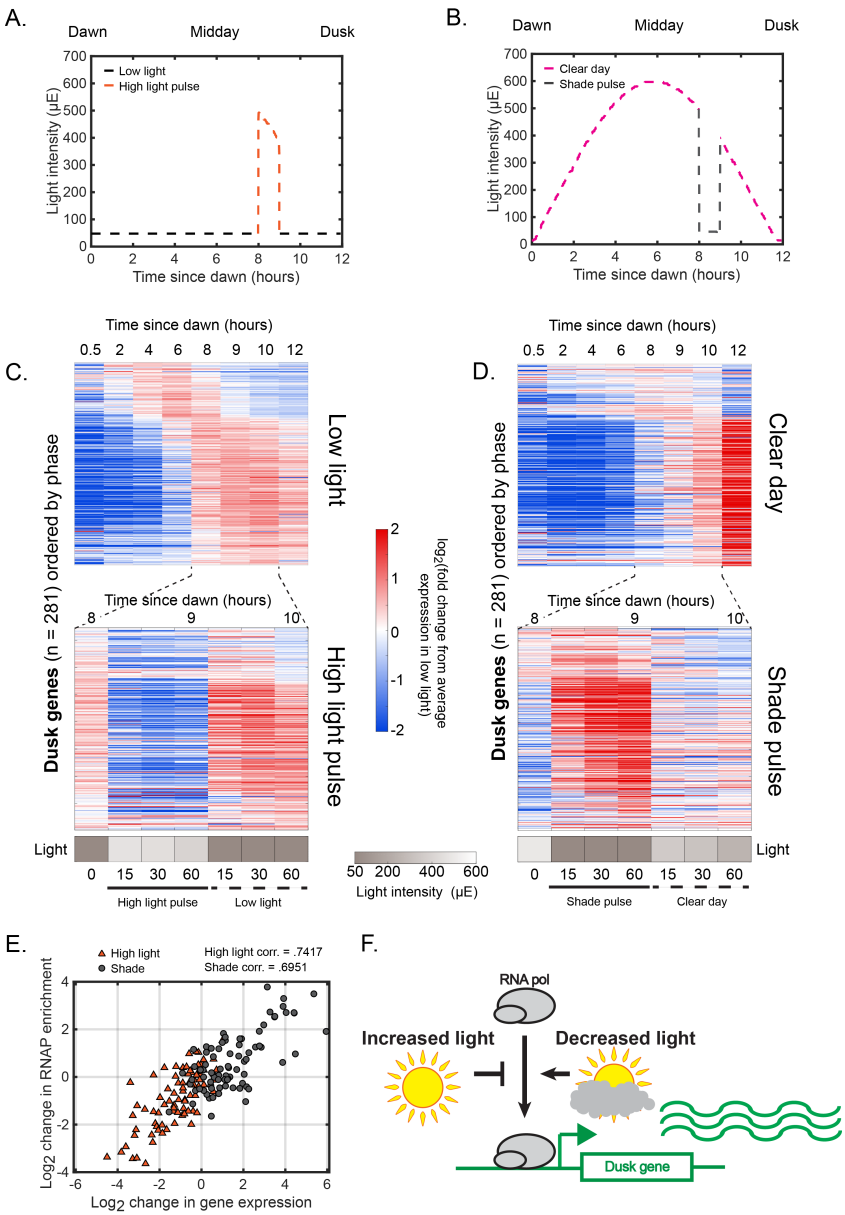

Figure 3.6: Rapid changes in light intensity modulate the recruitment of RNA polymerase to dusk genes to change dusk gene expression. A. Light intensity profiles of Low light (black) and High light pulse (orange) conditions, in units of E (μmol photons m^{-2} s^{-1}). B. Light intensity profiles of Clear day (magenta) and Shade pulse (gray) conditions, in units of μE. C. Gene expression dynamics of all dusk genes (n = 281) under Low light (top) and High light pulse (bottom) conditions. Gene expression was measured using RNA sequencing and expressed as the Log2 fold change from the average expression of the gene over all time points in the Low light condition. Genes are plotted in the same order in both heat maps and were sorted by phase under constant light conditions [90]. Light intensity at each time point in the High light pulse condition is indicated in a heat map adjacent to the corresponding time point. D. Gene expression dynamics of all dusk genes (n = 281) under Clear day (top) and Shade pulse (bottom) conditions, plotted as in C. E. Correlation between change in dusk gene expression and the change in enrichment of RNAP upstream of that gene after rapid changes in light intensity. The change in gene expression of a dusk gene (x-axis) and the corresponding change in RNAP enrichment upstream of that gene (y-axis) from the original condition after 60 minutes in High light (orange triangles) or Shade (gray circles), plotted for the 82 dusk genes with detectable RNAP peaks in their promoters (Methods). The correlation coefficient between change in RNAP enrichment and change in downstream gene expression for High light and Shade data is indicated above the plot. The correlation between RNAP enrichment change and downstream dusk gene expression also holds after 15 minutes of exposure to High light or Shade (High light correlation = .6386, Shade correlation = .6806, changes after 15 minutes). F. Regulation of RNAP recruitment to dusk genes by changes in light intensity. High light conditions repress the recruitment of RNAP to dusk genes (High light, Clear day —midday), while conditions where light intensity decreases (Shade, Clear day —sunset) promote the recruitment of RNAP to dusk genes.

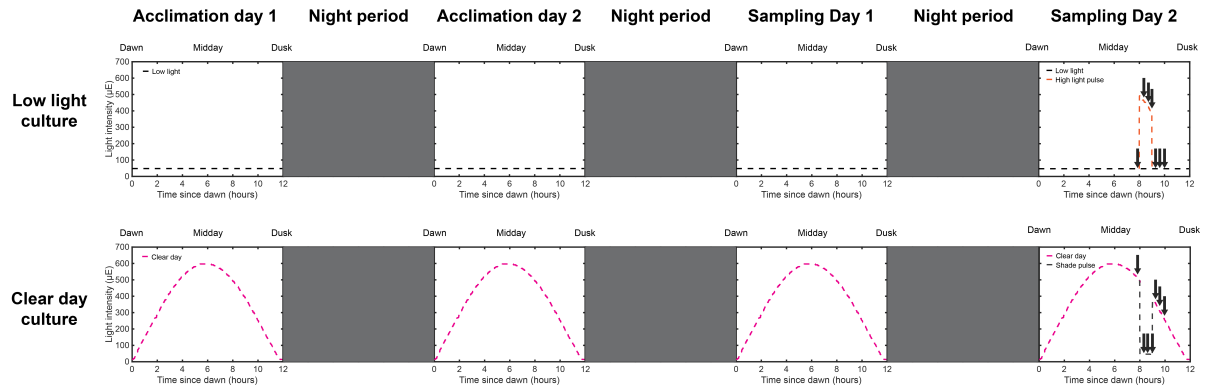

Figure 3.7: Culture set up for comparison of effects of High light and Shade pulse conditions. Cultures were maintained in exponential growth with periodic dilution while receiving constant bubbling of 1% CO_2 in air. Cultures were grown in either Low light (black, top) or Clear day (magenta, bottom) conditions for 12 hours, followed by 12 hours of darkness, for 3 cycles (days) to acclimatize the cultures to the light conditions. Cultures were grown in their respective condition for a fourth day and then exposed at 8 hours after dawn to either the High Light pulse (orange, top) or Shade Pulse (gray, bottom) prior to returning the culture to its original condition (Sampling day 2). Cultures were sampled at the times indicated with arrows for measurement of RNA levels (Figure 3.6), levels of phosphorylated RpaA and RpaB (Figure 3.10), and ChIP of RNA polymerase, RpaA, and RpaB (Figures 3.6,3.12).

to Low light, Clear day to Shade, Clear day —sunset) to elicit changes in levels of dusk gene mRNAs (Figure 3.6F).

3.3.2 Regulation of dusk gene expression by RpaA and RpaB under dynamic light regimes

We next explored the role of the known regulators of dusk gene expression - RpaA and RpaB - in controlling changes in expression of dusk genes under our dynamic light regimes. RpaA and RpaB are both OmpR-type response regulator transcription factors that are expected to activate or repress gene expression in a phosphorylation-dependent manner [18]. To determine whether our light regimes changed levels of RpaA∼P and RpaB∼P, we measured relative levels of RpaA∼P and RpaB∼P from cells grown under different light regimes and correlated these levels with changes in dusk gene expression (Figure 3.10A-I, Figure 3.11).

Levels of RpaA∼P increased from dawn to dusk in a similar manner in cells grown in Low light or Clear day conditions, even during exposure to abrupt changes in light intensity

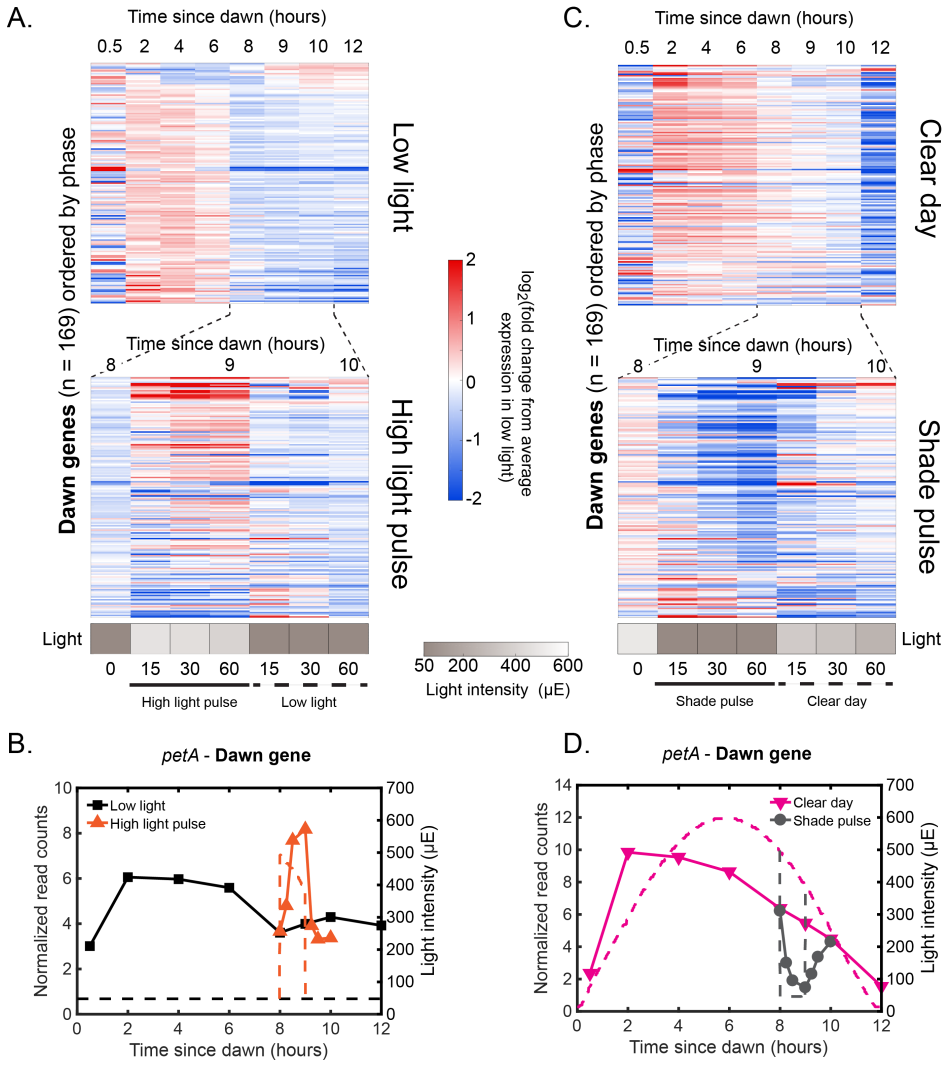

Figure 3.8: Rapid changes in light intensity affect dawn gene expression. A. Gene expression dynamics of all dawn genes (n = 169) under Low light (top) and High light pulse (bottom) conditions. Gene expressed was measured at each time point using RNA sequencing and expressed as the Log2 fold change from the average expression of the gene over all time points in the Low light condition. Genes are plotted in the same order in both heat maps and were sorted by phase under constant light conditions [90]. Light intensity at each time point in the High light pulse condition is indicated in a heat map adjacent to the corresponding time point. Dawn genes tend to increase in expression after exposure to the High light pulse. B. Gene expression dynamics of the representative dawn gene *petA* under Low light (black) and High light pulse (orange) conditions (data from A, left y-axis). The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis. C. Gene expression dynamics of all dawn genes (n = 169) under Clear day (top) and Shade pulse (bottom) conditions, plotted as in C. Dawn genes tend to decrease in expression after exposure to the Shade pulse. D. Gene expression dynamics of the representative dawn gene *petA* under Clear day (magenta) and Shade pulse (gray) conditions (data from C, left y-axis). The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis.

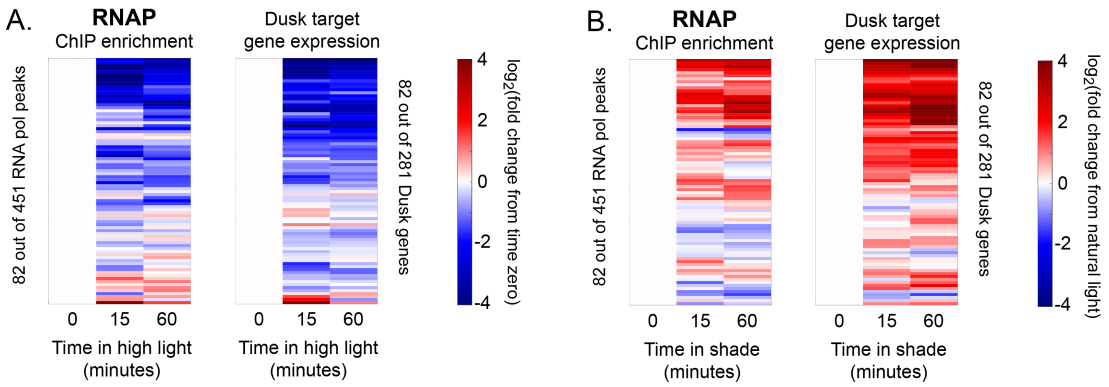

Figure 3.9: Changes in RNAP enrichment upstream of dusk genes after rapid changes in light intensity. A. Changes in enrichment of RNAP upstream of dusk genes during High light pulse conditions (left heat map) and corresponding changes in target dusk gene expression (right heat map). ChIP enrichment (left heat map) is expressed as the Log2 fold change of enrichment at the given time point in High light from enrichment in Low light conditions (time zero). The right heat map shows the change in expression of the gene target of the corresponding RNAP peak, measured as in figure 3.6. Gene expression was expressed as the Log2 fold change of expression at the given time point in High light from expression in Low light conditions (time zero). An RNAP peak and its target gene are aligned horizontally in the two heat maps. B. Changes in enrichment of RNAP upstream of dusk genes during Shade pulse conditions (left heat map) and corresponding changes in target dusk gene expression (right heat map). ChIP enrichment (left heat map) is expressed as the Log2 fold change of enrichment at the given time point in Shade from enrichment in Clear day conditions (time zero). The right heat map shows the change in expression of the gene target of the corresponding RNAP peak. Gene expression was expressed as the Log2 fold change of expression at the given time point in Shade from expression in Clear day conditions (time zero). An RNAP peak and its target gene are aligned horizontally in the two heat maps. Genes are ordered the same in A and B.

(Figure 3.10A-C). This suggests that conversion of timing information from the core oscillator to RpaA∼P levels is not affected by changes in light intensity (Figure 3.10J). Under our dynamic light regimes, the expression of the RpaA∼P-dependent dusk gene *Synpcc7942_1567* substantially changes in response to changes in light intensity (Figure 3.10G-I, Figure 3.2B), highlighting that the expression of dusk genes is de-coupled from levels of RpaA∼P when light intensity changes.

In contrast, levels of RpaB∼P changed rapidly in response to changes in light intensity in a manner correlating with changes in dusk gene expression. RpaB∼P levels were lower under Clear day conditions compared to Low light conditions up until 10 hours after dawn and then increase to higher levels at sunset when light intensity is sharply decreasing, mirroring the pattern of expression of the dusk gene *Synpcc7942_1567* (Figure 3.10D,G). Further, RpaB∼P levels rapidly increased when cells were exposed to decreases in light intensity (Figure 3.10E —High light to Low light, Figure 3.10F —Clear day to Shade), and rapidly decreased when

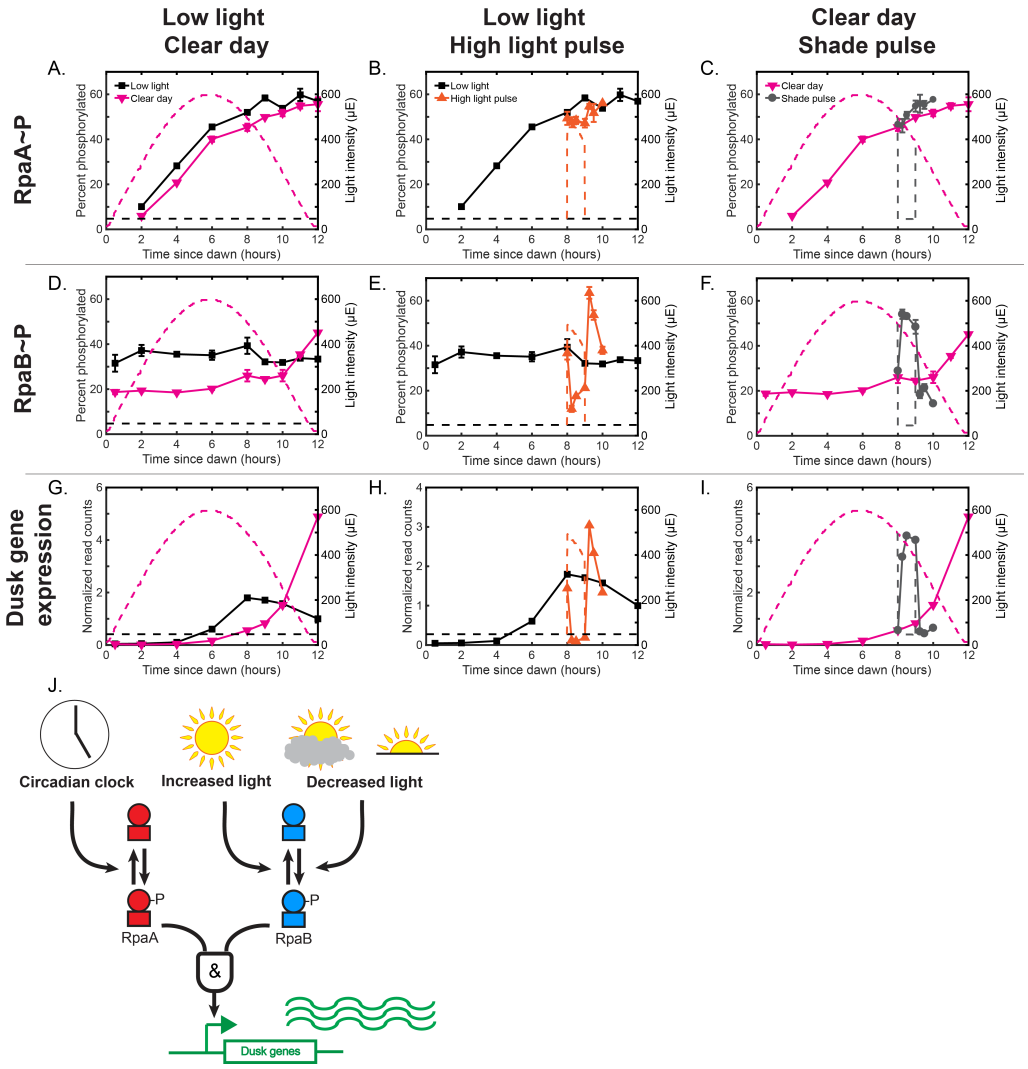

Figure 3.10: Phosphorylation dynamics of RpaA and RpaB under dynamic light regimes. A-C. Phosphorylation dynamics of RpaA under Low light vs Clear day (A), Low light vs High light pulse (B), or Clear day vs Shade pulse (C). Relative levels of phosphorylated RpaA were measured using Phos-tag Western blotting (left y-axis). Each point represents the average of values measured in two independent Western blots, with error bars displaying the range of the measured values. The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis. D-F. Phosphorylation dynamics of RpaB under Low light vs Clear day (D), Low light vs High light pulse (E), or Clear day vs Shade pulse (F), measured as in A-C. The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis. G-I. Comparison of gene expression dynamics of the representative dusk gene *Synpcc7942_1567* under Low light vs Clear day (G), Low light vs High light pulse (H), or Clear day vs Shade pulse (I), as measured in Figures 3.1 and 3.6. The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis. J. Model of the control of dusk gene expression by RpaA∼P and RpaB∼P. The circadian clock controls levels of RpaA∼P independent of changes in environmental light intensity. Decreases in light intensity favor the phosphorylation of RpaB (Shade, Clear day —sunset), while increases in light intensity favor the dephosphorylation of RpaB (High light, Clear day —midday). The strong correlation of RpaB∼P levels with dusk genes, as well as the fact that dusk genes can decrease in expression even when RpaA∼P are high, suggests that both RpaA∼P and RpaB∼P are required for robust expression of most dusk genes.

cells were exposed to increases in light intensity (Figure 3.10E —Low light to High light, Figure 3.10F —Shade to Clear day) in concert with changes in dusk gene expression (Figure 3.10H,I). These results suggest that decreases in light intensity favor the production of RpaB∼P, while increases in light intensity favor the production of unphosphorylated RpaB (Figure 3.10J). Given the implication of RpaB as a regulator of circadian gene expression and the strong correlation between levels of RpaB∼P and the levels of dusk gene mRNA, we hypothesize that RpaB∼P plays a role as an activator of dusk genes. Taken together with our understanding of the role of RpaA∼P in activating dusk genes, our results suggest that dusk gene expression is dependent on BOTH RpaA∼P and RpaB∼P (Figure 3.10J).

When phosphorylated, OmpR-type response regulators like RpaA and RpaB bind to promoters to regulate gene expression [18]. To explore how association of these proteins with promoters relates to changes in dusk gene expression, we carried out ChIP-seq of RpaA and RpaB before and after exposure to the High light and Shade pulses. We detected binding of RpaA and RpaB to only a subset of dusk gene promoters (RpaA - 56/281 dusk genes, RpaB - 42/281 dusk genes, Figure 3.12A), suggesting that RpaA and RpaB directly regulate a subset of dusk genes. However, most dusk genes are not bound by either regulator (200/281 dusk genes) and any changes in gene expression elicited by RpaA∼P or RpaB∼P levels must occur indirectly. RpaA and RpaB both bind upstream of several dusk genes, suggesting that they may directly co-regulate expression at some promoters (17 genes, Figure 3.12A). Among the genes bound by both RpaA and RpaB are three genes encoding sigma transcription factors whose expression is dependent on both the circadian clock and changes in light intensity, suggesting that RpaA and RpaB might indirectly change dusk gene expression by changing expression of sigma factors (Figure 3.13). Not all strongly light-responsive genes were bound by RpaA and RpaB, highlighting that other transcriptional regulators must be involved in

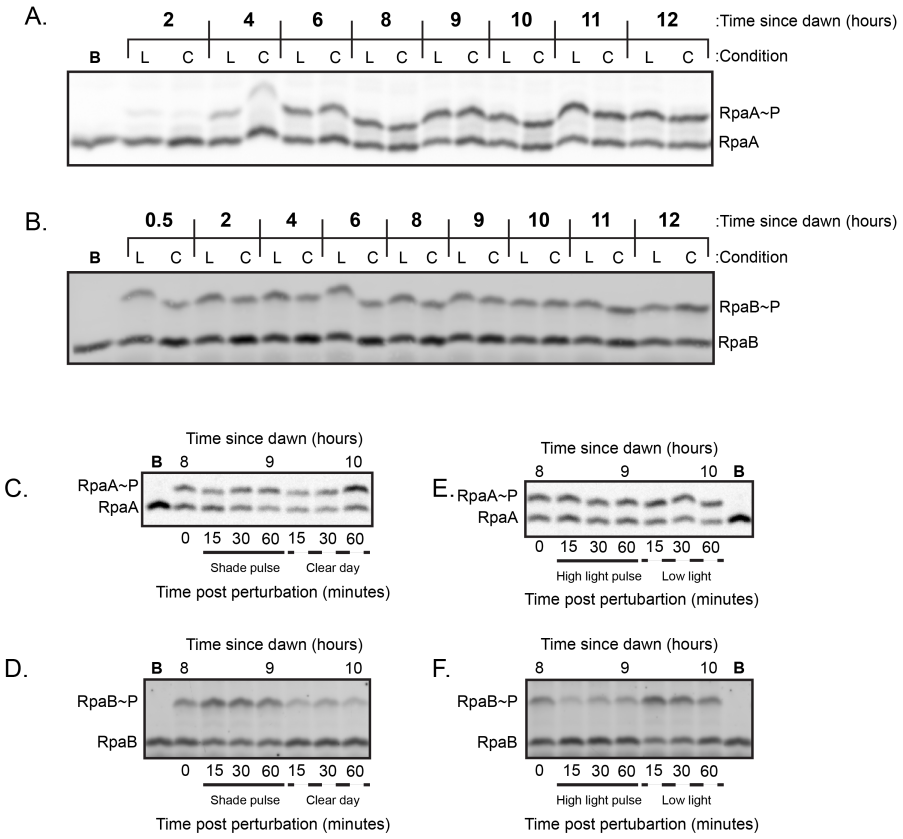

Figure 3.11: Representative Western blots used to quantify relative levels of RpaA∼P and RpaB∼P. A-B. Representative Western Blot used to quantify levels of phosphorylated RpaA (A) and RpaB (B) under Low light and Clear day conditions. Lysates were prepared from cells harvested from either Low light (L) or Clear day (C) conditions at the indicated time and subject to Phos-tag electrophoresis and Western blotting with an anti-RpaA antibody (A) or an anti-RpaB antibody (see Methods) (B). One sample was boiled prior to loading (Lane indicated with 'B') to identify the heat-labile band on the gel corresponding to phosphorylated RpaA or RpaB. C-D. Representative Western Blot used to quantify levels of phosphorylated RpaA (C) and RpaB (D) under High light pulse conditions. Time 0 refers to 8 hours since dawn under Low light conditions. E-F. Representative Western Blot used to quantify levels of phosphorylated RpaA (C) and RpaB (D) under Shade pulse conditions. Time 0 refers to 8 hours since dawn under Clear day conditions.

changing the expression of dusk genes in response to light (Figure 3.12B).

We next assessed how changes in transcription factor binding related to changes in the phosphorylation state of the factor after changes in light intensity. We find that RpaB enrichment tended to increase in response to the Shade pulse, while RpaB enrichment tended to decrease in response to the High light pulse, correlating with the corresponding changes in RpaB∼P levels (Figure 3.12C,E, Figure 3.14A,C). These results are consistent with a model in which changes in the levels of RpaB∼P lead to changes in RpaB binding to the genome (Figure 3.12H, [23]).

Strikingly, we find that RpaA binding to the genome changed in response to changes in light intensity at many sites, despite the fact that RpaA∼P levels do not change in these conditions (Figure 3.12D,E, Figure 3.14B,C). RpaA binding increased in response to the shade pulse, but decreased in response to the high light pulse (Figure 3.12D). Taken together with previous work, this suggests that the phosphorylation of RpaA allows it to associate with DNA, but changes in light intensity can further modulate the ability of RpaA∼P to bind DNA (Figure 3.12H, [48]). RpaA binding is sensitive to changes in light intensity at promoters where RpaB also binds (Figure 3.12D, yellow triangles, Figure 3.13) and at promoters where we do not detect RpaB binding (Figure 3.12D, red squares; Figure 3.15, *digC*). Interestingly, RpaA binding and downstream gene expression is insensitive to change in light intensity for some dusk genes, highlighting that RpaA activity can be isolated from changes in light intensity at some promoters (Figure 3.15, *Synpcc7942_2267*).

Given that changes in dusk gene expression occur through changes in RNAP recruitment, we asked how the association of RpaA and RpaB upstream of dusk genes related to changes in RNAP enrichment. RpaA and RpaB enrichment correlated with changes in RNAP enrichment, with association of the factors increasing in response to the Shade pulse and decreasing

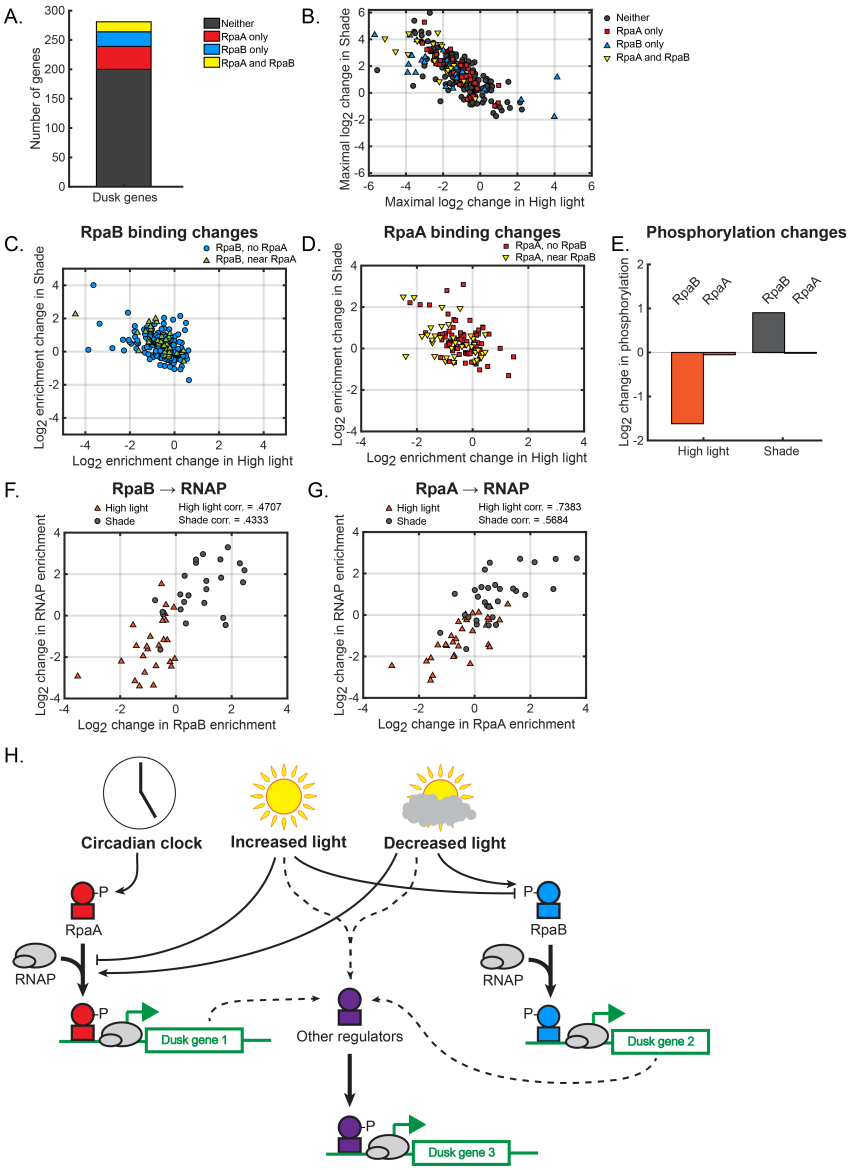

Figure 3.12: Regulation of dusk genes by RpaA and RpaB. A. Number of dusk genes targets of RpaA and/or RpaB. Target genes of binding sites of RpaA and RpaB were determined using Chromatin immunoprecipitation followed by sequencing under several different light conditions (see Methods). B. Light-responsive changes in gene expression of dusk genes. For each dusk gene, we calculated the maximal log2 change in expression in High light from Low light at 8 hours after dawn (x-axis) and the maximal log2 change in expression in Shade from Clear day at 8 hours after dawn (y-axis). C. Changes in binding of RpaB at all identified binding sites (n = 218, see Methods) in response to rapid changes in light intensity. The change in RpaB enrichment at each peak is expressed as the Log2 fold change of enrichment after 15 minutes in High light (x axis) or Shade (y-axis). RpaB peaks within 500 bp of an RpaA peak are displayed as green triangles, while RpaB peaks not close to an RpaA peak are shown as blue circles. D. Changes in binding of RpaA at all identified binding sites (n = 114, see Methods) in response to rapid changes in light intensity, expressed as in C. RpaA peaks within 500 bp of an RpaB peak are displayed as yellow triangles, while RpaA peaks not close to an RpaB peak are shown as red squares. E. Change in phosphorylation of RpaB (left bars) and RpaA (right bars) after 15 minutes in High light (orange) or Shade (gray) (data from Figure 3.10). F. Correlation between change in RpaB enrichment and the change in enrichment of RNAP upstream of the same dusk gene after rapid changes in light intensity. The change in enrichment of an RpaB at a given peak upstream of a dusk gene (x-axis) and the corresponding change in RNAP enrichment upstream of that gene (y-axis) from the original condition after 60 minutes in High light (orange triangles) or Shade (gray circles), plotted for the 27 dusk genes with detectable RpaB and RNAP peaks in their promoters (Methods). The correlation coefficient for High light and Shade data is indicated above the plot. G. Correlation between change in RpaA enrichment and the change in enrichment of RNAP upstream of the same dusk gene (n = 33) after rapid changes in light intensity, plotted as in D. H. Model of regulation of dusk genes by RpaA and RpaB. Although RpaA∼P levels are controlled by the circadian clock, changes in light intensity can affect recruitment of RpaA∼P to promoters to change expression of dusk genes. Changes in light intensity modulate recruitment of RpaB to dusk gene promoters by changing levels of RpaB∼P to change dusk gene expression. RNAP associates with promoters in conjunction with RpaA∼P or RpaB∼P to modulate the expression of dusk genes. Other light responsive regulators, some of which might be regulated by RpaA/B, must regulate the expression of dusk genes not directly controlled by RpaA/B.

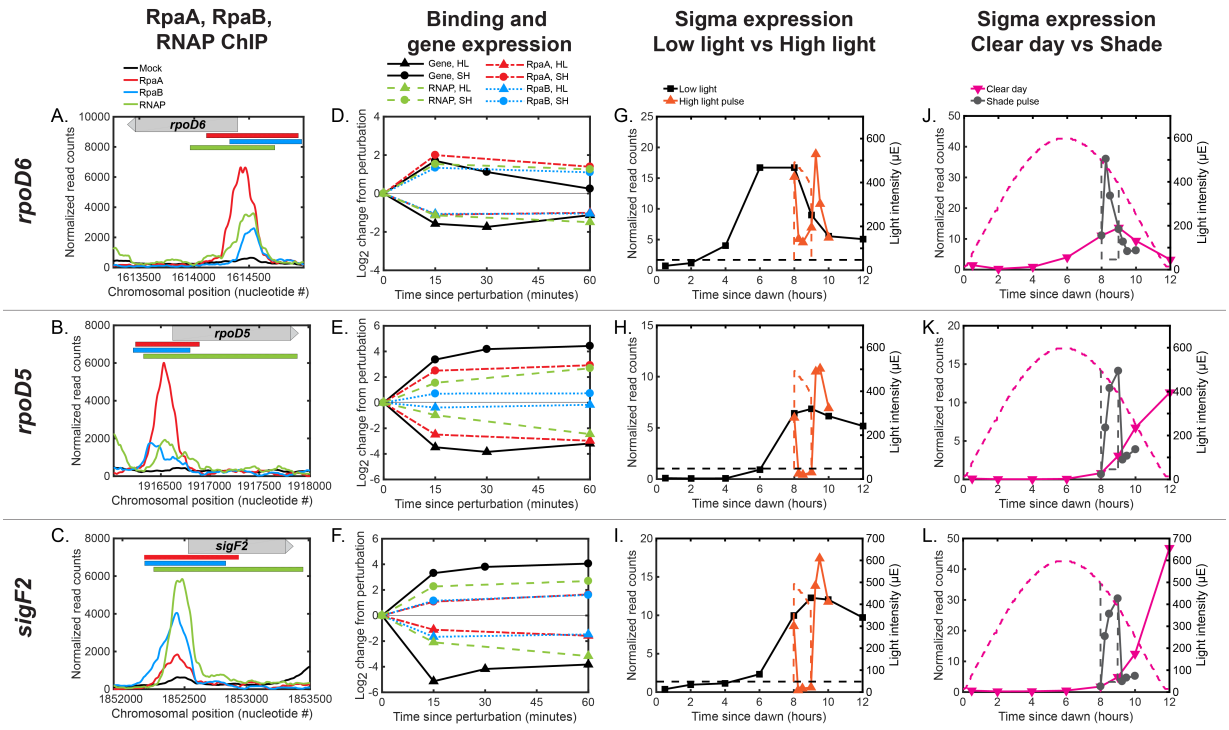

Figure 3.13: Regulation of dusk sigma factor gene expression by RpaA and RpaB. A-C. Normalized ChIP-seq signal of RpaA (red), RpaB (blue), RNAP (green) and mock IP (black) upstream of the sigma factor genes *rpoD6* (A), *rpoD5* (B), and *sigF2* (C). The location of the gene is located on the plot with a gray bar with an arrow indicating directionality of the gene. The location of RpaA, RpaB, and RNAP peaks (see Methods) are indicated on top of the blot with red (RpaA), blue (RpaB), and green (RNAP) bars. D-F. Changes in enrichment of RpaA (red), RpaB (blue), and RNAP (green) and downstream sigma factor gene expression (black) after exposure to the High light pulse (HL, triangles) or the Shade pulse (SH, circles) upstream of *rpoD6* (D), *rpoD5* (E), and *sigF2* (F). G-L. Gene expression dynamics of *rpoD6* (G,J), *rpoD5* (H,K), and *sigF2* (I,L) under Low light vs High light pulse (G-I) and Clear day vs Shade pulse (J-L) conditions. RpaA and RpaB binding changes in a correlated manner upstream of these genes. RpaA and RpaB binding also correlates with changes in RNAP enrichment and the change in downstream sigma factor expression.

in response to the High light pulse (Figure 3.12F,G). Further, RpaA and RpaB enrichment upstream of dusk genes correlates with changes in expression of the downstream gene (Figure 3.14D-I).

Taken together, our results suggest that changes in light intensity modulate the expression of some dusk genes by influencing RpaA and RpaB binding. The correlation between changes in RpaA/B, and RNAP enrichment suggest that association of RpaA/B and RpaB with dusk promoters are linked, highlighting that they might be recruited to promoters together to change the expression of dusk genes (Figure 3.12H). This is consistent with a role for RpaA and RpaB in modulating RNAP occupancy upstream of dusk genes to regulate gene expression. Other light-regulated transcription factors must be involved in controlling the expression of dusk genes that are not direct targets of RpaA or RpaB (Figure 3.12H).

3.3.3 Dusk genes group into co-expressed clusters which have distinct responses to circadian and light inputs

We sought to explore in more detail the dynamics of the response of dusk genes to changes in light intensity. We previously found that circadian genes cluster into groups of genes which show correlated and distinct changes in gene expression in response to perturbations of RpaA [48], and we wondered whether these groups also showed correlated expression in response to our dynamic light regimes. To this end, we used K-means clustering with gene expression data from our dynamic light regimes (Low light, Clear day, Shade pulse, and High light pulse), as well as from perturbations of RpaA (Constant light conditions, *rpaA* deletion, RpaA phosphomimetic overexpression, data from [48] to separate dusk genes into groups with correlated expression under these conditions.

We identify three major groups of dusk genes (35-80 genes) which show distinct coordi-

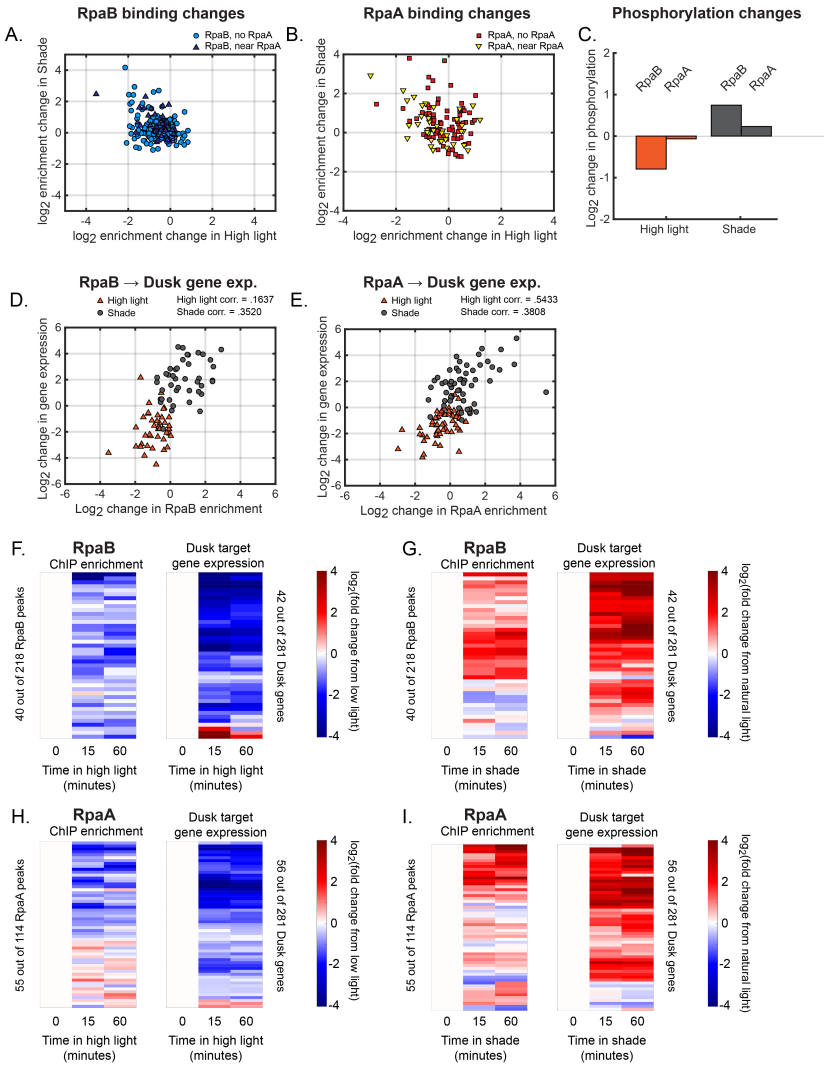

Figure 3.14: Changes in RpaB and RpaA enrichment after rapid changes in light intensity. A. Changes in binding of RpaB at all identified binding sites (n = 218, see Methods) in response to rapid changes in light intensity. The change in RpaB enrichment at each peak is expressed as the Log2 fold change of enrichment after 60 minutes in High light (x axis) or Shade (y-axis) compared to enrichment in Low light or Clear day conditions, respectively. RpaB peaks within 500 bp of an RpaA peak are displayed as green triangles, while RpaB peaks not close to an RpaA peak are shown as blue circles. B. Changes in binding of RpaA at all identified binding sites (n = 114, see Methods) in response to rapid changes in light intensity, expressed as in A. RpaA peaks within 500 bp of an RpaB peak are displayed as yellow triangles, while RpaA peaks not close to an RpaB peak are shown as red squares. C. Change in phosphorylation of RpaB (left bars) and RpaA (right bars) after 60 minutes in High light (orange) or Shade (gray) (data from Figure 3.10). D. Correlation between change in RpaB enrichment and the change in expression of the downstream dusk gene after rapid changes in light intensity. The change in enrichment of an RpaB at a given peak upstream of a dusk gene (x-axis) and the corresponding change expression of the downstream dusk gene (y-axis) from the original condition after 60 minutes in High light (orange triangles) or Shade (gray circles), plotted for the 42 dusk genes with detectable RpaB peaks in their promoters (Methods). The correlation coefficient for High light and Shade data is indicated above the plot. E. Correlation between change in RpaA enrichment and the change in expression of the downstream dusk gene after rapid changes in light intensity. Changes in RpaA enrichment upstream of the 57 dusk genes with detectable RpaA and RNAP peaks in their promoters after 60 minutes in the indicated condition, plotted as in D. Correlations after 15 minutes in the perturbed light condition were .5459 for the High light data and .3924 for the Shade data. F. Changes in enrichment of RpaB upstream of dusk genes during High light pulse conditions (left heat map) and corresponding changes in target dusk gene expression (right heat map) from Low light conditions at the 42 dusk genes with RpaB peaks in their promoters. An RpaB peak and its target gene are aligned horizontally in the two heat maps. G. Changes in enrichment of RpaB upstream of dusk genes during Shade pulse conditions (left heat map) and corresponding changes in target dusk gene expression (right heat map) from the Clear day condition, plotted as in F. Genes are ordered the same in F and G. H. Changes in enrichment of RpaA upstream of dusk genes during High light pulse conditions (left heat map) and corresponding changes in target dusk gene expression (right heat map) from Low light conditions at the 56 dusk genes with RpaB peaks in their promoters. An RpaA peak and its target gene are aligned horizontally in the two heat maps. I. Changes in enrichment of RpaA upstream of dusk genes during Shade pulse conditions (left heat map) and corresponding changes in target dusk gene expression (right heat map) from the Clear day condition, plotted as in H. Genes are ordered the same in H and I.

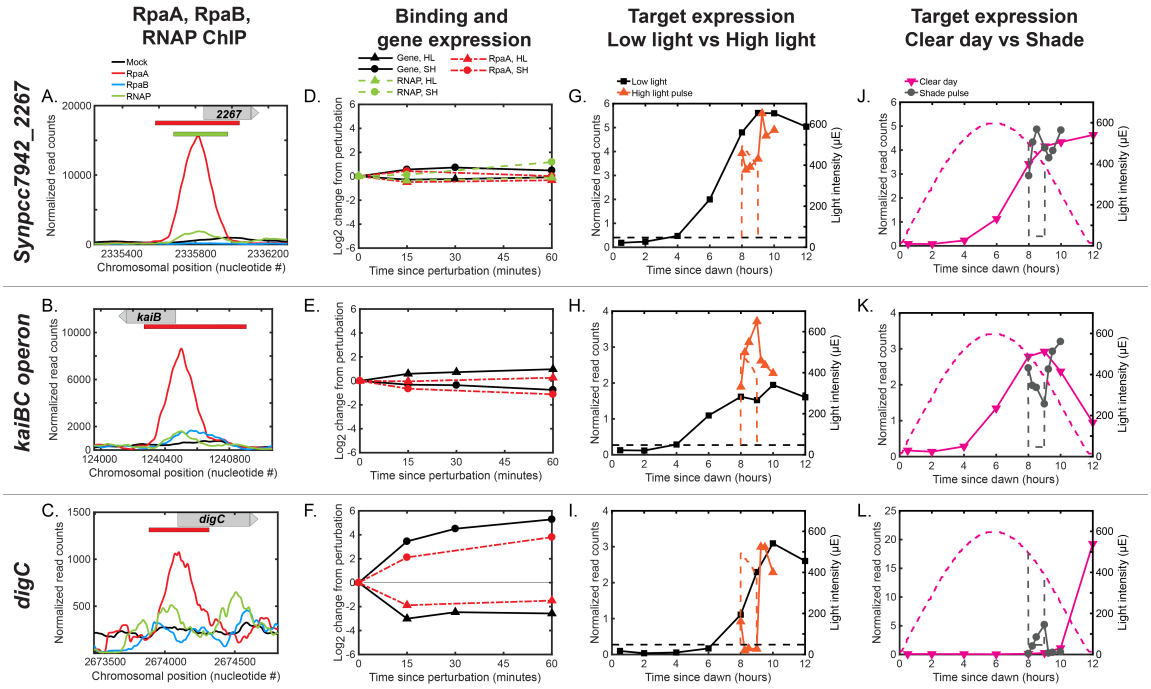

Figure 3.15: Binding behavior of RpaA under changes in light intensity at genes where RpaB does not bind. A-C. Normalized ChIP-seq signal of RpaA (red), RpaB (blue), RNAP (green) and mock IP (black) upstream of the genes *Synpcc7942_2267* (A), the *kaiBC* operon (B) and *digC* (C). The location of the gene is located on the plot with a gray bar with an arrow indicating directionality of the gene. The location of RpaA and RNAP peaks (see Methods) are indicated on top of the blot with red (RpaA) and green (RNAP) bars. No RpaB peaks were identified upstream of these genes. No RNAP peak was found upstream of *kaiB* or *digC*. D-F. Changes in enrichment of RpaA (red) and RNAP (green) and downstream gene expression (black) after exposure to the High light pulse (HL, triangles) or the Shade pulse (SH, circles) upstream of *Synpcc7942_2267* (D), the *kaiBC* operon (E) and *digC* (F). G-I. Gene expression dynamics of *Synpcc7942_2267* (G,J) *kaiB* (H,K) and *digC* (I,L) under Low light vs High light pulse (G-I) and Clear day vs Shade pulse (J-L) conditions. RpaA binding does not change upstream of the dusk gene *Synpcc7942_2267*, whose expression does not change signficantly in response to change in light intensity. In contrast, RpaA binding changes significantly upstream of the light-responsive dusk gene *digC*.

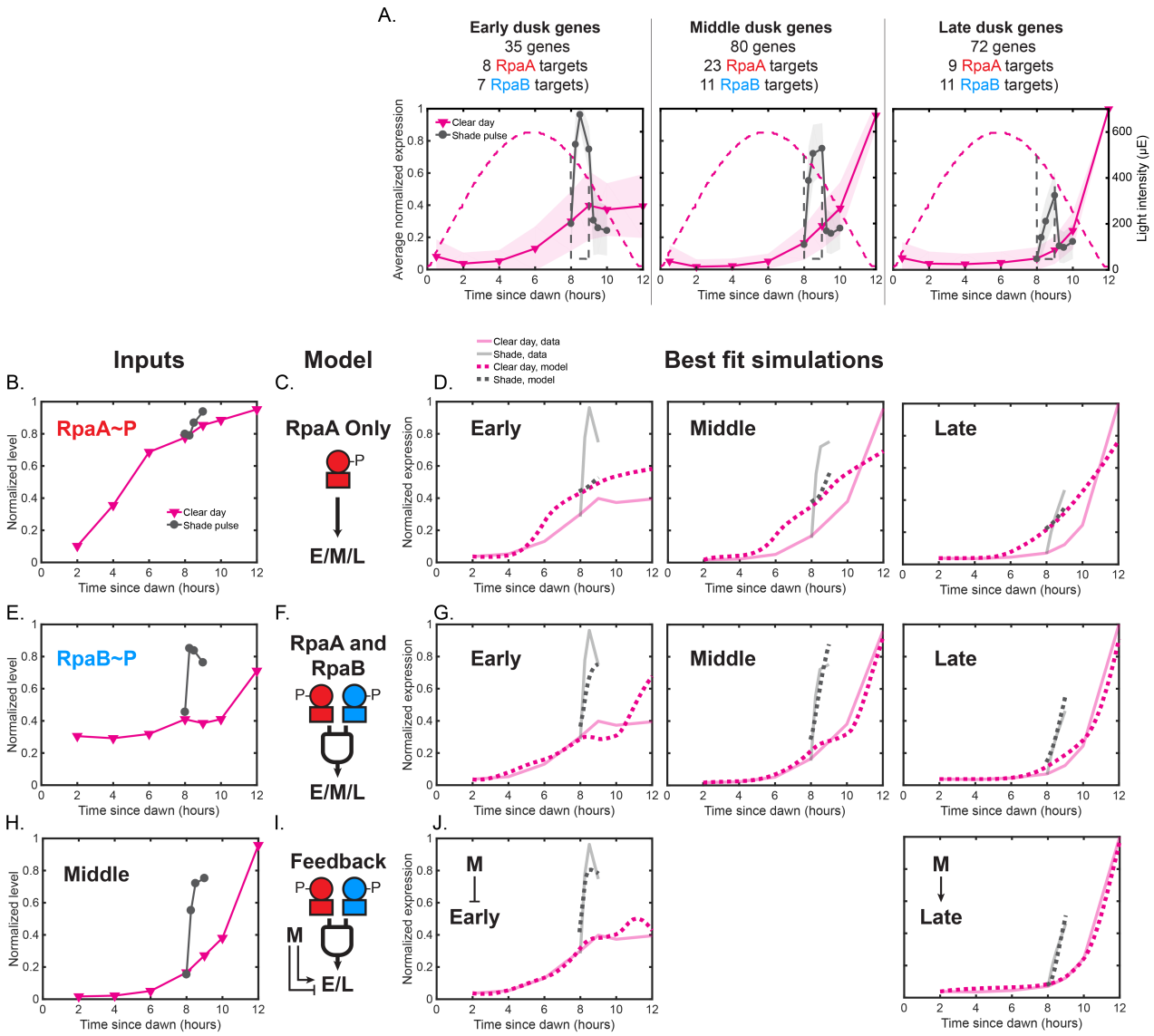

Figure 3.16: Regulation of the expression of clusters of light-responsive dusk genes under dynamic light regimes. A. Average expression profiles of genes within three major clusters identified by K-means clustering under Clear day (magenta) and Shade pulse (gray) conditions (left y-axis). Dusk genes were clustered with k-means clustering based off of their normalized expression in the four light conditions, as well as data from perturbations of RpaA activity in constant light conditions (Figure 3.17). Three major clusters emerge —Early (Left plot), Middle (middle plot), and Late (right plot) dusk genes. The number of genes within each cluster, as well as the number of genes with an RpaA or RpaB peak in their promoters (targets) is listed. The expression values of each gene across all 4 light conditions in this work were normalized to a range of 0 to 1, and the normalized expression values were averaged within each cluster. The shaded region of on the plot indicates the standard deviation of the normalized expression values within the cluster. The light intensity profile for each condition is plotted as dashed lines in the same color with values corresponding to the right y-axis. B. Normalized RpaA∼P levels under Clear day (magenta) and Shade pulse (gray) conditions used as input for mathematical models of dusk gene expression. RpaA∼P levels from all four light conditions were normalized to a range of 0 to 1. C. RpaA-only models of dusk gene expression model the expression of the Early, Middle, or Late cluster as an activation Hill function of normalized RpaA∼P levels. D. Simulations (dotted lines) of best fit RpaA-only models for Clear day and Shade pulse data (solid lines) for the Early (Left plot), Middle (middle plot), and Late (right plot) dusk genes. The average expression values of the cluster used for model fitting are plotted as solid transparent lines, and the simulated data of the best fit model are plotted as dotted lines. Data for Clear day conditions are plotted in magenta, and Shade pulse in gray. E. Normalized RpaB∼P levels under Clear day (magenta) and Shade pulse (gray) conditions used as input for mathematical models of dusk gene expression. RpaB∼P levels from all four light conditions were normalized to a range of 0 to 1. F. RpaA and RpaB models of dusk gene expression model the expression of the Early, Middle, or Late cluster as the product of an activation Hill function of normalized RpaA∼P levels and an activation Hill function of RpaB∼P levels. G. Simulations of best fit RpaA and RpaB models for the Early (Left plot), Middle (middle plot), and Late (right plot) dusk genes, plotted as in D. H. Normalized Middle cluster expression levels under Clear day (magenta) and Shade pulse (gray) conditions used as input for mathematical models of dusk gene expression. I. Feedback models add upon the RpaA and RpaB models by making the expression of the Early or Late cluster an activation (pointed arrow) or repression (blunt arrow) Hill function of Middle cluster expression levels. J. Simulations of best fit Feedback models for the Early (Left plot, Middle represses Early), and Late (right plot, Middle activates Late) dusk genes, plotted as in D. Parameters for best-fit models are listed in Table 3.1.

nated changes in gene expression over circadian time, and in response to changes in light intensity (Early, Middle, and Late dusk genes, Figure 3.16A, Figure 3.17). All three clusters show RpaA-dependent activation of gene expression (Figure 3.17B). Further, these genes increase in expression when cells experience decreases in light intensity (Figure 3.16A —Clear day to Shade, Clear day - sunset; Figure 3.17A —High light to Low light), and decrease in expression in response to increases in light intensity (Figure 3.17A —Low light to High light), correlating well with changes in RpaB∼P levels. RpaA and RpaB bind upstream of several genes within each cluster, suggesting that these regulators act directly on some genes in each group to modulate gene expression in response to circadian time and changes in light intensity (Figure 3.16A).

The three major dusk gene clusters show distinct dynamics in response to situations where dusk gene expression is activated. Over circadian time under Low light conditions, the three groups increase in expression in an ordered pattern, with Early dusk genes increasing in expression first, and Late dusk genes increasing in expression last (Figure 3.17A). Further, the genes respond differently to decreases in light intensity, with the Early gene cluster showing strongest activation to the Shade pulse and minimal effect in response to the decrease in light intensity at sunset in Clear day conditions (Figure 3.16A —left plot). Conversely, the late gene cluster responds strongly to the decrease in light under variable light, but does not strongly respond to the shade pulse (Figure 3.16A —right plot). This observation suggests that dusk genes will respond differently to a similar stimulus provided at different times of day —a phenomenon known as circadian gating of environmental responses [25]. The coordinated dynamics of these groups of genes under so many different conditions suggest that they are co-regulated in response to our perturbations.

To explore the regulation of the gene expression dynamics of these clusters, we used

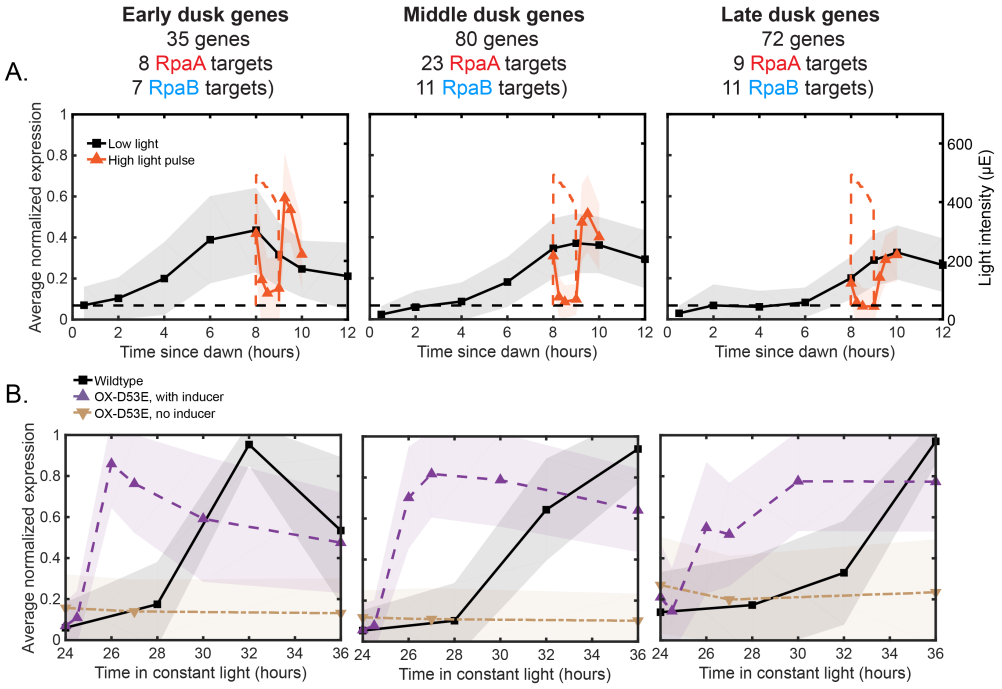

Figure 3.17: Average expression profiles of the major dusk gene clusters under various conditions. A. Average expression profiles of the Early (left plot), Middle (middle plot), and Late (right plot) dusk gene clusters under Low light (black) and High light pulse (orange) conditions (left y-axis). The expression values of each gene across all 4 light conditions in this work were normalized to a range of 0 to 1, and the normalized expression values were averaged within each cluster. The shaded region of on the plot indicates the standard deviation of the normalized expression values within the cluster. The light intensity profile for each condition is plotted as dashed lines in the same color with values corresponding to the right y-axis. B. Average expression profiles of the Early (left plot), Middle (middle plot), and Late (right plot) dusk gene clusters under Constant light conditions (left y-axis, data from [48]). Plotted are average cluster expression in wildtype cells in Constant light (black squares lines), OX-D53E cells without inducer (*rpaA*-, *kaiBC*-, Ptrc::*rpaA*(D53E), brown downward triangles), and OX-D53E cells with inducer (*rpaA*-, *kaiBC*-, Ptrc::*rpaA*(D53E), RpaA phosphomimetic induced, purple upward triangles). The expression values of each gene across all 3 conditions in Constant light were normalized to a range of 0 to 1, and the normalized expression values were averaged within each cluster. The shaded region of on the plot indicates the standard deviation of the normalized expression values within the cluster.

mathematical models to assess how well different regulatory schemes can explain the observed dynamics. To explore the basis of the response of these clusters to decreases in light intensity, we asked how well different models fit the gene expression data of the clusters under Clear day and in response to the Shade pulse (See Methods). Given that all three clusters are activated by RpaA∼P (Figure 3.17B), we began by modeling the activation of the expression of each cluster as a Hill function of RpaA∼P levels under Clear day and Shade pulse conditions (RpaA-only models, Figure 3.16B,C, Table 3.1). RpaA-only models cannot explain the light responsive dynamics of the dusk gene clusters, highlighting that dusk gene expression is not a sole function of levels of RpaA∼P (Figure 3.16D).

We sought to incorporate light-dependence of dusk gene expression into our model. Given that RpaB binds upstream of several genes within each cluster, and that RpaB∼P levels correlate with the expression of dusk genes, we model that dusk gene expression is additionally dependent on a Hill function of RpaB∼P levels (RpaA and RpaB models, Figure 3.16E,F, Table 3.1). Models of dusk gene activation by RpaA∼P and RpaB∼P can describe the activation of dusk gene expression in response to decreases in light intensity, highlighting that RpaB∼P is indeed a variable that can explain the light-dependent changes in dusk gene expression (Figure 3.16G). However, RpaA and RpaB models cannot describe the apparent circadian gating of the response of the Early cluster to decreases in light intensity —this model predicts that Early cluster gene expression should increase in response to both the Shade pulse and sunset in Clear day (Figure 3.16G —left plot).

It is known that additional regulatory interactions between members of a regulatory network can diversify the types of responses of the network to different stimuli [3]. As such, we asked whether regulatory interactions between the Early, Middle, and Late clusters could better describe the gene expression dynamics we see. We generated variations of our

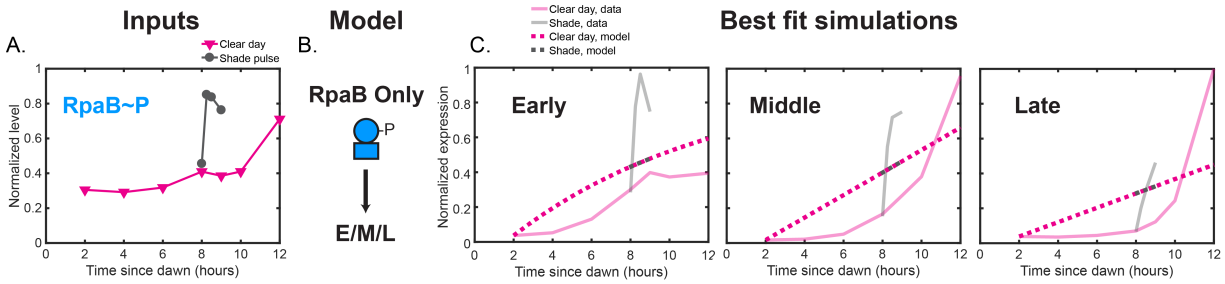

Figure 3.18: Best fit simulations of RpaB-only models in which the expression of the dusk cluster expression under Clear day and the Shade pulse is an activation Hill function of RpaB∼P levels. A. Normalized RpaB∼P levels under Clear day (magenta) and Shade pulse (gray) conditions used as model input. B. Model schematic. Dusk gene expression under Clear day and Shade pulse conditions was modelled as an activation Hill function of RpaB∼P levels ONLY. C. Best fit simulations of RpaB-only models for Early (left plot), Middle (middle plot), and Late (right plot) dusk gene clusters. The average expression values of the cluster used for model fitting are plotted as solid transparent lines, and the simulated data of the best fit model is plotted as dotted lines. Data for Clear day conditions are plotted in magenta, and Shade pulse in gray. Parameters for best-fit models are listed in Table 3.1.

RpaA and RpaB models in which the expression of a cluster is also described as positive or negatively regulated by one of the other major clusters (Feedback models, Figure 3.16H-J, Figure 3.19, Table 3.1). Interestingly, we find that a model in which Early gene expression is negatively dependent on Middle gene cluster levels and positively dependent on RpaA∼P and RpaB∼P levels can recapitulate the differences in response of the Early cluster response to the Shade pulse and sunset of Clear day (Figure 3.16G). Such an interaction is known as an incoherent feed forward loop, which can allow gene expression networks to respond faster to inputs and allow for more complex outputs [3,32,43,44]. Only models with an incoherent feed forward architecture between clusters can reproduce the gating effect (Figure 3.19). The difference in response of the Early gene cluster to the Shade pulse and sunset of Variable light in the model results from differences in Middle gene cluster levels in the two conditions —during the Shade pulse, Middle cluster levels do not reach high enough levels to inhibit Early cluster levels (Figure 3.16H —Shade pulse), but at sunset in Clear day, Middle cluster levels reach higher amounts and repress the expression of the Early cluster 3.16H —Clear day - sunset). In contrast, feedback models of Middle and Late cluster expression explain the dynamics of these clusters similarly to the RpaA and RpaB models (Figure 3.16G, data not

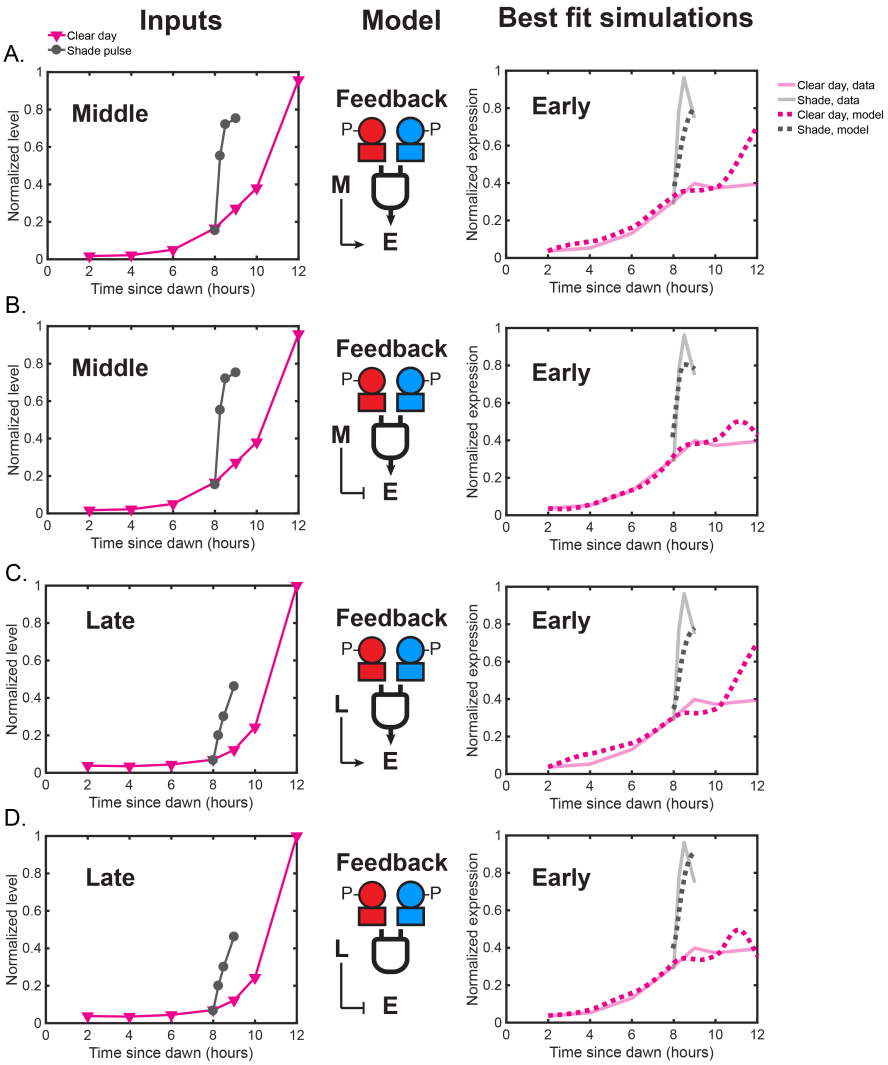

Figure 3.19: Feedback models of the expression of the Early dusk cluster. A. Feedback model in which the expression of the Early dusk cluster is a repression Hill function of Middle gene expression. The left plot shows normalized Middle cluster expression levels under Clear day (magenta) and Shade pulse (gray) conditions used as model input. The right plot shows the average expression values of the Early cluster used for model fitting (solid transparent lines), and the simulated data of the best fit model (dotted lines). Data for Clear day conditions are plotted in magenta, and Shade pulse in gray. B. Feedback model in which Early gene cluster expression is an activation Hill function of Middle cluster expression levels, presented as in A. C. Feedback model in which Early cluster expression is a repressive Hill function of Late cluster expression levels. The left plot shows normalized Late cluster expression levels under Clear day (magenta) and Shade pulse (gray) conditions used as model input. The right plot shows the average expression values of the Early cluster used for model fitting (solid transparent lines), and the simulated data of the best fit model (dotted lines). Data for Clear day conditions are plotted in magenta, and Shade pulse in gray. D. Feedback model in which Early cluster expression is an activation Hill function of Late cluster expression levels, presented as in C. Parameters for best-fit models are listed in Table 3.1.

shown). These results highlight that simple interactions between genes that respond both to circadian time (modeled as RpaA∼P levels) and changes in light intensity (modelled as RpaB∼P levels) can allow for different outputs in response to similar stimuli depending on the time of day the stimulus is applied.

3.4 Discussion

3.4.1 Understanding the integration of light intensity and circadian time to control expression

Circadian clocks provide time-of-day information to generate oscillations in physiological outputs, but it is not clear how this information is integrated with information from the environment. We take advantage of the circadian clock of the cyanobacterium *S. elongatus* to dissect how a circadian output —in this case the expression of large groups of genes —is affected by as dynamic light intensity changes that mimic those in a natural environment. We demonstrate that naturally-relevant changes in light intensity affect the transcription of circadian-clock regulated genes in cyanobacteria in a predictable manner. Situations where environmental light is increasing, such as the increase in light leading up until noon on a clear day, or when the sun emerges from behind a cloud, suppress the expression of dusk genes. In contrast, the decreasing light intensity at sunset, or in response to a cloud passing in front of the sun, leads to the up-regulation of dusk genes. Although genome-wide expression patterns have been measured in cyanobacteria grown under simulated natural light conditions [92,96], our work is novel in its comparison of Low light and Clear day conditions. These results highlight that cyanobacteria consider both information on time-of-day, and on the changes in environmental light ability, to make decisions about the levels of a set of mRNAs.

Our results suggest that decreases in light intensity are an important signal for cyanobacteria to upregulate the set of dusk gene mRNAs, but the purpose of such regulation remains unclear. The circadian clock promotes the survival of cyanobacteria under Light/Dark conditions by promoting rhythmic glycogen accumulation and breakdown to allow for alternative energy usage at night [10, 11, 65, 67]. Many RpaA-dependent dusk genes are essential for viability of cyanobacteria in Light/Dark conditions and encode enzymes involved in glycogen breakdown, glycolysis, and the oxidative pentose phosphate pathway [67]. We find that these clock-controlled carbon metabolism genes show light-responsive gene expression (Figure 3.20). In particular, an essential operon of dusk genes encoding *glgP*, a glycogen-breakdown enzyme, and *gap1*, a glycolysis enzyme, are strongly induced by both the Shade pulse and sunset in Clear day (Figure 3.20). The up-regulation of this operon and other dusk genes in response to shade onset might represent an attempt by cyanobacteria to access carbon utilization pathways to cope with a sudden decrease in light to power photosynthesis. Further, the decrease in light intensity during sunset under Clear day conditions could serve as an important signal that night is approaching. This regulatory logic would allow for dusk genes important for survival of night to be activated in advance of darkness onset, even if the timing of sunset varies due to changes in day length.

Changes in light intensity appear to orchestrate a switch between dawn and dusk gene expression, with the increase in dusk gene mRNAs always coinciding with a decrease in dawn gene mRNAs, and vice versa. Many dawn genes encode proteins involved in translation, carbon fixation, and ATP synthesis, which are critical for conversion of photosynthetic energy to increased biomass [27, 90]. The expression of dawn genes Is highest when cells experience increases in light intensity (Figure 3.5, Clear day —morning; Figure 3.8, High light pulse). The opposing effects of light intensity on dawn and dusk genes might allow cyanobacteria

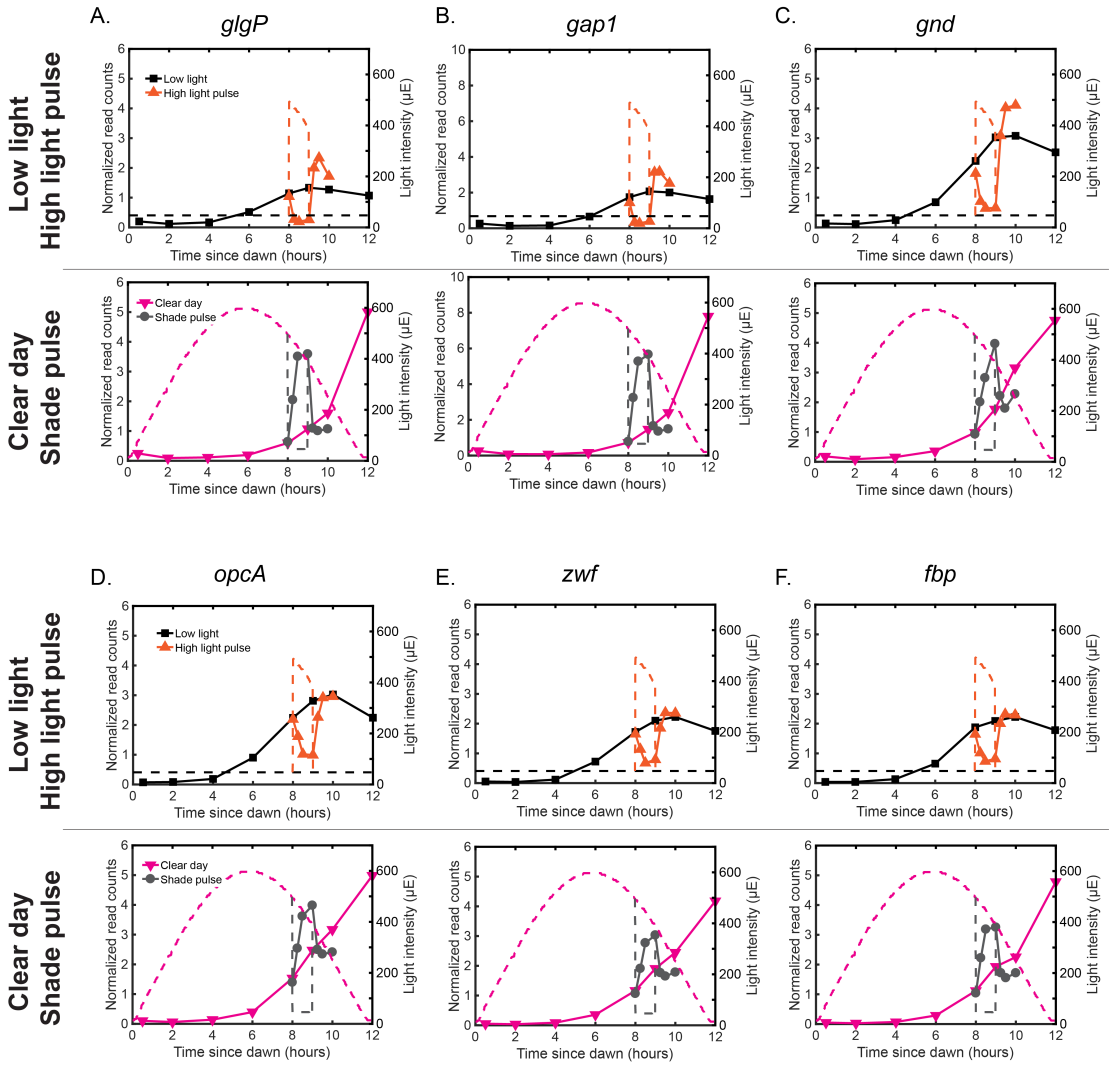

Figure 3.20: Light-dependent expression of RpaA-dependent genes that are essential for viability under L/D conditions. A. Comparison of gene expression dynamics of the dusk genes *glgP* (A), *gap1* (B), *gnd* (C), *opcA* (D), *zwf* (E), and *fbp* (F) under Low light vs High light pulse (top plot), or Clear day vs Shade pulse (bottom plot), as measured in figures 3.1 and 3.6. The light profile for each condition is plotted as dashed lines of the same color with values corresponding to the right y-axis. The expression of these dusk genes, especially *glgP* and *gap1*, is induced by shade and expressed most highly at sunset in Clear day.

to increase biomass through photosynthesis under high light conditions, while activating alternative carbon metabolism pathways for cellular maintenance in diminished light conditions. However, we observe that dynamic light regimes induced more substantial changes to dusk genes compared to dawn genes, suggesting that changes in light have a larger role in regulating enzymes involved in glycogen breakdown, glycolysis, and the oxidative pentose phosphate pathway which are expressed at dusk.

Interestingly, one subset of RpaA-dependent dusk genes is not drastically affected by changes in light intensity (Figure 3.21, Figure 3.15, *Synpcc7942_2267*, *kaiBC* operon), suggesting that the clock can regulate time-of-day dependent expression of some genes independent of light-responsive pathways. The *kaiB* and *kaiC* genes are among dusk genes which do not significantly respond to changes in light intensity in the manner of the major dusk gene clusters. The relatively minor changes in *kaiB* and *kaiC* expression under dynamic light regimes (Figure 3.15) might serve to maintain proper transcriptional feedback to the Kai oscillator in the face of changing growth rates and metabolic conditions [86].

3.4.2 Mechanisms of integration of distinct signal transduction pathways

We find that the expression of clock-regulated genes changes in cells grown under dynamic light conditions, but RpaA∼P levels are not drastically affected (Figures 3.1, 3.6, 3.10). Thus, other variables must be invoked to describe the dynamics of clock-regulated genes under more natural conditions (Figure 3.16, RpaA-Only models). The dynamics of large groups of dusk genes under dynamic light regimes can be explained by a model in which expression is a function of RpaA∼P levels, which encode information about the time-of-day, AND RpaB∼P levels, which encode information about changes in light intensity (Figure

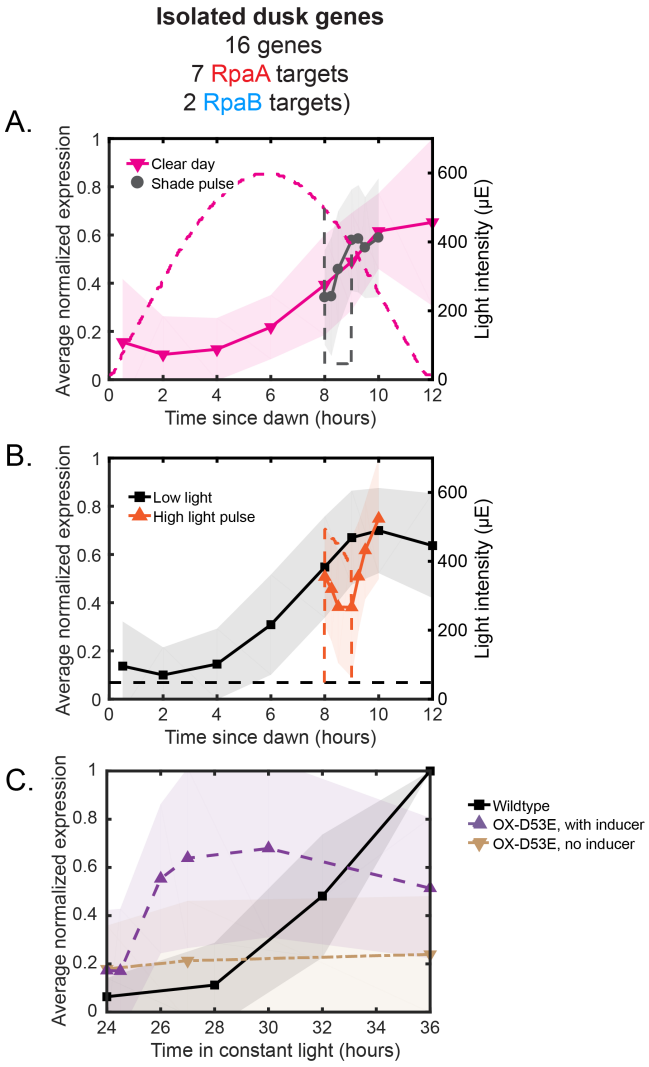

Figure 3.21: Average expression profiles of a cluster of light-intensity independent dusk genes ('Isolated' cluster, identified in the same k-means clustering clustering analysis in Figure 3.16. A-B. Average expression profiles of the Isolated dusk gene cluster under Clear day (magenta) vs Shade pulse (gray) conditions (left y-axis, A) and Low light (black) vs High light pulse (orange) conditions (left y-axis, B). The expression values of each gene across all 4 light conditions in this work were normalized to a range of 0 to 1, and the normalized expression values were averaged within each cluster. The shaded region of on the plot indicates the standard deviation of the normalized expression values within the cluster. The light intensity profile for each condition is plotted as dashed lines in the same color with values corresponding to the right y-axis. C. Average expression profiles of the Isolated dusk gene cluster under Constant light conditions (left y-axis, data from [48]). Plotted are average cluster expression in wildtype cells in Constant light (black squares lines), OX-D53E cells without inducer (*rpaA*-, *kaiBC*-, Ptrc::*rpaA*(D53E), brown downward triangles), and OX-D53E cells with inducer (*rpaA*-, *kaiBC*-, Ptrc::*rpaA*(D53E), RpaA phosphomimetic induced, purple upward triangles). The expression values of each gene across all 3 conditions in Constant light were normalized to a range of 0 to 1, and the normalized expression values were averaged within each cluster. The shaded region of on the plot indicates the standard deviation of the normalized expression values within the cluster.

3.16). Thus, we propose that RpaA∼P and RpaB∼P cooperate to regulate the expression of environmentally responsive dusk gene expression. Our results suggest several possible mechanisms for the cooperative control of gene expression by RpaA and RpaB under dynamic light regimes.

Our results support a model in which RpaB∼P binds with RNAP to promoters to directly activate the expression of some dusk genes (Figure 3.12). Thus, light-induced changes in RpaB∼P levels can directly cause changes in the expression of dusk genes. Taken together with previous studies, our results suggest that increases in light intensity lead to diminished levels of RpaB∼P, while decreases in light intensity cause increased RpaB∼P levels [17, 54]. The phosphorylation state of RpaB is regulated by the membrane bound histidine kinase NblS [41]. Changes in light intensity might regulate RpaB∼P levels by modulating NblS activity through changes in the redox state of photosynthetic electron carriers [42, 50]. NblS has been implicated in mediating gene expression responses to a wide array of stresses including high light, cold, salt, and osmotic stress [26, 46, 50, 64, 74, 75, 87, 91]. Thus, it is possible that changes in environmental conditions like temperature or salt levels regulate RpaB∼P levels to change the dynamics of dusk gene expression in response to a wider range of environmental variables. Further, we find that RpaB binds upstream of many previously identified dark-induced genes, suggesting that increases in RpaB∼P in response to darkness onset could activate the expression of these genes (29/196 dark-induced genes, [24]). Our attempts to cleanly modulate RpaB activity to further explore its role as a regulator of dusk genes were unsuccessful, in part because the *rpaB* gene is essential (Lpez-Redondo et al., 2010).

Direct interactions between RpaA∼P and RpaB∼P at some promoters could explain the cooperation between these two factors in regulating gene expression. RpaA binding

changes in a phosphorylation-independent manner in conjunction with RpaB and RNAP upstream of several sigma factor genes (Figure 3.13). Thus, RpaB might interact with RpaA at these promoters to influence RpaA binding in response to changes in light intensity. It has been shown that two OmpR-type response regulators, PhoB and TctD, bind in tandem to some promoters in Pseodomonas aeruginosa, but the two regulators appear to have opposing effects on downstream gene expression [7]. In contrast, RpaA and RpaB appear to bind in tandem at the sigma factor promoters with RNAP to cooperatively activate gene expression (Figure 3.13). We attempted to explore this interaction further by abrogating the binding of RpaB to sigma factor promoters through mutation of identified HLR1 motifs [33], but even substantial mutation of the motifs did not fully abrogate association of RpaB with the promoters in vivo (data not shown). Further study of the interaction between RpaA and RpaB at sigma factor promoters could reveal the mechanistic principles underlying the integration of information encoded in two parallel OmpR-type response regulators to control gene expression in response to multiple inputs.

Control of expression of most dusk genes by RpaA and RpaB must occur indirectly, as RpaA and RpaB only bind to a subset of dusk gene promoters. We previously proposed that RpaA controls the expression of most dusk genes by regulating the expression of the dusk sigma factors *rpoD6*, *rpoD5*, and *sigF2* [48]. In this work we find that RpaA∼P and RpaB∼P appear to cooperate to generate gene expression patterns that are dependent on both time-of-day and dynamic light conditions (Figure 3.13). Thus, RpaA∼P and RpaB∼P might cooperate to indirectly regulate the expression of many dusk genes by controlling sigma factor expression levels. Transcriptional regulators like sigma factors are subject to much post-transcriptional regulation [20]. It is thus possible that changes in light intensity affect the activity of dusk transcriptional regulators in a manner independent of RpaB∼P

regulation.

3.4.3 Phosphorylation-independent regulation of RpaA recruitment

Changes in light intensity modulate dusk gene expression by causing phosphorylation-independent changes in RpaA and RNAP recruitment to dusk genes (Figure 3.12), but the mechanism of this phenomenon is unclear. The phosphorylation of OmpR-type response regulators like RpaA and RpaB promotes dimerization to promote the binding of the dimer to a bi-partite DNA motif [18]. Previously we showed that RpaA binds directly to the promoters of the dusk genes *rpoD6* and *kaiBC* in vitro when phosphorylated [48]. RpaA binding upstream of *rpoD6* is sensitive to changes in light intensity and correlates with light-responsive *rpoD6* expression (Figure 3.13), but RpaA binding upstream of *kaiB* is unaffected by light intensity (Figure 3.15). These results suggest that the association of RpaA with promoters in vivo is dependent on light intensity only at specific promoters. Direct cooperation with RpaB to bind DNA cannot explain the light-sensitive changes in binding of RpaA when RpaB does not bind nearby (Figure 3.12D; Figure 3.15, *digC*). As RpaA binding changes in conjunction with RNAP binding at most promoters (figure 3.12G), it is possible that RpaA∼P and RNAP associate with promoters cooperatively, and that light dependent regulation of RNAP recruitment could affect RpaA association with promoters. It has been proposed that RpaA and RpaB are regulated by cellular redox state through control of thiol oxidation states by thioredoxin [31], and such a mechanism might explain light-dependent effects on RpaA and/or RpaB DNA binding. Further exploration of how RpaA regulates gene expression at different promoters can provide information about how response regulator activity can be regulated independent of phosphorylation at specific sites.

3.4.4 Gating of responses of dusk genes to decreases in light intensity

We find that groups of dusk genes show quantitative differences in activation to decreases in light intensity applied at different times of day (Figure 3.16). Genes within the Early cluster respond strongly to the Shade pulse, but not sunset in Clear day. In contrast, genes in the Late cluster respond weakly to the Shade pulse but strongly to sunset in Clear day. Such a phenomenon, referred to as circadian gating of environmental responses, is well appreciated in plants [25] and has been described for the dark-induction of genes in *S. elongatus* [24]. Cyanobacteria might use circadian gating of the response to decreases in light intensity to access different sets of genes if nighttime is imminent (Late genes) than in response to a transient change in light in the afternoon (Early genes). We demonstrate that the gating of the Early cluster can be explained if this group of genes is negatively regulated by another group of dusk genes. Interestingly, the sigma factor genes *rpoD6*, *rpod5*, and *sigF2* belong to the Early, Middle, and Late dusk clusters, respectively, suggesting that these transcription factors could regulate the expression of genes within the respective cluster (Figure 3.13). It is possible that the regulatory interactions between dusk clusters described in our mathematical models could represent direct or indirect regulatory interactions between these three sigma factors [20]. Further study of the regulatory network controlling the expression of dusk genes can reveal further principles about how circadian clocks interact with the environmental responses of organisms.

3.4.5 Implications for understanding cyanobacterial growth in the real world

Our work demonstrates that the transcriptional state of cyanobacteria in the natural world will constantly fluctuate in the face of changing environmental conditions. Cyanobacteria are attractive synthetic biology targets for the production of useful compounds from ambient sunlight [15]. However, optimization of production of a desired molecule will require a systems-level understanding of how the regulatory pathways within a cyanobacteria interact with each other and with the constantly changing conditions of the environment to control growth [93]. We have defined simple environmental variables (increases/decreases in light intensity) and cellular variables (RpaA∼P and RpaB∼P levels) which can explain the dynamics of a large set of growth-related genes in cyanobacteria under real-world conditions. Such variables can be used to inform models of how cyanobacteria will function in nature, but a complete understanding will require an exploration of protein- and metabolite-level responses. Our study lays a groundwork for understanding how regulatory pathways within cyanobacteria interact to regulate growth under constantly changing natural environments.

3.5 Methods

3.5.1 Cyanobacterial strains

Most experiments were conducted in a pure wildtype background of *Synechococcus elongatus* PCC7942 (ATCC catalog number 33912).

For RNAP ChIP experiments, we used a strain in which the β' subunit of RNA polymerase (*Synpcc7942_1524*) was C-terminally tagged with a 3x FLAG epitope (a gift from Ania Puszynska). To make the strain, wildtype *S. elongatus* was transformed with a plasmid

encoding the *Synpcc7942_1524* gene with sequence encoding a 3X GS linker and a 3X FLAG epitope inserted before the stop codon, targeted to insert at the native locus of the gene. A downstream kanamycin resistance cassette was used for selection. The strain was confirmed by sequencing colony PCR fragments that amplified the modified regions of the gene, and the presence of the tagged subunit was confirmed with Western blotting.

3.5.2 Construction of light apparatus

To grow the cyanobacteria in different light profiles, we constructed an apparatus to control the intensity of four high powered LED arrays (parts list in Figure 3.22). 'Warm white' LED arrays (1 in. x 1 in., Bridgelux) were chosen because of maximal overlap with the phycobilisome absorption spectrum. An LED array was mounted on a heatsink (Nuventix) and powered by a Flexblock LED driver (LEDdynamics) wired in the 'boost only' configuration (Figure 3.23)). The intensity of the LEDs was controlled by varying the voltage input into the DIM line of the Flexblock between 0 and 10 V. We used a digital potentiometer (AD7376, Analog Devices) as a controllable 10 V source. The voltage output of the digipot was controlled via serial peripheral interface with an Arduino Uno board (Arduino) (see Figure 3.24). Each LED array was controlled separately, and a single array was sufficient to grow a single 750 mL culture of *S. elongatus*. All wires carrying substantial currents from the main power supply to the LED arrays were rated 18 AWG, and all other wires were rated 22 AWG. The relatively low voltage of the main power supply (18 V) is essential for being able to turn off the LED arrays completely.

Part name	Digikey part number	Current price ($)	Quantity
PWR SUP MEDICAL 18V 8.3A 150W	EPS439-ND	73.71	1
CONN RCPT 8CONT DIN SLD PNL MNT	SC2007-ND	5.64	1
LEDDynamics Flexblock BUCK BOOST 48V, 700 mA	788-1038-ND	19.99	4
AD7376 digital potentiometer	AD7376ARWZ10-ND	8.66	4
AC to DC power supply, 10VDC, 275 mA	993-1233-ND	4.68	2
BXRA-30E1200-B-03, Bridgelux, Warm white, LED	Not sold at Digikey. Need to order from: AMBIT ELECTRONICS, INC. 30 West Street Spring Valley, NY 10977-4760 Phone: +1 (845) 425-2254 Fax: +1 (845) 425-5625 E-mail: Bridgelux@ambitelectronics.com Website: http://www.ambitelectronics.com	10.47	4
Aavid thermalloy Spotlight 47W heat sink	1061-1092-ND	9.50	4
Arduino Uno Board Rev3	1050-1024-ND	21.49	1

Figure 3.22: Parts for variable light source. The table includes the parts chosen for their specific properties. The remaining parts, such as wires, heat shrink tubing, thermal paste for mounting the LEDs on the heat sinks, proto-boards, and housing are quite general and specific brands are not necessary.

Line	Connection
DIM GND	GND of 10 V power supply/Arduino
DIM	Wipe of AD7376 potentiometer (Pin 16)
Vin+	+ of 18V power supply AND + of LED array
Vin-	GND of 18V power supply
LED+	NC (not connected)
LED-	- of LED array

Figure 3.23: Wiring the FlexBlock LED driver. The FlexBlock LED driver needs to be connected in a 'boost only' configuration (see spec sheet for more details), with the connections as described below.

Pin	Connection
1	+ of 10 V power supply
2	GND (shared GND between that of 10V power supply and Arduino GND)
3	GND
4	GND
5	pin 10 on Arduino (or any other pin designated as a Slave Select, such as 5, 6, and 9)
6	+5V of Arduino
7	pin 13 on Arduino (SCLK)
8	NC (not connected)
9	NC
10	NC
11	pin 11 on Arduino (MOSI)
12	+5V of Arduino
13	NC
14	+ of 10 V power supply
15	NC
16	DIM line of FlexBlock

Figure 3.24: Wiring the AD7376 potentiometer. We used the SOIC-16 housing for the AD7376 potentiometer for ease of soldering to wires. The table indicates how each pin was connected. The length of the GND wire from the Arduino board to the shared ground needs to be kept short ($\sim$2 in. or less) for SPI communication.

3.5.3 Calibrating light conditions

A single LED was mounted to shine perpendicular to the ground and isolated from other light sources with light-blocking materials. A single 750 mL cyanobacterial culture in a 150 cm^2 BD Falcon Tissue culture flask (Fisher Scientific) was placed beneath the LED, tilted such that the broad face of the culture was almost perpendicular to the incoming light. Each LED was calibrated by passing a known voltage input to the LEDs and recording the intensity of the light in Einsteins at the position of the surface of the culture directly beneath the LED using a LI-COR LI-250A light meter equipped a quantum sensor. To access a greater dynamic range of light intensity values, we calibrated the lights to give known intensity values to the cultures when grown at either of two distances from the light source —raised towards the lights to access higher light intensities, or lowered away from the lights to access lower light intensities.

To define the Clear day conditions, we used Light intensity values measured by the Ground-based Atmospheric Monitoring Instrument Suite, Rooftop Instrument Group on March 23rd, 2013 ([66], Figure 3.1C). We used this light intensity profile to define the rate of change of light intensity in our Clear day condition, with a maximal light intensity of 600 μE —an intensity approaching the very bright light intensity on the Earth's surface in nature. The Shade pulse condition was defined by dividing the intensity value of our Clear day profile by 10 fold between 8 and 9 hours after dawn. The High light pulse was defined as the intensity of the Clear day condition between 8 and 9 hours after dawn. Low light cultures were grown continuously at 50 μE. We generated the dynamic changes in light intensity of our conditions by changing the intensity of the LED every three minutes by passing the calibrated voltage value corresponding to the appropriate light intensity of our defined profile. After the light profile, a low Voltage value was sent to the LEDs to turn them

off and the cells were kept in a light-tight enclosure for 12 hours during the dark period. Cultures were grown in BG-11M media supplemented with 10 mM HEPES pH 8.0 at 30° C and continuously bubbled with 1% CO_2 in air. Cells were not grown with antibiotics during the course of the experiment.

3.5.4 Purification of anti-RpaA and anti-RpaB antibodies

Recombinant RpaA was purified as previously described [82]. To purify recombinant RpaB, we cloned the rpaB into the pET48-b+ plasmid (Novagen) and overexpressed Trx-His-tagged RpaB in Novagen Tuner (DE3) competent cells carrying this plasmid by adding 300 μM IPTG to mid-log phase cultures. RpaB was purified from cell lysate using Ni-NTA chromatography as described previously [21]. The Trx-His tag was cleaved from RpaB and removed using a subsequent Ni-NTA step as described [21]. Purified, cleaved RpaB was dialyzed into a buffer containing 20 mM HEPES-KOH, pH 8.0, 150 mM KCl, 10% w/v glycerol, and 1 mM DTT. Protein concentration was measured with the Pierce BCA assay, and aliquots were flash frozen and stored at -80° C.

Anti-RpaB serum was generated by immunization of two rabbits with purified RpaB by Cocalico Biologicals (Reamstown, PA). RpaA- and RpaB- conjugated Affigel 10/15 resin (Bio-rad) was prepared following manufacturer's instructions as described previously [21]. Anti-RpaB serum was first passed over an RpaA-conjugated resin and the flowthrough collected to subtract cross-reacting antibodies. Anti-RpaB antibodies were then purified from the flowthrough using and RpaB-conjugated resin as described previously [21]. The same process was repeated to purify anti-RpaA antibodies using rabbit serum described previously [48], passing the serum over an RpaB-conjugated resin and purifying with an RpaA-conjugated resin. No cross reactivity of the purified anti-RpaA and anti-RpaB antibodies

for the opposite regulator was detected via ELISA assay.

3.5.5 Measurement of RpaA∼P and RpaB∼P levels

At the specified timepoint, 10 mL of OD750 0.3 cyanobacterial culture were collected on cellulose acetate filters and flash frozen prior to storage at -80° C. Cell lysates for Western blotting were prepared from the collected cells as described previously [48]. Equal amounts of cell lysate (10-15 μg) were resolved on Phos-tag (Wako Chemicals) acrylamide gels and transferred to nitrocellulose membranes as described previously [21]. Membranes were probed with 1/5000 dilution of purified anti-RpaA and anti-RpaB antibody. RpaA blots were then incubated with goat anti-rabbit HRP-conjugated secondary antibody and developed using the Pierce Femto chemiluminescence kit. The exposed blots were imaged with an Alpha Innotech Imaging station. RpaB blots were incubated with Goat anti-Rabbit Westerndot 585 (Thermo Fisher) antibody and imaged with a Typhoon Imager. The intensities of the bands corresponding to unphosphorylated and phosphorylated RpaA/B were quantified using Imagequant software (GE Healthcare Life Sciences) using rubber band background subtraction. Images of exposed RpaA blots were inverted in Adobe Photoshop prior to quantification. The percent of RpaA (or RpaB) phosphorylated was quantified as the intensity of the RpaA∼P band divided by the sum of the intensities of the RpaA and RpaA∼P bands, multiplied by 100. Values reported in Figure 3.10 are the average of two separate measurements from replicate Western blots. The trends seen in Figure 3.10 were reproducibly observed between separate biological replicates of the light condition time courses.

3.5.6 RNA sequencing

At the specified timepoint, 25 mL of OD750 0.3 cyanobacterial culture were collected on cellulose acetate filters and flash frozen prior to storage at -80° C. Cells were resuspended in RNAprotect Bacteria reagent (Qiagen), and 1/3 of the cells were resuspended in a buffer containing 15 mg/mL lysozyme, 10 mM Tris-Cl, 1 mM EDTA pH 8, and 50 mM NaCl and incubated for 10 minutes. RNA was purified from the lysed cells using the Qiagen RNeasy Mini Kit. Ribosomal RNA was depleted from 1.25 μg of purified RNA using the Ribo-Zero bacteria rRNA removal kit (Illumina). Strand-specific RNAseq libraries were prepared from the depleted RNA using the Truseq Stranded mRNA Sample prep kit (Illumina) and sequenced on an Illumina HiSeq machine by the Bauer Core Facility at the Harvard FAS Center for Systems Biology. Sequencing reads were aligned to the *S. elongatus* genome as described previously [48], with samples averaging 8 million aligned reads. We quantified expression of a gene by counting the number of aligned sequencing reads between the start and stop of each gene, and normalized these values between all samples from the light conditions in this work using median normalization as described previously [4, 48].

3.5.7 Definition of Circadian Genes

We defined a subset of previously identified circadian genes to focus our analysis on. We began with a list of 856 previously described reproducibly circadian genes [48, 90]. We next required that these genes have a Cosiner amplitude [37] of greater that 0.15 under Constant light conditions [90]. We also required that the gene display expression of at least 1 read per nucleotide in at least one time point of the RNA sequencing experiments in this study. These filters produce a list of 450 high confidence circadian genes.

We noted that genes classified as dawn (class 2) and dusk (class 1) genes under Constant

light conditions [90] showed maximal expression at a different time of day under our Low light/Dark conditions. As such, we redefined dawn genes as those genes with a phase of 40° to 189° under Constant light conditions [90], and dusk genes as those with a phase of 190° to 360° and 0° to 39°, as determined by the Cosiner algorithm [37]. These definitions produce a list of 169 high confidence dawn genes, and 281 high confidence dusk genes. The expression of our redefined circadian genes under Constant light conditions is plotted in Figure 3.2.

3.5.8 ChIP Sequencing

At the specified time point, 120 mL of OD750 0.3 cyanobacterial culture were removed and crosslinked with 1% formaldehyde at 30° C for 5 minutes in front of a light source. Crosslinking was quenched with 125 mM glycine. Crosslinked cells were washed twice with phosphate buffered saline, pelleted, and flash frozen prior to storage at -80° C.

Pellets were resupended in 1 mL of BG-11M supplemented with 500 mM L-proline and 1 mg/mL lysozyme and incubated at 30° C for 1 hour to digest the cell wall. Cells were collected and resuspended in a Lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 1x Roche Complete EDTA-free Protease Inhibitor Cocktail) prior to shearing in a Covaris E220 Adaptive Focus System (Peak Incident Power = 175; Duty Factor = 10%; Cycles per burst = 200; Time = 160 s). The lysates were cleared via centrifugation, and concentration was determined via the Pierce BCA Assay.

For a given pulldown, 800 μg of lysate was incubated overnight at 4° C in 500 μL of lysis buffer with 8 μg of anti-RpaA, anti-RpaB, or FLAG M2 mouse monoclonal antibody (Sigma-Aldrich) for RNAP pulldowns. A mock pulldown was carried out in which equal amounts of lysate from every time point of the time course (Shade 0, 15, 60 minutes, High

light 0, 15, 60 minutes) in a total of 800 μg was incubated with 8 μg of rabbit Igg. Next, 35 μL of Dynabeads protein G (Thermo Fischer Scientific) equilibrated in lysis buffer were added and the sample was incubated with mixing for 2 hours at 4° C. The beads were washed and DNA was eluted and purified as described previously [48].

Sequencing libraries were prepared from the purified ChIP DNA using the NEBNext Ultra II DNA Library Prep Kit (New England Laboratory, Ipswich, MA). Libraries were sequenced on an Illumina HiSeq2000 instrumentby the by the Bauer Core Facility at the Harvard FAS Center for Systems Biology. We created sequencing libraries of ChIP experiments from two separate biological repeats of the time course experiment. Reads were aligned to the S. elongatus genome as described previously [48]. resulting in an average of 3 million aligned reads for replicate 1, and 5 million aligned reads for replicate 2.

3.5.9 ChIP-seq Analysis

The aligned read data per genomic position was smoothed with a Gaussian filter (window size = 400 base pairs, standard deviation = 50). Each data set was normalized to the Mock ChIP-seq experiment and peaks which were significantly enriched above the Mock were identified in each data set using a previously described [48] custom-coded form of the Peak-seq algorithm [70]. Within each replicate time course for a given protein, we compiled a list of peaks which were enriched at least 3.5 fold over the Mock experiment at the position of highest ChIP signal. Finally, we required that a peak be detected in both replicates for it to be considered. This analysis generated 114 RpaA peaks, 218 RpaB peaks, and 451 RNAP peaks. To calculate enrichment for a peak, we determined the ChIP signal at a given time point at the genomic position of the highest ChIP signal detected for that peak and divided this by the value of the Mock experiment at that position. The data plotted in this

manuscript are from replicate 2, but all trends hold in replicate 1. We assigned a gene as a target of a peak if : 1. the start codon of the gene was within 500 bp of the position of maximal ChIP signal within a peak; 2. the peak resided upstream of the gene; 3. The gene was the closest gene to that peak on the same strand.

3.5.10 K-means clustering

We calculated normalized expression values of high confidence dusk genes under our dynamic light conditions, as well as in previously described RpaA perturbations in Constant light [48]. We separately normalized the data from set of dynamic light conditions (Low light, Clear day, High light pulse, Shade pulse) and the Constant light data (Wildtype, OX-D53E cells (*rpaA*-, *kaiBC*-, Ptrc::*rpaA*(D53E)) without inducer, OX-D53E with inducer) using z-score normalization, and used this data to separate the dusk genes into 8 groups with K-means clustering using sample correlation as the distance metric. We focused our analysis on the three largest clusters which accounted for most of the dusk genes (187/281 genes).

3.5.11 Mathematical modelling

We coarse-grained each of the three groups of circadian dusk genes to a single effective gene with the average dynamics of the group (Fig. 3.16A, triangles). We modeled the interaction between these groups and RpaA∼P and RpaB∼P using a simple kinetic model of an AND gate based on that used in [43],

$$\frac{d\mathrm{X}}{dt} = B_{\mathrm{X}} + \beta_{\mathrm{X}} f(\mathrm{RpaA}{\sim}\mathrm{P}, K_{\mathrm{AX}}) f(\mathrm{RpaB}{\sim}\mathrm{P}, K_{\mathrm{BX}}) f(Y, K_{\mathrm{YX}}) - \alpha_{\mathrm{X}}\mathrm{X} \tag{3.1}$$

where X is either the Early, Mid, or Late cluster; B_X is the basal transcription rate; f is a function of the interaction of X with RpaA∼P, RpaB∼P, or another cluster Y; and β_X is the max transcription rate of that cluster. Activating interactions were treated using a simple Hill function,

$$f(u, K) = \frac{(u/K)^H}{1 + (u/K)^H} \tag{3.2}$$

where H is the Hill coefficient of interaction. Repressive interactions were treated using

$$f(u, K) = \frac{1}{1 + (u/K)^H} \tag{3.3}$$

All the parameters were fit using the range of values shown in Table 3.1 using MATLAB. We used the quality of fit to determine the ability of a model to describe the data.

Table 3.1: The results of fitting a simple model of gene expression regulation with the dynamical mRNA data from RNA-seq and clustering. The definitions of the variables are given in eq. 3.1-3.3. The error is defined as the square root of the sum of the squared deviations between simulation and data.

Model	Cluster	Figure	H	β_X	α_X	B_X	K_{AX}	K_{BX}	K_{YX}	Error
RpaA-only	Early	3.16D	4.97	1.16	1.71	0.06	0.7	-	-	0.72
RpaB-only	Early	3.18C	0.43	0.07	0.1	0.02	-	0.6	-	0.81
RpaA and RpaB	Early	3.16D	3.55	2.76	2.98	0.09	0.34	0.5	-	0.48
Feedback, M act.	Early	3.19A	2.43	2.14	1.42	0.09	0.15	0.68	0.04	0.53
Feedback, M rep.	Early	3.16J, 3.19B	4.74	2.64	0.81	0.02	0.31	0.59	0.6	0.22
Feedback, L act.	Early	3.19C	3.09	2.44	2.39	0.11	0.22	0.53	0.03	0.51
Feedback, L rep.	Early	3.19D	4.26	2.3	1.9	0.08	0.39	0.51	0.73	0.3
RpaA-only	Middle	3.16D	4.94	2.86	1.85	0.07	1	-	-	0.62
RpaB-only	Middle	3.18C	3.08	0.06	0	0	-	0.98	-	0.77
RpaA and RpaB	Middle	3.16D	4.78	1.84	1.05	0.03	0.68	0.51	-	0.24
Feedback, E act.	Middle	Not shown	4.41	1.91	1.36	0.05	0.65	0.45	0.22	0.23
Feedback, E rep.	Middle	Not shown	4.67	2.57	1.08	0.05	0.58	0.62	0.98	0.22
Feedback, L act.	Middle	Not shown	3.45	1.98	1.45	0.05	0.48	0.49	0.07	0.24
Feedback, L rep.	Middle	Not shown	4.97	1.61	0.45	0.02	0.81	0.49	0.99	0.27
RpaA-only	Late	3.16D	5	0.4	0	0	1	-	-	0.45
RpaB-only	Late	3.18C	0.92	0.01	0	0.03	-	0.08	-	0.73
RpaA and RpaB	Late	3.16D	5	1.44	0.01	0	1	0.53	-	0.17
Feedback, E act.	Late	Not shown	4.95	1.55	0	0	1	0.52	0.23	0.12
Feedback, E rep.	Late	Not shown	5	2.78	0.09	0	0.96	0.64	0.83	0.08
Feedback, M act.	Late	3.16J	2.53	1.9	0.48	0.03	0.74	0.13	0.65	0.1
Feedback, M rep.	Late	Not shown	5	1.63	0	0	1	0.53	1	0.23

Chapter 4

Conclusions and perspectives

4.1 Discussion

4.1.1 RpaA is a master regulator of circadian gene expression

In Chapter 1, we discussed the general question of understanding how information about time of day encoded in a circadian clock is used to control physiological outputs. In Chapter 2, we use a combination of approaches to demonstrate that genome-wide oscillations in gene expression in cyanobacteria are generated by controlling the phosphorylation state of a single transcription factor —RpaA. Using genetics, we show that the *rpaA* gene is required to produce oscillations in expression of circadian genes in cyanobacteria (Figure 2.1, 2.3). By inducing RpaA∼P activity in cells lacking the *rpaA* gene, we demonstrate that increases in the amount of RpaA∼P lead to the activation of dusk genes, and a repression of dawn genes, driving the dynamic expression of hundreds of genes (Figure 2.10). Further, we find that RpaA binds DNA when phosphorylated Figure 2.5, 2.7) to activate gene expression (Figure 2.5, 2.6) at a subset of dusk genes (Figure 2.9). Our results support a model where RpaA binds to promoters when phosphorylated to regulate upstream transcriptional regulators

which enact gene expression changes to the group of circadian genes as a whole (Figure 2.12).

4.1.2 Regulation of gene expression downstream of RpaA

RpaA does not bind upstream of all high amplitude dusk genes (Figure 2.9), but RpaA∼P activity leads to the induction of many high-amplitude non-target dusk genes (Figure 2.10). Thus, RpaA∼P exerts indirect control of non-target dusk genes. Interestingly, dusk genes cluster into large groups which show correlated expression patterns in response to perturbations of RpaA∼P levels (Figure 2.10). These results suggest that information on RpaA∼P levels can be converted into diverse dynamics of downstream gene expression. It is possible that these co-expressed clusters of genes represent the regulons of other clock-regulated transcription factors (Figure 2.12). Thus, by identifying factors which regulate non-RpaA target circadian genes, we can begin to understand how a single circadian input can be converted into diverse outputs.

Interestingly, the expression of three sigma factor genes (*rpoD6*, *rpoD5*, and *sigF2*) are regulated in a circadian manner by RpaA (Figure 2.9). Thus, it is possible that these sigma factors regulate the expression of non-target dusk genes to allow for indirect regulation by RpaA∼P (Figure 2.12). Kathleen Fleming in the O'Shea lab has characterized the regulons of these sigma factors in *S. elongatus* and found that they directly regulate the expression of many circadian genes (Personal communcation). Interactions between downstream regulators can diversity the outputs of signal-transduction systems [3]. In future studies, it will be worth exploring molecular interactions between the RpaA-regulated sigma factors and how they influence each other's activity [20]. By understanding the principles underlying the regulation of sigma factor activity, we can identify mechanisms behind the converstion

of RpaA∼P levels to diverse outputs. Interestingly, we find that these co-expressed groups of RpaA-dependent dusk genes also show coordinated gene expression under dynamic light conditions (Figure 3.16). This suggests that mechanistic principles governing expression of these groups of genes under constant light conditions might also be operating in more realistic dynamic environments.

4.1.3 Integration of environmental information into circadian outputs

In Chapter 1, we discussed the general question of understanding how naturally relevant changes in light intensity affect the transcriptional output of the clock. By growing cyanobacteria in controlled dynamic light regimes, we uncover that naturally-relevant changes in light intensity impinge upon the expression of clock-regulated dusk genes by influencing the recruitment of RNAP to these genes (Figures 3.1, 3.6). We find that situations in which light intensity decreases —like the decrease in light intensity during sunset on a clear day (Figure 3.1), or if a cloud passes in front of the sun to shade the earth (Figure 3.6) —lead to the activation of dusk genes. In contrast, situations where light intensity increases —like the high light intensity reached at midday on a clear day (Figure 3.1) or the rapid increase in light that occurs if shade is relieved (Figure 3.6) —lead to the suppression of the expression of dusk genes and the favoring of the expression of dawn genes (Figure 3.5, 3.8).

Our results suggest that circadian clock output pathway interacts with responses to decreases in light intensity to control the expression of a large group of dusk genes in response to environmental cues that occur prior to night-time, when clock output through RpaA∼P is high. However, it is unclear how the changes in expression of these genes affects physiological outputs in cyanobacteria. Many clock-regulated environmentally responsive genes control

processes involving glycogen processing and alternative metabolism (Figure 3.20), suggesting that these changes in gene expression might be involved in activating alternative carbon metabolism pathways. However, understanding the conversion of changes in levels of mRNAs to changes in physiological outputs requires an understanding of the translation and post-translational regulation of the factors encoded by these genes. As such, it is unlikely that we can predict the changes in physiological outputs caused by changes in gene expression data when assessing systems in complex environments.

Instead, we propose that future studies of the interaction between the clock and changes in light intensity identify physiological processes that are affected by light conditions similar to those used in this dissertation (Figure 3.1, 3.6). It is known that rates of cell division are regulated by light intensity [52] and the circadian clock [39, 77]. Further, increases in growth rate will also incur increased demand on translation of the genome to support increases in biomass. Also, changes in light intensity likely impact the usage of energy production pathways like photosynthesis or respiration. As such, we propose that future studies assess rates of translation, cell division, oxygen evolution (photosynthetic rate) and oxygen consumption (respiration rate) in response to our dynamic light conditions. These studies will clearly identify broad scale physiological outputs that are affected by our light conditions. Once relevant physiological outputs are identified, screening studies can be devised to identify potential clock-regulated genes which are involved in modulating these outputs.

4.1.4 Mechanisms of integration of signal transduction pathways

In Chapter 3, we identify several mechanisms underlying the integration of responses to light intensity with circadian output pathways to control gene expression. First, we find that changes in light intensity modulate the recruitment of RpaA to promoters independent

of changes in phosphorylation (Figure 3.12). Thus, changes in light intensity could cause pervasive changes in gene expression by inhibiting the activity of the most upstream regulator of dusk gene expression. However, the mechanistic underpinnings behind this apparent light-based regulation of RpaA binding are unclear. In Chapter 2, we demonstrate the RpaA∼P binds to specific sequences in the *kaiBC* (FIgure 2.5) and *rpoD6* (Figure 2.8) promoters. We attempted to further explore the regulation of RpaA binding to other promoters using in vitro footprinting methods. However, we could not detect binding to several promoters where we detect robust RpaA association with ChIP-seq (Data not shown). Further, RpaA activity exhibits different behaviors at different promoters in vivo (Figure 3.13, 3.15). As such, the nature of the interaction of RpaA with DNA at different promoters is unclear.

Instead of attempting to reconstitute specific interactions from purified components, we propose that future studies probe the identity of all of the factors associating with specific RpaA-regulated promoters to illuminate all of the players involved in controlling the expression of a given gene. We propose the use of in vitro promoter complex assembly from cyanobacteria cell extracts on specific promoter sequences to purify the components involved in regulating a specific promoter. We hypothesize that RpaA binding to some promoters is dependent on RNAP recruitment. Additional layers of regulation could be imparted by the requirement of an alternative sigma factor for RNAP recruitment to that promoter. Thus, changes in RNAP recruitment could affect RpaA association with promoters. A more thorough characterization of the molecules associating with specific promoters will better inform an understanding of how RpaA regulates different genes in diverse ways.

Further, we identify that the transcription factor RpaB controls the expression of clock-regulated genes by directly binding their promoters when phosphorylated (Figure 3.12). Our results suggest that RpaB∼P, in conjuction with RpaA∼P, may indirectly influence

the expression of dusk genes to which it does not bind (Figure 3.16). We hypothesized that RpaB interacts with RpaA to regulate sigma factor expression in order to indirectly regulate the expression of other circadian genes (Figure 3.13). However, our attempts to impair RpaB recruitment to these promoters through mutation of RpaB binding motifs were unsuccessful (Data not shown), and thus we were unable to further probe this hypothesis. Single molecule experiments in vitro might reveal principles governing the recruitment of RpaA, RpaB, and RNAP to the sigma factor promoters which can inform models of the cooperative recruitment of these factors to control gene expression.

4.1.5 Conclusions

In this work, we identify mechanisms governing the regulation of expression of hundreds of genes in cyanobacteria by a circadian clock (Chapter 2). Then, we identify how naturally-relevant enviromental changes affect the expression of these clock-regulated genes, and find clues on the mechanisms behind this regulation (Chapter 3). Our work lays the foundation for understanding how a simple circadian clock in a photosynthetic organism interacts with the dynamic conditions of nature to control physiology.

Bibliography

[1] Katharine Compton Abruzzi, Joseph Rodriguez, Jerome S. Menet, Jennifer Desrochers, Abigail Zadina, Weifei Luo, Sasha Tkachev, and Michael Rosbash. Drosophila CLOCK target gene characterization: implications for circadian tissue-specific gene expression. *Genes & Development*, 25(22):2374–2386, November 2011.

[2] Yagut Allahverdiyeva, Marjaana Suorsa, Mikko Tikkanen, and Eva-Mari Aro. Photoprotection of photosystems in fluctuating light intensities. *Journal of Experimental Botany*, 66(9):2427–2436, May 2015.

[3] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC, July 2006.

[4] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106, 2010.

[5] M. K. Ashby and C. W. Mullineaux. Cyanobacterial ycf27 gene products regulate energy transfer from phycobilisomes to photosystems I and II. *FEMS microbiology letters*, 181(2):253–260, December 1999.

[6] T. L. Bailey and C. Elkan. Fitting a mixture model by expectation maximization to

discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 2:28–36, 1994.

[7] Piotr Bielecki, Vanessa Jensen, Wiebke Schulze, Julia Gdeke, Janine Strehmel, Denitsa Eckweiler, Tanja Nicolai, Agata Bielecka, Thorsten Wille, Roman G. Gerlach, and Susanne Hussler. Cross talk between the response regulators PhoB and TctD allows for the integration of diverse environmental signals in Pseudomonas aeruginosa. *Nucleic Acids Research*, 43(13):6413–6425, July 2015.

[8] B. J. Binder and S. W. Chisholm. Relationship between DNA cycle and growth rate in Synechococcus sp. strain PCC 6301. *Journal of Bacteriology*, 172(5):2313–2319, May 1990.

[9] EugeniaM. Clerico, JaynaL. Ditty, and SusanS. Golden. Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. In Ezio Rosato, editor, *Circadian Rhythms*, number 362 in Methods in Molecular Biology, pages 155–171. Humana Press, January 2007. DOI: 10.1007/978-1-59745-257-1_11.

[10] Spencer Diamond, Darae Jun, Benjamin E. Rubin, and Susan S. Golden. The circadian oscillator in Synechococcus elongatus controls metabolite partitioning during diurnal growth. *Proceedings of the National Academy of Sciences*, 112(15):E1916–E1925, April 2015.

[11] Spencer Diamond, Benjamin E. Rubin, Ryan K. Shultzaberger, You Chen, Chase D. Barber, and Susan S. Golden. Redox crisis underlies conditional lightdark lethality in cyanobacterial mutants that lack the circadian regulator, RpaA. *Proceedings of the National Academy of Sciences*, 114(4):E580–E589, January 2017.

[12] Guogang Dong, Qiong Yang, Qiang Wang, Yong-Ick Kim, Thammajun L. Wood, Katherine W. Osteryoung, Alexander van Oudenaarden, and Susan S. Golden. Elevated ATPase activity of KaiC applies a circadian checkpoint on cell division in Synechococcus elongatus. *Cell*, 140(4):529–539, February 2010.

[13] Christine Dubowy and Amita Sehgal. Circadian Rhythms and Sleep in Drosophila melanogaster. *Genetics*, 205(4):1373–1397, April 2017.

[14] Raphaelle Dubruille and Patrick Emery. A Plastic Clock: How Circadian Rhythms Respond to Environmental Cues in Drosophila. *Molecular Neurobiology*, 38(2):129–145, October 2008.

[15] Daniel C. Ducat, Jeffrey C. Way, and Pamela A. Silver. Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology*, 29(2):95–103, February 2011.

[16] Jay C. Dunlap, Jennifer J. Loros, and Patricia J. DeCoursey. *Chronobiology: Biological Timekeeping*. Sinauer Associates, Inc., Sunderland, Mass., first edition edition, December 2009.

[17] Javier Espinosa, Joseph S. Boyd, Raquel Cantos, Paloma Salinas, Susan S. Golden, and Asuncion Contreras. Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output. *Proceedings of the National Academy of Sciences of the United States of America*, 112(7):2198–2203, February 2015.

[18] Rong Gao and Ann M Stock. Molecular Strategies for Phosphorylation-Mediated Regulation of Response Regulator Activity. *Current opinion in microbiology*, 13(2):160–167, April 2010.

[19] Charles E. Grant, Timothy L. Bailey, and William Stafford Noble. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018, April 2011.

[20] Tanja M. Gruber and Carol A. Gross. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology*, 57(1):441–466, 2003.

[21] Andrian Gutu and Erin K. OShea. Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Molecular cell*, 50(2):288–294, April 2013.

[22] Mitsumasa Hanaoka, Naoki Takai, Norimune Hosokawa, Masayuki Fujiwara, Yuki Akimoto, Nami Kobori, Hideo Iwasaki, Takao Kondo, and Kan Tanaka. RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in Synechococcus elongatus PCC 7942. *The Journal of Biological Chemistry*, 287(31):26321–26327, July 2012.

[23] Mitsumasa Hanaoka and Kan Tanaka. Dynamics of RpaBpromoter interaction during high light stress, revealed by chromatin immunoprecipitation (ChIP) analysis in Synechococcus elongatus PCC 7942. *The Plant Journal*, 56(2):327–335, October 2008.

[24] Norimune Hosokawa, Tetsuhiro S. Hatakeyama, Takashi Kojima, Yoshiyuki Kikuchi, Hiroshi Ito, and Hideo Iwasaki. Circadian transcriptional regulation by the posttranslational oscillator without de novo clock gene expression in Synechococcus. *Proceedings of the National Academy of Sciences*, 108(37):15396–15401, September 2011.

[25] Carlos T. Hotta, Michael J. Gardner, Katharine E. Hubbard, Seong Jin Baek, Neil Dalchau, Dontamala Suhita, Antony N. Dodd, and Alex a. R. Webb. Modulation of environmental responses of plants by circadian clocks. *Plant, Cell & Environment*, 30(3):333–349, March 2007.

[26] Hui-Yi Hsiao, Qingfang He, Lorraine G. van Waasbergen, and Arthur R. Grossman. Control of Photosynthetic and High-Light-Responsive Genes by the Histidine Kinase DspA: Negative and Positive Regulation and Interactions between Signal Transduction Pathways. *Journal of Bacteriology*, 186(12):3882–3888, June 2004.

[27] Hiroshi Ito, Michinori Mutsuda, Yoriko Murayama, Jun Tomita, Norimune Hosokawa, Kazuki Terauchi, Chieko Sugita, Mamoru Sugita, Takao Kondo, and Hideo Iwasaki. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proceedings of the National Academy of Sciences*, 106(33):14168–14173, August 2009.

[28] Natalia B. Ivleva, Tiyu Gao, Andy C. LiWang, and Susan S. Golden. Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46):17468–17473, November 2006.

[29] Carl Hirschie Johnson, Phoebe L. Stewart, and Martin Egli. The cyanobacterial circadian system: from biophysics to bioevolution. *Annual Review of Biophysics*, 40:143–167, 2011.

[30] Sarah Joshua and Conrad W. Mullineaux. The rpaC gene product regulates phycobilisome-photosystem II interaction in cyanobacteria. *Biochimica Et Biophysica Acta*, 1709(1):58–68, August 2005.

[31] Taro Kadowaki, Yoshitaka Nishiyama, Toru Hisabori, and Yukako Hihara. Identification of OmpR-Family Response Regulators Interacting with Thioredoxin in the Cyanobacterium Synechocystis sp. PCC 6803. *PLOS ONE*, 10(3):e0119107, March 2015.

[32] Shai Kaplan, Anat Bren, Erez Dekel, and Uri Alon. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular Systems Biology*, 4:203, 2008.

[33] Anthony D. Kappell and Lorraine G. van Waasbergen. The response regulator RpaB binds the high light regulatory 1 sequence upstream of the high-light-inducible hliB gene from the cyanobacterium Synechocystis PCC 6803. *Archives of Microbiology*, 187(4):337–342, April 2007.

[34] Hiroaki Kato, Tomoyuki Kubo, Maiko Hayashi, Ikki Kobayashi, Tatsuya Yagasaki, Taku Chibazakura, Satoru Watanabe, and Hirofumi Yoshikawa. Interactions between histidine kinase NblS and the response regulators RpaB and SrrA are involved in the bleaching process of the cyanobacterium Synechococcus elongatus PCC 7942. *Plant & Cell Physiology*, 52(12):2115–2122, December 2011.

[35] Yong-Ick Kim, David J. Vinyard, Gennady M. Ananyev, G. Charles Dismukes, and Susan S. Golden. Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):17765–17769, October 2012.

[36] Nobuya Koike, Seung-Hee Yoo, Hung-Chung Huang, Vivek Kumar, Choogon Lee, Tae-Kyung Kim, and Joseph S. Takahashi. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science (New York, N.Y.)*, 338(6105):349–354, October 2012.

[37] Ken-ichi Kucho, Kazuhisa Okamoto, Yuka Tsuchiya, Satoshi Nomura, Mamoru Nango, Minoru Kanehisa, and Masahiro Ishiura. Global Analysis of Circadian Expression

in the Cyanobacterium Synechocystis sp. Strain PCC 6803. *Journal of Bacteriology*, 187(6):2190–2199, March 2005.

[38] Shinsuke Kutsuna, Yoichi Nakahira, Mitsunori Katayama, Masahiro Ishiura, and Takao Kondo. Transcriptional regulation of the circadian clock operon kaiBC by upstream regions in cyanobacteria. *Molecular Microbiology*, 57(5):1474–1484, September 2005.

[39] Guillaume Lambert, Justin Chew, and Michael J. Rust. Costs of Clock-Environment Misalignment in Individual Cyanobacterial Cells. *Biophysical Journal*, 111(4):883–891, August 2016.

[40] Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10:R25, 2009.

[41] Maria Luisa Lpez-Redondo, Felix Moronta, Paloma Salinas, Javier Espinosa, Raquel Cantos, Ray Dixon, Alberto Marina, and Asuncin Contreras. Environmental control of phosphorylation pathways in a branched two-component system. *Molecular Microbiology*, 78(2):475–489, October 2010.

[42] Eugene G. Maksimov, Kirill S. Mironov, Marina S. Trofimova, Natalya L. Nechaeva, Daria A. Todorenko, Konstantin E. Klementiev, Georgy V. Tsoraev, Eugene V. Tyutyaev, Anna A. Zorina, Pavel V. Feduraev, Suleyman I. Allakhverdiev, Vladimir Z. Paschenko, and Dmitry A. Los. Membrane fluidity controls redox-regulated cold stress responses in cyanobacteria. *Photosynthesis Research*, January 2017.

[43] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21):11980–11985, October 2003.

[44] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The Incoherent Feed-forward Loop Accelerates the Response-time of the gal System of Escherichia coli. *Journal of Molecular Biology*, 356(5):1073–1081, March 2006.

[45] Martial Marbouty, Cyril Saguez, Corinne Cassier-Chauvat, and Franck Chauvat. Characterization of the FtsZ-interacting septal proteins SepF and Ftn6 in the spherical-celled cyanobacterium Synechocystis strain PCC 6803. *Journal of Bacteriology*, 191(19):6178–6185, October 2009.

[46] Kay Marin, Iwane Suzuki, Katsushi Yamaguchi, Kathrin Ribbeck, Hiroshi Yamamoto, Yu Kanesaki, Martin Hagemann, and Norio Murata. Identification of histidine kinases that act as sensors in the perception of salt stress in Synechocystis sp. PCC 6803. *Proceedings of the National Academy of Sciences*, 100(15):9061–9066, July 2003.

[47] Joseph S. Markson and Erin K. O'Shea. The molecular clockwork of a protein-based circadian oscillator. *FEBS letters*, 583(24):3938–3947, December 2009.

[48] Joseph S. Markson, Joseph R. Piechura, Anna M. Puszynska, and Erin K. OShea. Circadian Control of Global Gene Expression by the Cyanobacterial Master Regulator RpaA. *Cell*, 155(6):1396–1408, December 2013.

[49] J. Marra. Phytoplankton photosynthetic response to vertical movement in a mixed layer. *Marine Biology*, 46(3):203–208, September 1978.

[50] Koji Mikami, Yu Kanesaki, Iwane Suzuki, and Norio Murata. The histidine kinase Hik33 perceives osmotic stress and cold stress in Synechocystis sp. PCC 6803. *Molecular Microbiology*, 46(4):905–915, November 2002.

[51] Beronda L. Montgomery. The Regulation of Light Sensing and Light-Harvesting Impacts

the Use of Cyanobacteria as Biotechnology Platforms. *Frontiers in bioengineering and biotechnology*, 2, July 2014.

[52] Alireza Moosavi Zenooz, Farzin Zokaee Ashtiani, Reza Ranjbar, and Najvan Javadi. Synechococcus sp. (PTCC 6021) cultivation under different light irradiances-Modeling of growth rate-light response. *Preparative Biochemistry & Biotechnology*, 46(6):567–574, August 2016.

[53] T. Mori, B. Binder, and C. H. Johnson. Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proceedings of the National Academy of Sciences of the United States of America*, 93(19):10183–10188, September 1996.

[54] Flix Moronta-Barrios, Javier Espinosa, and Asuncin Contreras. In vivo features of signal transduction by the essential response regulator RpaB from Synechococcus elongatus PCC 7942. *Microbiology (Reading, England)*, 158(Pt 5):1229–1237, May 2012.

[55] Tulika Munshi, Antima Gupta, Dimitrios Evangelopoulos, Juan David Guzman, Simon Gibbons, Nicholas H. Keep, and Sanjib Bhakta. Characterisation of ATP-dependent Mur ligases involved in the biogenesis of cell wall peptidoglycan in Mycobacterium tuberculosis. *PloS One*, 8(3):e60143, 2013.

[56] Masayuki Muramatsu and Yukako Hihara. Acclimation to high-light conditions in cyanobacteria: from gene expression to physiological responses. *Journal of Plant Research*, 125(1):11–39, January 2012.

[57] Yoriko Murayama, Tokitaka Oyama, and Takao Kondo. Regulation of circadian clock gene expression by phosphorylation states of KaiC in cyanobacteria. *Journal of Bacteriology*, 190(5):1691–1698, March 2008.

[58] Masato Nakajima, Keiko Imai, Hiroshi Ito, Taeko Nishiwaki, Yoriko Murayama, Hideo Iwasaki, Tokitaka Oyama, and Takao Kondo. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science (New York, N.Y.)*, 308(5720):414–415, April 2005.

[59] Norihito Nakamichi, Takatoshi Kiba, Mari Kamioka, Takamasa Suzuki, Takafumi Yamashino, Tetsuya Higashiyama, Hitoshi Sakakibara, and Takeshi Mizuno. Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42):17123–17128, October 2012.

[60] Mitsuteru Nakao, Shinobu Okamoto, Mitsuyo Kohara, Tsunakazu Fujishiro, Takatomo Fujisawa, Shusei Sato, Satoshi Tabata, Takakazu Kaneko, and Yasukazu Nakamura. CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Research*, 38(suppl_1):D379–D381, January 2010.

[61] Taeko Nishiwaki, Yoshinori Satomi, Yohko Kitayama, Kazuki Terauchi, Reiko Kiyohara, Toshifumi Takao, and Takao Kondo. A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *The EMBO journal*, 26(17):4029–4037, September 2007.

[62] Y. Ouyang, C. R. Andersson, T. Kondo, S. S. Golden, and C. H. Johnson. Resonating circadian clocks enhance fitness in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8660–8664, July 1998.

[63] Mark L. Paddock, Joseph S. Boyd, Dawn M. Adin, and Susan S. Golden. Active output state of the Synechococcus Kai circadian oscillator. *Proceedings of the National Academy of Sciences of the United States of America*, 110(40):E3849–E3857, October 2013.

[64] Kalyanee Paithoonrangsarid, Maria A. Shoumskaya, Yu Kanesaki, Syusei Satoh, Satoshi Tabata, Dmitry A. Los, Vladislav V. Zinchenko, Hidenori Hayashi, Morakot Tanticharoen, Iwane Suzuki, and Norio Murata. Five histidine kinases perceive osmotic stress and regulate distinct sets of genes in Synechocystis. *The Journal of Biological Chemistry*, 279(51):53078–53086, December 2004.

[65] Gopal K. Pattanayak, Connie Phong, and Michael J. Rust. Rhythms in Energy Storage Control the Ability of the Cyanobacterial Circadian Clock to Reset. *Current Biology*, 24(16):1934–1938, August 2014.

[66] Grant Petty and George Weidner. Ground-based Atmospheric Monitoring Instrument Suite, Rooftop Instrument Group: solar radiation data. Technical report, University of Wisconsin-Madison, Space Science and Engineering Center and Department of Atmospheric and Oceanic Sciences., Madison, WI., 2017.

[67] Anna M. Puszynska and Erin K. O'Shea. Switching of metabolic programs in response to light availability is an essential function of the cyanobacterial circadian output pathway. *eLife*, 6:e23210, April 2017.

[68] Ximing Qin, Mark Byrne, Yao Xu, Tetsuya Mori, and Carl Hirschie Johnson. Coupling of a core post-translational pacemaker to a slave transcription/translation feedback loop in a circadian system. *PLoS biology*, 8(6):e1000394, June 2010.

[69] Guillaume Rey, Franois Cesbron, Jacques Rougemont, Hans Reinke, Michael Brunner, and Felix Naef. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS biology*, 9(2):e1000595, February 2011.

[70] Joel Rozowsky, Ghia Euskirchen, Raymond K. Auerbach, Zhengdong D. Zhang, Theodore Gibson, Robert Bjornson, Nicholas Carriero, Michael Snyder, and Mark B. Gerstein. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotechnology*, 27(1):66–75, January 2009.

[71] Michael J. Rust, Joseph S. Markson, William S. Lane, Daniel S. Fisher, and Erin K. O'Shea. Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science (New York, N.Y.)*, 318(5851):809–812, November 2007.

[72] Yurie Seino, Tomoko Takahashi, and Yukako Hihara. The Response Regulator RpaB Binds to the Upstream Element of Photosystem I Genes To Work for Positive Regulation under Low-Light Conditions in Synechocystis sp. Strain PCC 6803. *Journal of Bacteriology*, 191(5):1581–1586, March 2009.

[73] Asako Seki, Mitsumasa Hanaoka, Yuki Akimoto, Susumu Masuda, Hideo Iwasaki, and Kan Tanaka. Induction of a Group 2 Factor, RPOD3, by High Light and the Underlying Mechanism in Synechococcus elongatus PCC 7942. *Journal of Biological Chemistry*, 282(51):36887–36894, December 2007.

[74] Yohei Shimura, Yoshihiro Shiraiwa, and Iwane Suzuki. Characterization of the subdomains in the N-terminal region of histidine kinase Hik33 in the cyanobacterium Synechocystis sp. PCC 6803. *Plant & Cell Physiology*, 53(7):1255–1266, July 2012.

[75] Maria A. Shoumskaya, Kalyanee Paithoonrangsarid, Yu Kanesaki, Dmitry A. Los, Vladislav V. Zinchenko, Morakot Tanticharoen, Iwane Suzuki, and Norio Murata. Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in synechocystis. *The Journal of Biological Chemistry*, 280(22):21531–21538, June 2005.

[76] Roshi Shrestha, Jorge Gmez-Ariza, Vittoria Brambilla, and Fabio Fornara. Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals. *Annals of Botany*, 114(7):1445–1458, November 2014.

[77] Ryan K. Shultzaberger, Joseph S. Boyd, Spencer Diamond, Ralph J. Greenspan, and Susan S. Golden. Giving Time Purpose: The Synechococcus elongatus Clock in a Broader Network Context. *Annual Review of Genetics*, 49(1):485–505, 2015.

[78] R. A. Singer and W. F. Doolittle. Control of gene expression in blue-green algae. *Nature*, 253(5493):650–651, February 1975.

[79] Kristina M. Smith, Gencer Sancar, Rigzin Dekhang, Christopher M. Sullivan, Shaojie Li, Andrew G. Tag, Cigdem Sancar, Erin L. Bredeweg, Henry D. Priest, Ryan F. McCormick, Terry L. Thomas, James C. Carrington, Jason E. Stajich, Deborah Bell-Pedersen, Michael Brunner, and Michael Freitag. Transcription factors in light and circadian clock signaling networks revealed by genomewide mapping of direct targets for neurospora white collar complex. *Eukaryotic Cell*, 9(10):1549–1556, October 2010.

[80] Rachelle M. Smith and Stanly B. Williams. Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium Synechococcus elongatus. *Proceedings of the National Academy of Sciences of the United States of America*, 103(22):8564–8569, May 2006.

[81] Tomoko Takahashi, Nanako Nakai, Masayuki Muramatsu, and Yukako Hihara. Role of Multiple HLR1 Sequences in the Regulation of the Dual Promoters of the psaAB Genes in Synechocystis sp. PCC 6803. *Journal of Bacteriology*, 192(15):4031–4036, August 2010.

[82] Naoki Takai, Masato Nakajima, Tokitaka Oyama, Ryotaku Kito, Chieko Sugita, Mamoru Sugita, Takao Kondo, and Hideo Iwasaki. A KaiC-associating SasARpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proceedings of the National Academy of Sciences*, 103(32):12109–12114, August 2006.

[83] Sotaro Takano, Jun Tomita, Kintake Sonoike, and Hideo Iwasaki. The initiation of nocturnal dormancy in Synechococcus as an active process. *BMC Biology*, 13:36, 2015.

[84] Yasuhito Taniguchi, Mitsunori Katayama, Rie Ito, Naoki Takai, Takao Kondo, and Tokitaka Oyama. labA: a novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC. *Genes & Development*, 21(1):60–70, January 2007.

[85] Yasuhito Taniguchi, Naoki Takai, Mitsunori Katayama, Takao Kondo, and Tokitaka Oyama. Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian kaiBC expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7):3263–3268, February 2010.

[86] Shu-Wen Teng, Shankar Mukherji, Jeffrey R. Moffitt, Sophie de Buyl, and Erin K. O'Shea. Robust circadian oscillations in growing cyanobacteria require transcriptional feedback. *Science (New York, N.Y.)*, 340(6133):737–740, May 2013.

[87] Chao-Jung Tu, Jeffrey Shrager, Robert L. Burnap, Bradley L. Postier, and Arthur R. Grossman. Consequences of a Deletion in dspA on Transcript Accumulation in Synechocystis sp. Strain PCC6803. *Journal of Bacteriology*, 186(12):3889–3902, June 2004.

[88] Vikram Vijayan, Isha H. Jain, and Erin K. O'Shea. A high resolution map of a cyanobacterial transcriptome. *Genome Biology*, 12(5):R47, 2011.

[89] Vikram Vijayan and Erin K. O'Shea. Sequence determinants of circadian gene expression phase in cyanobacteria. *Journal of Bacteriology*, 195(4):665–671, February 2013.

[90] Vikram Vijayan, Rick Zuzow, and Erin K. O'Shea. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences*, 106(52):22564–22568, December 2009.

[91] Lorraine G. van Waasbergen, Nadia Dolganov, and Arthur R. Grossman. nblS, a Gene Involved in Controlling Photosynthesis-Related Gene Expression during High Light and Nutrient Stress inSynechococcus elongatus PCC 7942. *Journal of Bacteriology*, 184(9):2481–2490, May 2002.

[92] Jacob R. Waldbauer, Sbastien Rodrigue, Maureen L. Coleman, and Sallie W. Chisholm. Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle. *PLoS ONE*, 7(8), August 2012.

[93] Stefanie Westermark and Ralf Steuer. Toward Multiscale Models of Cyanobacterial Growth: A Modular Approach. *Frontiers in Bioengineering and Biotechnology*, 4, December 2016.

[94] Mark A. Woelfle, Yao Xu, Ximing Qin, and Carl Hirschie Johnson. Circadian rhythms of superhelical status of DNA in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47):18819–18824, November 2007.

[95] Qiong Yang, Bernardo F. Pando, Guogang Dong, Susan S. Golden, and Alexander van Oudenaarden. Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science (New York, N.Y.)*, 327(5972):1522–1526, March 2010.

[96] Erik R. Zinser, Debbie Lindell, Zackary I. Johnson, Matthias E. Futschik, Claudia

Steglich, Maureen L. Coleman, Matthew A. Wright, Trent Rector, Robert Steen, Nathan McNulty, Luke R. Thompson, and Sallie W. Chisholm. Choreography of the Transcriptome, Photophysiology, and Cell Cycle of a Minimal Photoautotroph, Prochlorococcus. *PLoS ONE*, 4(4), April 2009.

[97] David Zwicker, David K. Lubensky, and Pieter Rein ten Wolde. Robust circadian clocks from coupled protein-modification and transcription-translation cycles. *Proceedings of the National Academy of Sciences*, 107(52):22540–22545, December 2010.