Supplemental Text

Supplemental Figure Legends

Figure S1. *rpaA* mutant gene expression timecourse (related to Figure 1)

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 (Vijayan et al., 2009). The timecourses of those same genes in the *rpaA* mutant strain are shown in green, also relative to the wild-type time average (see Experimental Procedures). 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 is Figure S1A (above) 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.

Figure S2. Rescue experiments (related to Figure 2)

A. Western blot of KaiC phosphorylation for the control and *ΔrpaA* clock rescue timecourses shown in Figure 2. Quantification of these blots is shown below the heatmaps in Figure 2. 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 Extended Experimental Procedures for details.

B. Microarray-based gene expression timecourses in the rescue experiments (Figure 2) for the subjective dusk (class 1) genes with the highest amplitude in the wild-type strain (same set as in Figure S1A). 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 S1 (see Experimental Procedures), except that the dye-swap experiment was not performed.

C. Microarray-based gene expression timecourses in the rescue experiments (Figure 2) for the subjective dawn (class 2) genes with the highest amplitude in the wild-type strain (same set as in Figure S1B). This plot is constructed in the same manner as is Figure S2A (above) but using the highest-amplitude subjective dawn genes in the wild-type strain.

Figure S3. RpaA binds to the *kaiBC* promoter in vivo and in vitro and promotes *kaiBC* expression in a phosphorylation-dependent manner (related to Figure 3)

A. Analysis of affinity-purified anti-RpaA antibody specificity by Western blot. Samples from wild-type and *rpaA* mutant circadian timecourses were lysed in urea lysis buffer. Equal masses of total protein from each lysate were separated on a 4-20% gradient SDS-PAGE gel. Proteins were transferred from the gel to nitrocellulose, which was then blocked in milk and

probed with the same affinity-purified anti-RpaA antibody used for ChIP-qPCR and ChIP-Seq experiments (0.17 μg/ml in TBST buffer containing 2.5% milk). Locations of Kaleidoscope Prestained molecular weight markers (Bio-Rad) are shown to the left of the gel.

B. Western blot analysis of the phosphorylation state of RpaA for Figure 3A. 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 (Gutu and O'Shea, 2013).

C. Correlation between HA-RpaA phosphorylation and HA-RpaA enrichment at the *kaiBC* promoter. RpaA phosphorylation was measured by Phos-tag Western blot (Gutu and O'Shea, 2013), while association with P*kaiBC* was measured by ChIP-qPCR performed with anti-HA antibody (see Extended Experimental Procedures). 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 Extended Experimental Procedures. The identity of the band corresponding to phosphorylated RpaA (RpaA~P) was revealed by boiling a reaction that was expected to contain RpaA~P for several minutes at 95 °C to hydrolyze

phosphoryl groups from RpaA (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 (Kutsuna et al., 2005). *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.

F. Western blot analysis of the abundance of ectopically-expressed RpaA in the strains employed in Figure 3C. Strains were induced with indicated concentrations of IPTG, grown for 24 h, entrained with two 12-h dark pulses, and then released to constant light. Samples were collected after 36 h in constant light. Samples were lysed in denaturing buffer, run on a standard 4-20% gradient SDS-PAGE gel, and analyzed by Western blotting. Under this gel condition, RpaA and RpaA~P co-migrate.

G. Western blot analysis of the phosphorylation state of RpaA in the complementation strains employed in Figure 3C using Phos-tag Western blotting. Samples are the same as those described in Figure S3F. Where indicated ("heat"), samples were heated at 95 °C for four

minutes. 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.

Figure S4. Identification of RpaA binding sites by ChIP-Seq (related to Figure 4)

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. Peak RpaA binding with the HA-RpaA strain occurred at subjective dusk (36 h), within the range observed for wild-type RpaA (32 - 40 h, Figures 3A and 4).

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 4; see Extended Experimental Procedures), 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 (anti-RpaA antibody with the *rpaA* mutant strain or anti-HA antibody with the wild-type strain). The anti-RpaA ChIP-Seq experiment was carried out on a sample from the 32 h timepoint (the time of maximum RpaA binding in the wild-type strain) while the anti-HA ChIP-Seq experiments was conducted on a samples from the 36 h timepoint (the time of maximum HA-RpaA binding in the HA-RpaA strain). 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 Extended Experimental Procedures. RpaA phosphorylation was measured by Phos-tag Western blotting (Gutu and O'Shea, 2013).

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, Table S3A and Figure 4A) and those identified in the HA-RpaA strain using the anti-HA antibody (2-fold enrichment cutoff, Table S3B and Figure S4A). A lower threshold was used for calling binding sites the HA-RpaA strain because of the weaker RpaA binding observed in that strain (Figure S4B); overlap with the wild-type strain increases for lower thresholds but with a concomitant increase in the number of non-overlapping false positives. The highest (least significant) *Q*-value for a peak obtained with the anti-RpaA antibody was 2.4×10^{-79} (Table S3A) and for the anti-HA antibody was 2.5×10^{-63} (Table S3B).

E. 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 (Vijayan and O'Shea, 2013). Labeled as in Figure S3E with the phase of each promoter (subjective dawn or subjective dusk) indicated.

Figure S5. RpaA orchestrates global circadian gene expression (Related to Figure 6)

A. RpaA protein abundance in samples used for RNA sequencing for Figure 6 and related experiments. We prepared whole protein lysates from samples acquired contemporaneously with those used for RNA-seq (Figure 6A-6E and S5B-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 *ΔrpaA ΔkaiBC* strain in which the P*trc* construct was inserted into the genome, but without *rpaA* inserted downstream of P*trc* (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 6D). Normalized expression is shown on the *y*-axis as described in the Extended Experimental Procedures. 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 6E. Cells were imaged using their intrinsic red autofluorescence. Scale bar = 10 μm.

F. Western blot analysis of the phosphorylation state of RpaA in the strains employed in Figure 6E. Indicated strains were grown for four days in constant light (74 μE m^{-2} s^{-1}) with the 100 μM IPTG. Western blots were performed with two different secondary antibody concentrations in order to capture the full range of RpaA abundances present. The top row shows blots probed with the higher secondary antibody concentration, while the bottom row

shows a blot probed with the lower secondary antibody concentration; samples in the first column were analyzed with the higher secondary antibody concentration only. 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. When the same samples are analyzed with a lower secondary antibody concentration (bottom right image), the IPTG-dependent increase in unphosphorylated RpaA levels becomes more apparent. Also, the band of unknown identity (*) is not visible in this image, demonstrating that it is present at very low levels relative to unphosphorylated RpaA.

Supplemental Table Legends

Table S1. Characteristics of genes in wild-type and *rpaA* mutant gene expression timecourses and in RpaA ChIP-seq analysis (Related to Figures 1, 5, 6, and S1)

Columns A-J, General gene characteristics. Column A, *synpcc7942* gene ID; **column B**, JGI (Joint Genome Institute) ID; **column C**, Gene name; **column D**, Gene description from Cyanobase for chromosome and plasmid pANL and from JGI for plasmid pANS; **column E**, Gene function. Functional categories were taken from Cyanobase COGs. The category "Not in COGs" was changed to "Unclassified" except for those RpaA targets whose function could be discerned from database searches. Such targets were manually assigned to the appropriate COG category.; **column F**, Chromosome or plasmid on which the gene resides; **column G**, Gene start location; **column H**, Gene end location; **column I**, Strand; **column J**, RNA-seq-derived transcript identification number (Table S5). Genes on the plasmids do not have a transcript number because the plasmids were not included in the identification of transcripts (Vijayan et al., 2011). Genes on the chromosome without a transcript number were not detectably expressed (Vijayan et al., 2011).

Columns K-S, Circadian characteristics in the wild-type strain (Related to Figures 1 and S1). Column K: Is the gene reproducibly circadian (1 – yes, 0 – no)? A gene was considered reproducibly circadian if it (a) had a Cosiner period (Kucho et al., 2005) of 22 - 26 h in the 60-h timecourse described in (Vijayan et al., 2009), (b) had a Cosiner period of 20 – 28 h in an independent 48-h timecourse performed by Vikram Vijayan (V. Vijayan, personal communication), and (c) had the same phase-based class (1 or 2) in both timecourses.; **column L**, Period in the 60-h timecourse as determined by Cosiner analysis (best fit period between 12

and 36 h) (Vijayan et al., 2009); **column M**, Cosiner amplitude at the period listed in column L in the 60-h timecourse (Vijayan et al., 2009); **column N**, Cosiner phase at the period listed in column L in the 60-h timecourse (Vijayan et al., 2009); **column O**, Is the gene circadian in the 60-h timecourse (1 – yes, 0 – no)? A gene was considered circadian if its period was between 22 and 26 h, inclusive (Vijayan et al., 2009).; **column P**, Class of gene expression based on phase in 60-h timecourse (Vijayan et al., 2009). If the gene was not circadianly-expressed (column O), it was assigned a class of 0. Class 1 genes are subjective dusk genes; class 2 genes are subjective dawn genes; **column Q**, Period in the 48-h timecourse as determined by Cosiner analysis (best fit period between 12 and 36 h) (V Vijayan, personal communication); **column R**, Is the gene circadian in the 48-h timecourse (1 – yes, 0 – no)? A gene was considered circadian if its period was between 20 and 28 h, inclusive. This period window is wider than that for the 60-h timecourse (column O) because this 48-h timecourse was shorter and therefore the period was less precisely determined; **column S**, Class of gene expression based on phase in 48-h timecourse. If the gene was not circadianly-expressed (column R), it was assigned a class of 0.

Columns T-W, Gene properties in the *rpaA* mutant timecourse (related to Figures 1, 5, and S1). Column T, Log_2 of the expression change in the *rpaA* mutant relative to the wild-type time average (see Experimental Procedures).; **column U**, Period of expression as determined by Cosiner analysis (best fit period between 12 and 36 h) (Kucho et al., 2005). Note that a period is assigned even if a gene is arrhythmic.; **column V**, Cosiner amplitude; **column W**, Cosiner phase.

Columns X-Y, **ChIP targets of RpaA (related to Figure 5 and 6). Column X**, Is the gene a ChIP target of RpaA (1 – yes, 0 – no; see Extended Experimental Procedures)?; **column Y**,

Identification number of peak (Table S3A) targeting this gene. If the gene is not a target, this field is left blank.

Table S2. Properties of genes in the clock rescue experiments (Related to Figures 2 and S2)

Columns A-H, General gene characteristics. Same as in Table S1.

Columns I-Q, Gene properties in the control clock rescue in each of two biological replicates. Data shown in Figures 2 and S2 are from Replicate 1. **Column I,** Is the gene reproducibly circadian in the control rescue (1 – yes, 0 – no)? A gene was considered reproducibly circadian if it was circadian in each of two 48-hour replicates (columns M and Q).; **column J**, Period of expression as determined by Cosiner analysis (best fit period between 12 and 36 h) (Kucho et al., 2005) for replicate 1. Note that a period is assigned even if a gene is arrhythmic.; **column K**, Cosiner amplitude for replicate 1; **column L**, Cosiner phase for replicate 1; **column M**, Is the gene circadian in replicate 1? A gene was considered circadian if it had a Cosiner period within 2 hours of the period of KaiC phosphorylation, which in this replicate was 23.0 h.; **columns N-Q**, Same as columns J-M but for Replicate 2, for which KaiC phosphorylation had a period of 22.8 h. Data from the 56 h timepoint of replicate 2 were not included in the Cosiner analysis because of poor data quality as assessed from the ratios of spike-in controls (Vijayan et al., 2009).

Columns R-Z, Gene properties in the *ΔrpaA* clock rescue strain in each of two biological replicates. Data shown in Figures 2 and S2 are from Replicate 1. **Column R,** Is the gene

reproducibly circadian in the *ΔrpaA* clock rescue (1 – yes, 0 – no)? A gene was considered reproducibly circadian if it was circadian in each of two 48-hour replicates (columns V and Z).; **columns S-V**, Same as columns J-M but for replicate 1 of the *ΔrpaA* clock rescue, for which KaiC phosphorylation had a period of 23.2 h; **columns W-Z**, Same as for columns J-M but for replicate 2 of the *ΔrpaA* clock rescue, for which KaiC phosphorylation had a period of 22.3 h. Data from the 56 h timepoint of replicate 1 and the 48 h timepoint of replicate 2 were not included in the Cosiner analysis because of poor data quality as assessed from the ratios of spike-in controls (Vijayan et al., 2009).

Table S3. RpaA binding sites (peaks) (related to Figures 4 and S4).

A. RpaA binding sites identified by anti-RpaA ChIP-Seq using a 3-fold enrichment cutoff. **Columns A-F, Characteristics of peaks. Column A**, Peak identification number; **column B**, Chromosome or plasmid on which the peak is located; **column C**, Left-hand boundary of the peak; **column D**, Right-hand boundary of the peak; **column E**, Location of the peak maximum in the ChIP-Seq signal itself, which is not necessarily identical to the location of the maximum in the enrichment signal (ChIP divided by the mock); **column F**, Multiple-hypothesis-corrected *Q*-value (Rozowsky et al., 2009) for the peak; **columns G**, Is the peak also found in the anti-HA ChIP-Seq data (Figures S4A, S4B, S4D, and Table S3B) (1 – yes, 0 – no)? A peak in the anti-HA ChIP-Seq analysis was considered to be the same as a peak in the anti-RpaA ChIP-Seq analysis if the distance between the peak maxima (column E in Tables S3A and S3B) was ≤ 100 bp.; **column H**, Number of instances of the RpaA binding motif (Figure 4D and Table S4A) identified by FIMO (Grant et al., 2011) within 125 bp of the maximum of the peak (column E)

with $p \leq 0.001$ (Table S4B). The + and – strands were searched separately and hits sometimes overlap on the same strand. **Columns I-P, Enrichments in the wild-type timecourse (Figures 4B and 4C) and in the anti-HA ChIP-Seq (Figures S4A, S4B, S4C, and S4D). Columns I-O**, Enrichments in the wild-type timecourse (Figures 4B and 4C). For each timepoint, the maximum enrichment within 50 bp of the peak maximum (column E) was determined. Enrichment was calculated by dividing the PeakSeq-normalized ChIP-Seq signal for the given timepoint in the wild-type strain by the anti-RpaA ChIP-Seq signal for the pooled *ΔrpaA* sample (pool of samples acquired every 4 hours for 24 hours).; **column P**, Maximum enrichment in the anti-HA ChIP-Seq data (36 h timepoint) within 50 bp of the anti-RpaA peak maximum (column E). Enrichment was calculated by dividing the PeakSeq-normalized anti-HA ChIP-Seq signal for the HA-tagged RpaA strain by the anti-HA ChIP-Seq signal for the wild-type strain.

B. HA-RpaA binding sites (peaks) identified by ChIP-Seq with the anti-HA antibody using a 2-fold enrichment cutoff. **Columns A-F, Characteristics of peaks.** Same as in Table S3A columns A – F but for peaks identified by ChIP-Seq with the anti-HA antibody using a 2-fold enrichment cutoff. **Column G**, Maximum enrichment in the anti-HA ChIP-Seq experiment (36 h timepoint) within a 50 bp window around the location of the peak maximum (column E). Enrichment was calculated by dividing the PeakSeq-normalized anti-HA ChIP-Seq signal for the HA-tagged RpaA strain by the anti-HA ChIP-Seq signal for the wild-type.

Table S4 –RpaA binding motif (Related to Figure 4 and Table S3A)

A. Position-specific probability matrix (PSPM) for the motif identified by MEME (Bailey and Elkan, 1994) and shown in Figure 4D.

B. RpaA binding motifs (Figure 4D) identified by FIMO (Grant et al., 2011) in RpaA binding sites/peaks (Figure 4C and Table S3A) within 125 bp of the peak maxima. **Column A**, RpaA binding site number in Table S3A.; **column B**, Strand on which the motif is located.; **column C**, Location of left-hand side of motif, where the location of the peak maximum (Table S3A, column E) is base pair 126; **column D**, Location of right-hand side of motif, where the location of the peak maximum (Table S3A, column E) is base pair 126; **column E**, Match score; **column F**, *p*-value; **column G**, *q*-value; **column H**, Sequence matching the motif.

Table S5. Transcripts identified by RNA-Seq (Vijayan et al., 2011) that are targeted by RpaA based on anti-RpaA ChIP-Seq data (related to Figure 5, Table S1, and Table S3A)

Columns A-K, Characteristics of transcripts (Vijayan et al., 2011). Only "annotated transcripts" and "high-confidence ncRNAs" from Vijayan et al., 2011, were included in this analysis. C**olumn A**, Transcript identification number; **column B**, Chromosome or plasmid on which transcript resides. Note that Vijayan et al., 2011, searched for transcripts on the chromosome only.; **column C**, Strand encoding the transcript.; **column D**, Location of 5' end of the transcript.; **column E**, Location of 3' end of the transcript.; **column F**, Is the transcript an "annotated transcript" (Vijayan et al., 2011) (1 – yes, 0 – no)? A transcript is considered "annotated" if it encodes mRNA, tRNA, and/or rRNA (Vijayan et al., 2011). Mutually exclusive with column G.; **column G**, Is the transcript a high-confidence non-coding RNA (Vijayan et al.,

2011) (1 – yes, 0 – no)? Mutually exclusive with column F.; **column H**, JGI number(s) of mRNA, tRNA, and/or rRNA fully encompassed within the transcript.; **column I**, JGI number (Table S1) of the mRNA, tRNA, and/or rRNA located closest to the 5' end of the transcript (from column H).; **column J**, JGI number of the mRNA, tRNA, and/or rRNA located closest to the 3' end of the transcript (from column H).

Columns K-M, ChIP targets of RpaA (related to Figure 5). Column K, Is the transcript a ChIP target of RpaA (1 – yes, 0 – no; see Extended Experimental Procedures)?; **column L**, Identification number of ChIP peak (Table S3A) targeting this gene. If the transcript is not a target, this field is left blank.; **column M**, Distance (bp) between peak maximum (Table S3A, column E) and 5' end of the transcript. If the transcript is not a target, this field is left blank.

Table S6. Analysis of the effect of RpaA(D53E) induction on global gene expression (Related to Figure 6)

Column A, *synpcc7942* gene ID; **column B**, JGI ID; **column C**, Change in expression of the gene from subjective dawn (24 h) to subjective dusk (36 h) in entrained wild-type cells as measured by RNA-seq, expressed as the $\log_2$ of the ratio of expression at subjective dusk divided by expression at subjective dawn. If a gene did not have any reads assigned to it at subjective dusk and/or subjective dawn, we report "NaN" here.; **column D**, Change in expression of the gene in OX-D53E after 12 hours of incubation with IPTG as measured by RNA-seq, expressed as the $\log_2$ of the ratio of expression after 12 hours of incubation with IPTG divided by expression just prior to addition of IPTG (time 0 h). If a gene did not have any reads assigned to

it at either time point, we report "NaN" here.; **column E**, Name of the wild-type cluster to which the gene was assigned using *K*-means clustering based on the expression timecourse of the gene in entrained wild-type cells. Only high-confidence circadian genes were clustered. Genes not considered in the clustering are denoted with "N/A."; **column F**, Name of the cluster to which the gene was assigned using *K*-means clustering based on its expression timecourse in OX-D53E after addition of IPTG. Only high-confidence circadian genes were clustered. Genes not considered in the clustering are denoted with "N/A."

Table S7. Strain list and sequencing statistics (Related to Experimental Procedures and Extended Experimental Procedures)

A. Strain list. 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).

B. ChIP-Seq sequencing statistics. Insert size (Column C) represents the most frequently-occurring input DNA length, determined from Bioanalyzer electrophoresis traces and/or the cross-correlation of the start positions of reads mapping to the forward and reverse DNA strands.

C. RNA-Seq sequencing statistics.

Extended Experimental Procedures

Cyanobacterial strains

Strains were constructed using standard procedures for genomic integration by homologous recombination (Clerico et al., 2007) and are described in Table S7A.

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(Km^r) and pNS2KmP*trc-kaiBC* were gifts from Takao Kondo (Nagoya University).

Strain EOC66 was constructed by transforming AMC408 with pDRpaA(Km^r) (Takai et al., 2006). EOC72 was constructed by transforming AMC408 with pNS2KmP*trc-kaiBC* (Murayama et al., 2008) and a plasmid deleting the *kaiBC* locus between positions +1 and +1813 (relative to the translation start site of *kaiB*) with the chloramphenicol resistance cassette from pAM1573 (Andersson et al., 2000). EOC101 was constructed by transforming AMC408 with pNS2KmP*trc-kaiBC,* a plasmid deleting the *kaiBC* locus between positions +1 and +1813 (relative to the translation start site of *kaiB*) with a gentamycin cassette (SmaI-PvuII fragment from pAM2055 (Katayama et al., 2003)), and a plasmid deleting the same region of *rpaA* targeted by pDRpaA(Km^r) but with the chloramphenicol cassette from pAM1573.

The P*kaiBC* reporter strain EOC113 was created by transforming AMC395 with the *luxAB*-encoding plasmid pAM1580 containing the *kaiBC* promoter (Chabot et al., 2007) inserted between the XbaI and SalI restriction sites. Strains EOC339, EOC341, EOC345, and EOC346 were constructed by transforming EOC113 with the chloramphenicol-marked *rpaA* deletion vector described above for EOC101 and with a gentamycin-marked neutral site 2.2 (NS 2.2) targeting vector carrying the P*trc* promoter driving expression of the appropriate gene. The

targeting vector was constructed by replacing the chloramphenicol cassette in EB2065 (Jain et al., 2012) with a gentamycin cassette from pAM2055. Specifically, primers CATAACTAGTGGATCTGGTAACCCCAGCGCGG and ATACACTAGTGGATCTAAGCTTGCTTCTTTGC were used to perform inverse PCR on EB2065, introducing an SpeI site into which was cloned the gentamycin cassette amplified from pAM2055 with primers CATAACTAGTGACGCACACCGTGGAAACG and ATACACTAGTGCGGCGTTGTGACAATTTAC.

To obtain the P*trc* constructs to place into this targeting vector, we first cloned the *rpaA* gene into the EcoRI and BamHI sites in vector pAM2991 (Mackey et al., 2008); for the empty-vector (mock / P*trc::*-) control, pAM2991 was used without modification. The D53A and D53E mutations then were introduced using the QuikChange II XL kit (Agilent) with primers ACCTGATCATGCTGGCTCTAATGCTGCCGCG and CGCGGCAGCATTAGAGCCAGCATGATCAGGT (D53A) or CCTGATCATGCTGGAGCTAATGCTGCCGCGG and CCGCGGCAGCATTAGCTCCAGCATGATCAGG (D53E). The *lacI*^{q}-P*trc*-*rpaA*/empty-*rrnB* region of these pAM2991 variants was amplified with primers P-GCGACATCTTCCTGCTCCA and P-GCGCGGCTTAACTCAAGC and cloned into the SmaI site of the NS2.2 (Gm^{r}) targeting vector.

Strains EOC370, EOC371, and EOC374 were constructed in a true wild-type background (*Synechococcus elongatus* PCC7942, ATCC catalog number 33912). The ATCC strain was transformed with pDRpaA(Km^{r}), the chloramphenicol-marked *kaiBC* deletion vector described above, and the NS2.2 (Gm^{r}) targeting vectors containing the appropriate P*trc* construct described above.

Strains EOC116 (*ΔsasA)* and EOC118 (*ΔcikA)* were described in Gutu and O'Shea, 2013.

The HA-RpaA strain EOC67 was constructed by knocking a KanR-3xHA-RpaA construct into the *rpaA* locus in AMC408 background, such that the endogenous *rpaA* gene was completely replaced with the 3xHA-*rpaA* construct. The kanamycin cassette was placed such that the region between the RpaA binding site (peak 6 in Table S3A) and the 3xHA tag was unperturbed. The sequence of the genome of EOC67 around the knocked-in KanR-3xHA-RpaA construct is:

synpcc7942_0096-KanR-3xHA(with linker)-*rpaA*

GGATTCTACAAACGAACTTGAGCTTTGCCGATGAAGAAAACCTCAAGGCAGCCCTA
GAAAAAGCCAAAGCCGAAGCCCAAGCCCTGCGTCTCAATTAATCTCCAACGCTTCG
ATCGCTCCTGATTACTTGAGTTGATCAAGGCTCAGATCAAGAGCGATCGCTACTTCA
ATACAGCTTACTTTTGGCCGCTAGTCGACCTGCATCCCTTAACTTACTTATTAAATAA
TTTATAGCTATTGAAAAGAGATAAGAATTGTTCAAAGCTAATATTGTTTAAATCGTC
AATTCCTGCATGTTTTAAGGAATTGTTAAATTGATTTTTTGTAAATATTTTCTTGTATT
CTTTGTTAACCCATTTCATAACGAAATAATTATACTTTTGTTTATCTTTGTGTGATATT
CTTGATTTTTTTCTACTTAATCTGATAAGTGAGCTATTCACTTTAGGTTTAGGATGAA
AACTAGAGGATCTCAATGAATATTGGTTGACACGGGCGTATAAGACATGTTATACTG
TTGAATAACAAGGACGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT
TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCG
GCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGT
CAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATG
AACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGC
GCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGA
AGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCAT
CATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGA
CCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTG
TCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTC
GCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGA
TGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTG
TGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATAT
TGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGC
CGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCG
GGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGAT
TTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGAC
GCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCGG
GGATCCCTCAACAGCAGTTCAATCCTAAGAAACTGCAGAAGCAGGGATTGAGTAAT
TTGATGGTTGACTGCAGTCCTTATCAGTTGCTTCTCAACAATTATCCAACCGCAATTA
CGGACAGAGCACTGGACTAGCACCACTGCTAAAACCGCAAAGCTCGGACAAATATA

AAGACAACTTAAAGATTTTCTACGCCCTGCGACCTCACCAGAACAGAGCCACTATCA
TTTTTAAGAAAACATAAATTGATCCGCTGTAAGCCGCTGCCAGCACAAAGCTTACCG
CGTTTTTGCTAAAGCCTAGTCCGCTGGTTTGTCCTCCCGGAATGTTACCCTATGTACC
CCTACGATGTGCCCGATTACGCCGGCTATCCCTATGACGTCCCGTATGCCGGTAGCT
ATCCATATGACGTGCCGGATTACGCTGGCAGCGGCAGCATGAAACCCCGCATCCTC
GTGATCGATGATGACTCAGCCATCTTGGAGCTGGTCGCCGTCAATCTGGAGATGTCT
GGCTATGACGTACGCAAAGCTGAGG

Cell culture

For the experiments described in Figures 1, 3A, 4, 5, S1, S3A, S4A, S4B, and S4D, cells were grown in a turbidostat as described previously (Vijayan et al., 2009). 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, 6, S2, S3B, S4C, and S5A-S5D, cultures were grown in tissue culture flasks illuminated with 100 µE m^{-2} s^{-1} (µmoles photons m^{-2} s^{-1}) of cool fluorescent light and bubbled continuously with 1% CO_2 in air, with the OD_{750} 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 and S2, and 6 and S5A-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 m^{-2} s^{-1}) concomitant with addition of IPTG to a final concentration of 6 µM (Figures 2 and S2) or 100 µM IPTG (Figure 6 and S5A-D). Culture conditions for cell length measurements are described in the Cell Length Measurements section below.

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 Vijayan et al., 2009, 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 to be reproducibly circadian if it had: (a) a Cosiner period (Kucho et al., 2005) of 22 – 26 h in the 60-h timecourse described in Vijayan et al., 2009; (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.

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 (Takai et al., 2006), 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 (*ΔrpaA*) lysate by Western blot (Figure S3A). Only a single band was observed for the wild-type lysate, while no reactivity was observed against the *rpaA* mutant lysate.

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 Gutu and O'Shea 2013, except that affinity-purified anti-RpaA antibody (0.17 µg/ml) was used in place of anti-RpaA serum.

Chromatin immunoprecipitation (ChIP)

For each ChIP reaction, approximately 18 $OD_{750\ nm}$ 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 ul 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 ul (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 ul 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 purification 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%.

qPCR for ChIP and gene expression

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

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)

RT-qPCR was performed as described previously (Vijayan and O'Shea, 2013), with the abundance of the *kaiBC* transcript normalized to that of *hslO* transcript, whose abundance is constant in time (Vijayan and O'Shea, 2013; Vijayan et al., 2009).

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.

ChIP-Seq library preparation and sequencing

Except for "replicate 2" samples in GSE50922, 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 ul 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 H_2O, 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 (Langmead et al., 2009), counting only those aligning uniquely to one location with up to three mismatches.

For “replicate 2” samples in GSE50922, libraries for sequencing were prepared using paired-end adapters and PCR primers from Illumina following the manufacturer’s protocol for ChIP-Seq sample preparation (part number 11257047 revision A) with modifications. All enzymes except for fast ligase were obtained from NEB instead of Illumina. First, 5 – 10 ng of ChIP or input DNA was blunted and phosphorylated with T4 DNA ligase, T4 DNA polymerase,

and T4 polynucleotide kinase. 3' dA overhangs were added using exo$^-$ Klenow fragment. Adapters (diluted 1:10 from the Illumina stock) were ligated for 15 min at room temperature using LigaFast DNA ligase (Promega). Adapter-ligated fragments were size-selected on 2% agarose/TAE gel containing 400 μg/ml ethidium bromide; a gel slice corresponding to the 200-300 bp region (based on 50 bp ladder, Promega) was excised. DNA was purified from the gel using the Qiagen gel extraction kit according to the manufacturer's protocol, except that the gel slice was dissolved at room temperature. Adapter-ligated fragments were then amplified by PCR using PE Primers 1.0 and 2.0 (Illumina). The resulting libraries was purified with the Qiagen MinElute kit and eluted in 15 ul of buffer EB. The concentration and fragment length distribution of the libraries was then determined using the DNA 1000 Bioanalyzer assay (Agilent). Only those reads in which the first 32 bases aligned perfectly and uniquely to the genome were counted.

The number of aligned reads and the most common insert size for each sample is listed in Table S7B.

RpaA binding motif

Enriched motifs were identified with MEME (Bailey and Elkan, 1994) 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* #1, Figures 4D and S4E). Only one statistically significant motif (Figure 4D, *E*-value 1.7 x 10^{-36}) was located; the position-

specific probability matrix (PSPM) for this motif is listed in Table S4A. We identified instances of this motif in the query sequences using FIMO (Grant et al., 2011) (Table S4B).

DNase I Footprinting

The P*kaiBC*, P*rpoD6, and PpurF* promoter sequences were amplified from *S. elongatus* genomic DNA with primers GACGGATCCTTTTACGAGGGCTCATACGC and GACGAATTCGTGAGATGTATCGACGGTCTATCC for P*kaiBC*, GACGGATCCATCTCTTGTTCGTCGCTGAG and GACGAATTCTCCCTCTTACATTTTCGACACA for P*rpoD6*, and GACGAATTCCATGCCCTTGCAGAGCTC and GACGGATCCGCGATCGAACGTCGTTTG for P*purF*. These primers added BamHI and EcoRI restriction sites (underlined) to the ends of the amplified promoters. Promoter sequences were ligated into BamHI/EcoRI-digested pUC18 to create pUC18-P*kaiBC*, pUC18-P*rpoD6*, and pUC18-P*purF*. To generate mutations in P*kaiBC*, we utilized the QuikChange II XL site-directed mutagenesis kit (Agilent) with primers TCTATCCCACGAGAAACCTGAAAAGGTTTAGGAGGTCTTAAGC and GCTTAAGACCTCCTAAACCTTTTCAGGTTTCTCGTGGGATAGA for m15, TCCCACGAGAAACCTGAAAAGGTAATCGAGGTCTTAAGCTC and GAGCTTAAGACCTCGATTACCTTTTCAGGTTTCTCGTGGGA for m16, and CTGAAAAGGTAAAGGAGGTCAAAAGCTCGGCTCAATTTCTCT and AGAGAAATTGAGCCGAGCTTTTGACCTCCTTTACCTTTTCAG for m20, using pUC18-P*kaiBC* as a template. We mutagenized P*purF* similarly using primers GGCTAATGCAAATTTAACAATGTCTTTACCCTAGCGGTTAGTCTTTAGC and

GCTAAAGACTAACCGCTAGGGTAAAGACATTGTTAAATTTGCATTAGCC for m1-5, GCTAATGCAAATTTAACAGTGACTTAACCCTAGCGGTTAGTCTTTA and TAAAGACTAACCGCTAGGGTTAAGTCACTGTTAAATTTGCATTAGC for m1-7, and AGCGGCTAATGCAAATTTAGCAATGACTTAACCCTAGCG and CGCTAGGGTTAAGTCATTGCTAAATTTGCATTAGCCGCT for m1-8, using pUC18-P*purF* as a template.

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-^{32}P 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 $MgCl_2$, 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 (Takai et al., 2006)), 1.5 μM CikA (purified as described previously (Gutu and O'Shea, 2013)), 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 $MgCl_2$, 5 mM $CaCl_2$, 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). Radiolabeled 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 $MnCl_2$ 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).

ChIP-seq data analysis

Data were analyzed using a custom-coded, modified form of the PeakSeq algorithm (Rozowsky et al., 2009) 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 ≥ 3-fold in one replicate and ≥ 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 (Vijayan et al., 2011)). 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 (Nakao et al., 2010) using manually-updated assignments.

RNA-seq

Ribosomal RNA was depleted from 6 μg of total RNA (purified as described in Vijayan et al., 2009) 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. The number of aligned reads is listed in Table S7C.

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 (Anders and Huber, 2010). 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 S5D, 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 6A and S5B, 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 (Table S1) 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.

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 (Vijayan and O'Shea, 2013). Replicate experiments produced qualitatively similar results.

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 $OD_{750\ nm}$ 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).

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Andersson, C.R., Tsinoremas, N.F., Shelton, J., Lebedeva, N.V., Yarrow, J., Min, H., and Golden, S.S. (2000). Application of bioluminescence to the study of circadian rhythms in cyanobacteria. Methods Enzymol *305*, 527-542.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB International Conference on Intelligent Systems for Molecular Biology *2*, 28-36.

Chabot, J.R., Pedraza, J.M., Luitel, P., and van Oudenaarden, A. (2007). Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. Nature *450*, 1249-1252.

Clerico, E.M., Ditty, J.L., and Golden, S.S. (2007). Specialized techniques for site-directed mutagenesis in cyanobacteria. Methods Mol Biol *362*, 155-171.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Jain, I.H., Vijayan, V., and O'Shea, E.K. (2012). Spatial ordering of chromosomes enhances the fidelity of chromosome partitioning in cyanobacteria. Proc Natl Acad Sci U S A *109*, 13638-13643.

Katayama, M., Kondo, T., Xiong, J., and Golden, S.S. (2003). ldpA encodes an iron-sulfur protein involved in light-dependent modulation of the circadian period in the cyanobacterium Synechococcus elongatus PCC 7942. J Bacteriol *185*, 1415-1422.

Kucho, K., Okamoto, K., Tsuchiya, Y., Nomura, S., Nango, M., Kanehisa, M., and Ishiura, M. (2005). Global analysis of circadian expression in the cyanobacterium Synechocystis sp. strain PCC 6803. J Bacteriol *187*, 2190-2199.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Mackey, S.R., Choi, J.S., Kitayama, Y., Iwasaki, H., Dong, G., and Golden, S.S. (2008). Proteins found in a CikA interaction assay link the circadian clock, metabolism, and cell division in Synechococcus elongatus. J Bacteriol *190*, 3738-3746.

Min, H., Liu, Y., Johnson, C.H., and Golden, S.S. (2004). Phase determination of circadian gene expression in Synechococcus elongatus PCC 7942. J Biol Rhythms *19*, 103-112.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T., and Nakamura, Y. (2010). CyanoBase: the cyanobacteria genome database update 2010. Nucleic Acids Res *38*, D379-381.

Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M.B. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol *27*, 66-75.

Supplemental Figure

Figure S1

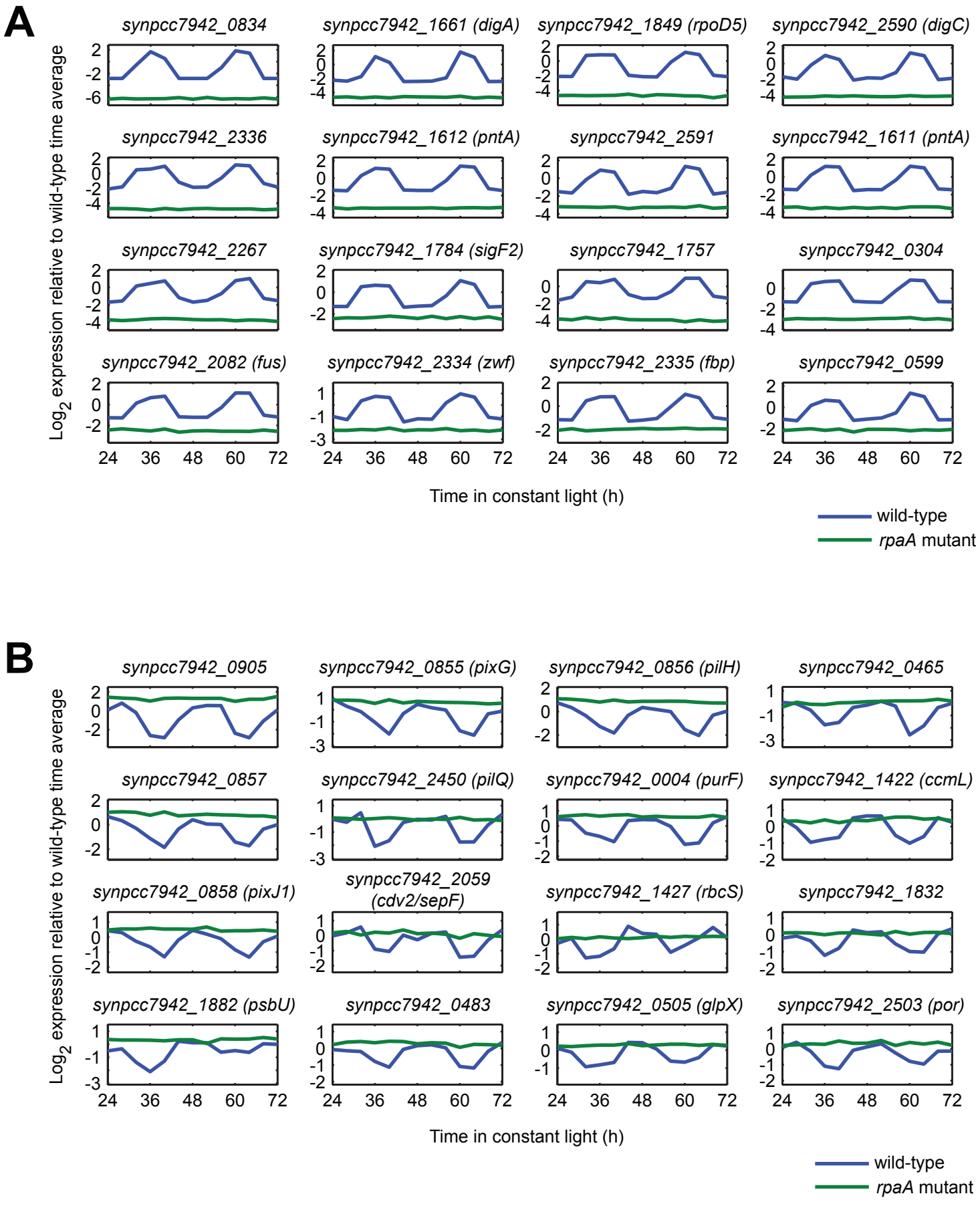


Figure S2

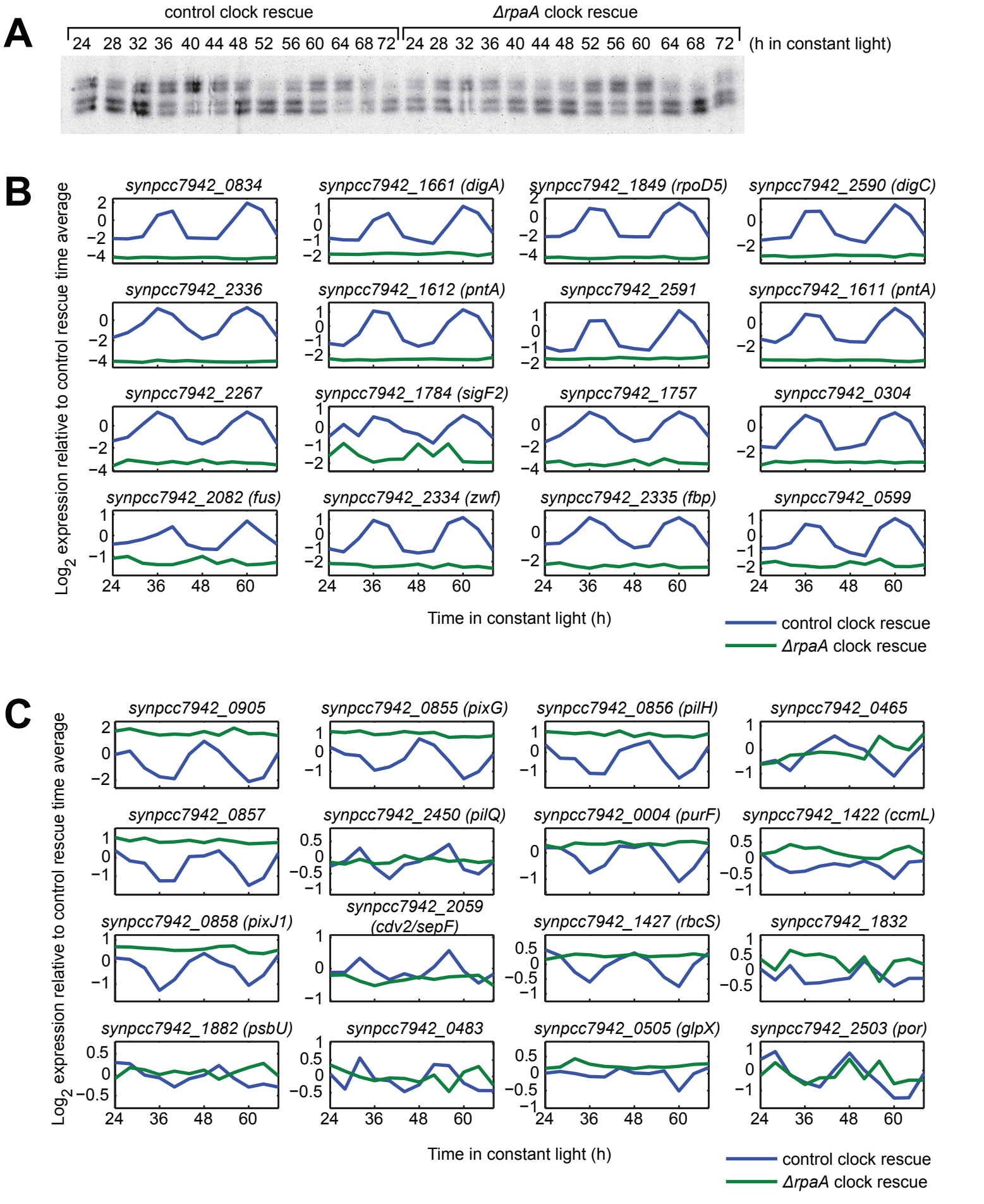
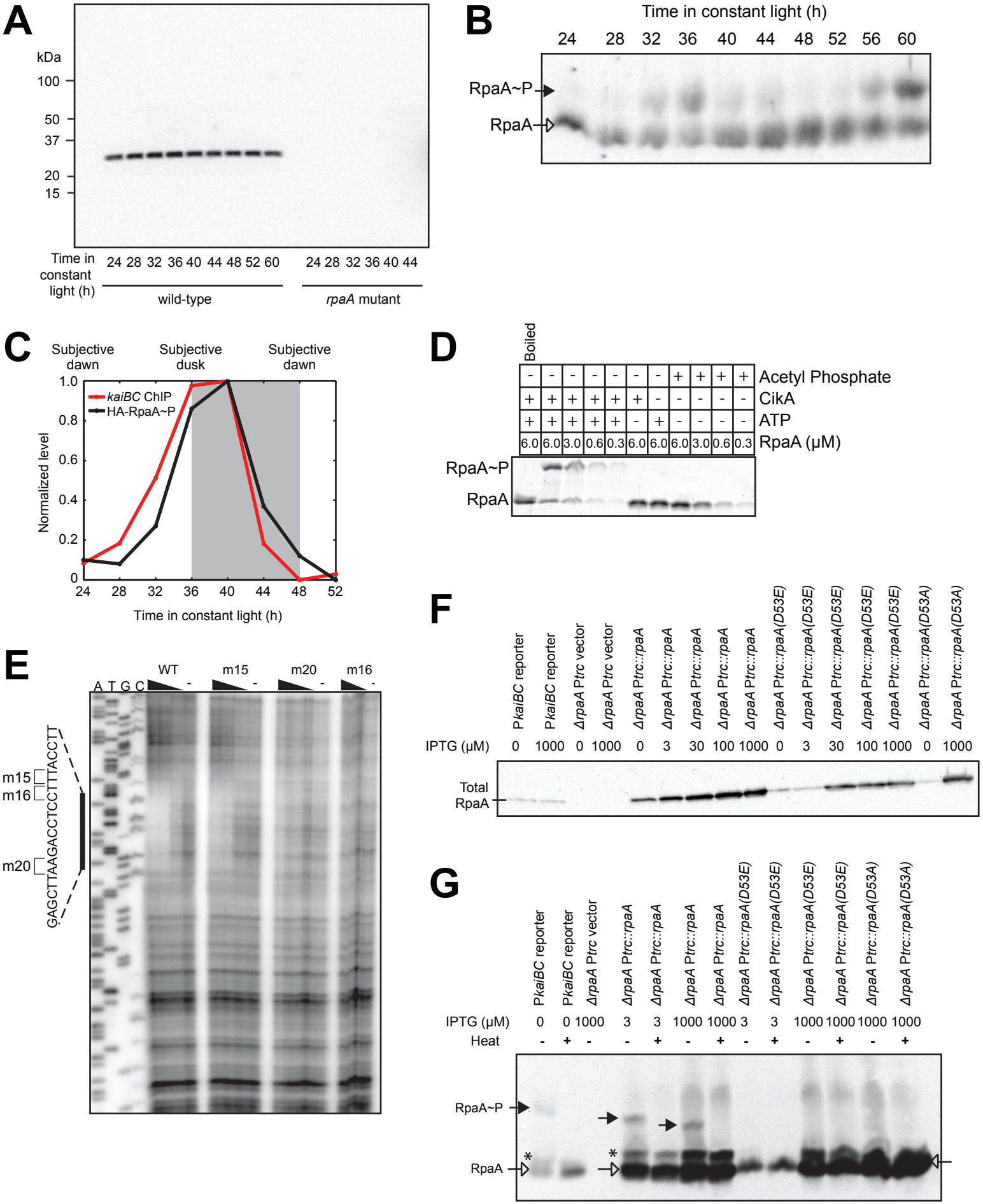
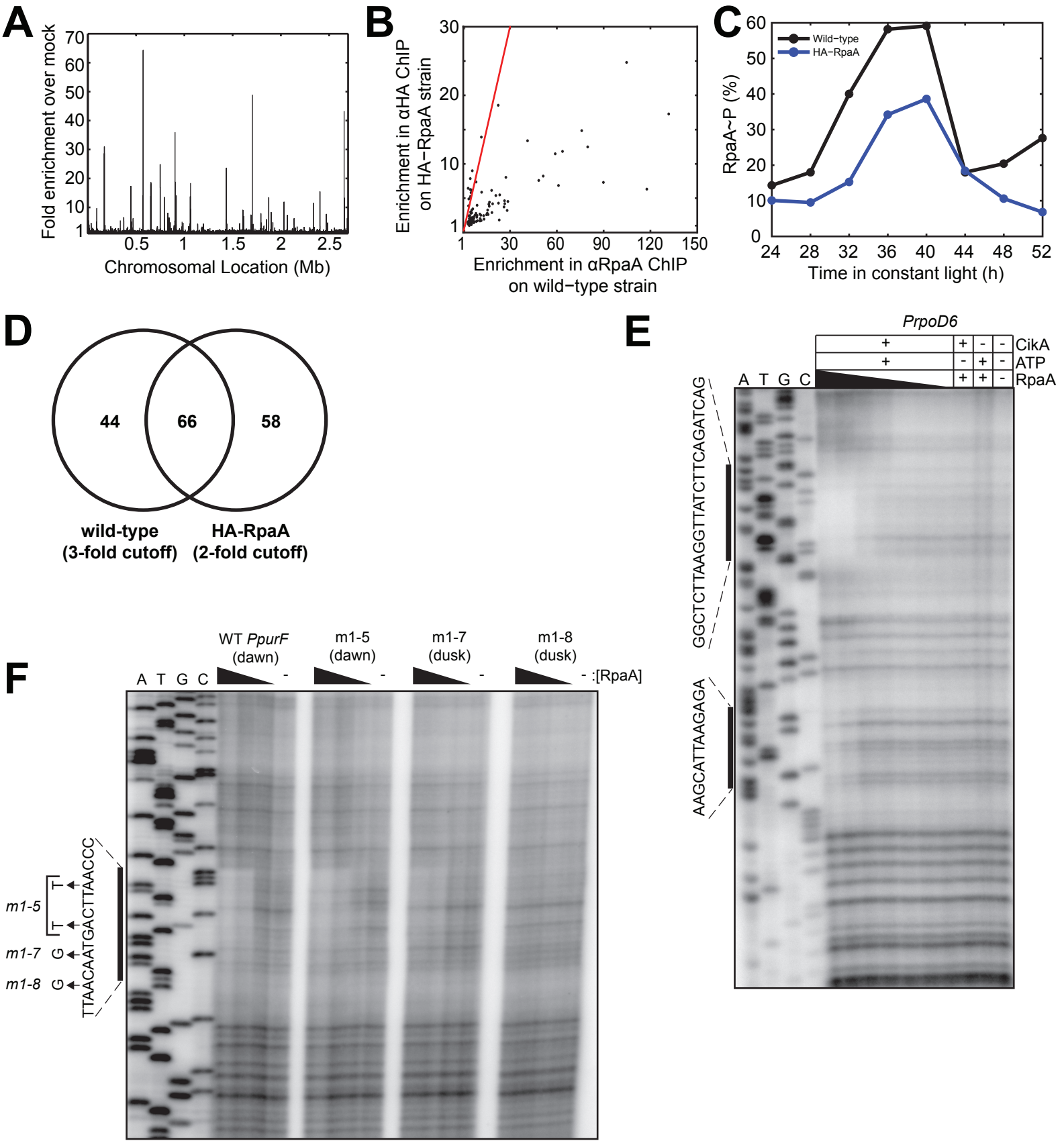

A
control clock rescue
ΔrpaA clock rescue
24 28 32 36 40 44 48 52 56 60 64 68 72
24 28 32 36 40 44 48 52 56 60 64 68 72
(h in constant light)
B
synpcc7942_0834
synpcc7942_1661 (digA)
synpcc7942_1849 (rpoD5)
synpcc7942_2590 (digC)
synpcc7942_2336
synpcc7942_1612 (pntA)
synpcc7942_2591
synpcc7942_1611 (pntA)
synpcc7942_2267
synpcc7942_1784 (sigF2)
synpcc7942_1757
synpcc7942_0304
synpcc7942_2082 (fus)
synpcc7942_2334 (zwf)
synpcc7942_2335 (fbp)
synpcc7942_0599
Log2 expression relative to control rescue time average
Time in constant light (h)
control clock rescue
ΔrpaA clock rescue
C
synpcc7942_0905
synpcc7942_0855 (pixG)
synpcc7942_0856 (pilH)
synpcc7942_0465
synpcc7942_0857
synpcc7942_2450 (pilQ)
synpcc7942_0004 (purF)
synpcc7942_1422 (ccmL)
synpcc7942_0858 (pixJ1)
synpcc7942_2059 (cdv2/sepF)
synpcc7942_1427 (rbcS)
synpcc7942_1832
synpcc7942_1882 (psbU)
synpcc7942_0483
synpcc7942_0505 (glpX)
synpcc7942_2503 (por)
Log2 expression relative to control rescue time average
Time in constant light (h)
control clock rescue
ΔrpaA clock rescue

Figure S3


A
kDa
100
50
37
20
15
Time in constant light (h)
24 28 32 36 40 44 48 52 60
24 28 32 36 40 44
wild-type
rpaA mutant
B
Time in constant light (h)
24 28 32 36 40 44 48 52 56 60
RpaA~P
RpaA
C
Subjective dawn
Subjective dusk
Subjective dawn
kaiBC ChIP
HA-RpaA~P
Normalized level
1.0
0.8
0.6
0.4
0.2
0
24 28 32 36 40 44 48 52
Time in constant light (h)
D
Boiled
Acetyl Phosphate
CikA
ATP
RpaA (μM)
6.0 6.0 3.0 0.6 0.3 6.0 6.0 6.0 3.0 0.6 0.3
RpaA~P
RpaA
E
WT m15 m20 m16
A T G C
m15
m16
m20
GAGCTTAAGACCTCCTTTACCTT
F
PkaiBC reporter
PkaiBC reporter
ΔrpaA Ptrc vector
ΔrpaA Ptrc vector
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53A)
ΔrpaA Ptrc::rpaA(D53A)
IPTG (μM) 0 1000 0 1000 0 3 30 100 1000 0 3 30 100 1000 0 1000
Total RpaA
G
PkaiBC reporter
PkaiBC reporter
ΔrpaA Ptrc vector
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53E)
ΔrpaA Ptrc::rpaA(D53A)
ΔrpaA Ptrc::rpaA(D53A)
IPTG (μM) 0 0 1000 3 3 1000 1000 3 3 1000 1000 1000 1000
Heat - + - - + - + - + - + - +
RpaA~P
RpaA
*
*

Supplemental Figure

Figure S4



Supplemental Figure

Figure S5