

Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Hurley, J. M., Dasgupta, A., Emerson, J. M., Zhou, X., Ringelberg, C. S., Knabe, N., ... & Dunlap, J. C. (2014). Analysis of clock-regulated genes in <i>Neurospora</i> reveals widespread posttranscriptional control of metabolic potential. <i>Proceedings of the National Academy of Sciences</i> , 111(48), 16995-17002. doi:10.1073/pnas.1418963111
DOI	10.1073/pnas.1418963111
Publisher	National Academy of Sciences
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Analysis of clock-regulated genes in *Neurospora* reveals widespread posttranscriptional control of metabolic potential

Jennifer M. Hurley^{a,1}, Arko Dasgupta^{a,1}, Jillian M. Emerson^a, Xiaoying Zhou^a, Carol S. Ringelberg^a, Nicole Knabe^a, Anna M. Lipzen^b, Erika A. Lindquist^b, Christopher G. Daum^b, Kerrie W. Barry^b, Igor V. Grigoriev^b, Kristina M. Smith^c, James E. Galagan^{d,e}, Deborah Bell-Pedersen^f, Michael Freitag^g, Chao Cheng^a, Jennifer J. Loros^{a,g}, and Jay C. Dunlap^{a,2}

^aDepartment of Genetics, Geisel School of Medicine, Hanover, NH 03755; ^bUS Department of Energy Joint Genome Institute, Walnut Creek, CA 94598; ^cDepartment of Biochemistry and Biophysics, Center of Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331; Departments of ^dBiomedical Engineering and ^eMicrobiology, Boston University, Boston, MA 02215; ^fDepartment of Biology, Texas A&M University, College Station, TX 77843; and ^gDepartment of Biochemistry, Geisel School of Medicine, Hanover, NH 03755

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2009.

Contributed by Jay C. Dunlap, October 2, 2014 (sent for review September 20, 2014; reviewed by John B. Hogenesch and Jerome Menet)

Neurospora crassa has been for decades a principal model for filamentous fungal genetics and physiology as well as for understanding the mechanism of circadian clocks. Eukaryotic fungal and animal clocks comprise transcription-translation-based feedback loops that control rhythmic transcription of a substantial fraction of these transcriptomes, yielding the changes in protein abundance that mediate circadian regulation of physiology and metabolism: Understanding circadian control of gene expression is key to understanding eukaryotic, including fungal, physiology. Indeed, the isolation of clock-controlled genes (*ccgs*) was pioneered in *Neurospora* where circadian output begins with binding of the core circadian transcription factor WCC to a subset of *ccg* promoters, including those of many transcription factors. High temporal resolution (2-h) sampling over 48 h using RNA sequencing (RNA-Seq) identified circadianly expressed genes in *Neurospora*, revealing that from ~10% to as much 40% of the transcriptome can be expressed under circadian control. Functional classifications of these genes revealed strong enrichment in pathways involving metabolism, protein synthesis, and stress responses; in broad terms, daytime metabolic potential favors catabolism, energy production, and precursor assembly, whereas night activities favor biosynthesis of cellular components and growth. Discriminative regular expression motif elicitation (DREME) identified key promoter motifs highly correlated with the temporal regulation of *ccgs*. Correlations between *ccg* abundance from RNA-Seq, the degree of *ccg*-promoter activation as reported by *ccg*-promoter-luciferase fusions, and binding of WCC as measured by ChIP-Seq, are not strong. Therefore, although circadian activation is critical to *ccg* rhythmicity, posttranscriptional regulation plays a major role in determining rhythmicity at the mRNA level.

circadian | RNA-Seq | transcription | *Neurospora* | clock-controlled genes

A well-recognized model for fungal genetics, photobiology, and circadian systems, *Neurospora* is the established model for nearly all aspects of growth and metabolism among the filamentous fungi. The core clock of *Neurospora crassa* has been well studied; its oscillator comprises a transcription-translation feedback loop involving a complex of five core proteins, White Collar 1 (WC-1), White Collar 2 (WC-2), Frequency (FRQ), Frequency Interacting RNA Helicase (FRH), and Casein Kinase 1 (CK1), as well as several ancillary factors. Transcription factors (TFs) WC-1 and WC-2 form the white collar complex (WCC), which drives the rhythmic expression of FRQ. FRQ binds to FRH to form the FRQ/FRH complex (FFC), which then acts with CK1 on the WCC to inhibit its activity, thus closing the loop. Throughout the circadian cycle, FRQ interacts with many partners that affect the number and location of its posttranslational modifications as well as its stability; it is these

posttranslational modifications of FRQ that set the length of the circadian period (1–3).

Although historically much interest has focused on clock mechanism, it is through circadian control of transcription that the clock exerts behavioral, physiological, and metabolic control over an organism. A substantial fraction of expressed genes are regulated by the circadian clock in a variety of organisms (refs. 4–7; reviewed in ref. 8), and previous work using ESTs or microarrays in *N. crassa* has reported ~5–15% of genes to be rhythmic clock-controlled genes (*ccgs*) (9–12). However, prior work is flawed in that there is serious concern in calling a gene rhythmic from expression data based on a single circadian day or even sampled over multiple days with a resolution less than 2 h; indeed, sampling only every 4 h may miss up to 90% of clock-regulated genes (13). No extended time series has been sampled at high temporal resolution in any fungus. High-throughput sequencing techniques such as RNA sequencing (RNA-Seq), being more quantitative and having greater dynamic range than microarrays

Significance

Circadian clocks regulate gene expression levels to allow an organism to anticipate environmental conditions. These clocks reside in all the major branches of life and confer a competitive advantage to the organisms that maintain them. The clock in the fungus *Neurospora crassa* is an excellent model for basic understanding of core circadian architecture as well as for filamentous fungi. Here, we identify genes whose expression is clock regulated; indeed, as much as 40% of the transcriptome may be clock regulated, broadly directing daytime catabolism and nighttime growth. Both transcriptional control and posttranscriptional regulation play major roles in control of cycling transcripts such that DNA binding of transcription factors alone appears insufficient to set the phase of circadian transcription.

Author contributions: J.M.H., A.D., J.J.L., and J.C.D. designed research; J.M.H., A.D., J.M.E., X.Z., C.S.R., N.K., A.M.L., E.A.L., C.G.D., K.W.B., I.V.G., K.M.S., M.F., and C.C. performed research; J.E.G. and D.B.-P. contributed new reagents/analytic tools; J.M.H., A.D., J.M.E., X.Z., C.C., J.J.L., and J.C.D. analyzed data; and J.M.H., A.D., J.J.L., and J.C.D. wrote the paper.

Reviewers: J.B.H., University of Pennsylvania School of Medicine; and J.M., Texas A&M University.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the National Center for Biotechnology Information Short Read Archive (SRA) (accession nos. SRP045821 and SRP046458).

¹J.M.H. and A.D. contributed equally to this work.

²To whom correspondence should be addressed. Email: jay.c.dunlap@dartmouth.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1418963111/-DCSupplemental.

(14, 15), have been used in mammalian and fly systems to improve the qualitative and the quantitative nature of the identification of rhythmic genes (4, 6, 16). Although these studies provided novel insights, they were either completed under conditions of diurnal entrainment by light–dark cycles (which will, as noted by ref. 16, confound acute responses to light with endogenous circadian control) or were sampled at temporal resolution >2 h.

To profit from conceptual and technical advances and to avoid recognized problems, we have used RNA-Seq to identify rhythmic transcripts (mRNA) in *N. crassa* in triplicate time courses, each spanning 48 h with a 2-h resolution, all collected in constant-dark conditions. We identified 872 genes as being consistently rhythmic, constituting roughly 10% of the *N. crassa* genome, and a larger set of 4,112 genes, most of which are likely rhythmic. We report that rhythmic genes are enriched for specific pathways at different times of day and furthermore have identified promoter motifs that may be important for the temporal control of the activation of these genes. Analyses of the correlation between clock-regulated promoter activation and mRNA levels reveal that mRNA levels are controlled by the clock both via rhythmic gene activation at the promoter and via posttranscriptional controls that can separately govern rhythmic steady-state transcript levels. To discern the mechanism through which the WCC controls *cgc* expression, we performed ChIP on WC-2 over a circadian time course and show that it displays a single peak of DNA binding during the day, regardless of the time of the peak of the *cgc* it activates. These data are consistent with the action of the WCC as the initial rhythm generating TFs in a hierarchical cascade (17).

Results

Genome-Wide Sequencing of *Neurospora* Circadian Transcription. To identify mRNAs that cycle in a circadian manner, we performed three circadian time courses, in each case extracting RNA from tissue every 2 h over a period of 48 h (Fig. S1A). These RNA samples were then analyzed for *frq* mRNA levels by real-time PCR over circadian time to ensure that the samples were oscillating as expected. *frq* mRNA levels oscillated with an amplitude of twofold to threefold over the circadian day (Fig. S1B), confirming that the extracted RNA samples constituted a rhythmic RNA set. These samples were then used to prepare Illumina libraries using standard protocols and subjected to Illumina high-throughput sequencing. The raw data were then normalized to reads per kilobase of exon model per million mapped reads (RPKM) values and underwent further normalization and smoothing (Materials and Methods).

To determine how many of the genes were rhythmic at the transcript level, we analyzed the normalized data using JTK cycle, a nonparametric algorithm for detecting rhythmic components in genome-scale datasets (18). For each dataset, we identified roughly 1,600 genes (roughly 17% of the genome) that were cycling according to JTK cycle with a q value (the minimum false discovery rate) of <0.05 as a conservative estimate (Fig. 1A and Dataset S1). In this study, we considered mRNA that was identified as rhythmic in two of the three datasets as consistently rhythmic, and using this cutoff we identified 872 consistently cycling genes (Fig. 1B and Dataset S1). In all, 4,112 genes were found to be rhythmic in at least one dataset.

When the peaks of expression of mRNA levels for the significantly rhythmic genes were analyzed by assessing their expression levels on a heat map, it was apparent that peak times of mRNA expression were not evenly spaced but biphasic, peaking during the late night to early circadian day [circadian time (CT) 20–5 h], as previously seen in *Neurospora* and as would be expected from the activation due to the WCC (Fig. 1A and C) (1, 19); this is also seen in insect and mammalian models (6, 16). Roughly 35% of the genes identified as rhythmic peaked between CT 11–15 h, which would be assigned to the evening gene category (Fig. 1C).

Following Gene Activation with Luciferase Reporters Affords High-Resolution Tracking of Circadian Promoter Activation. To correlate rhythmic promoter activation of circadian genes with rhythmic

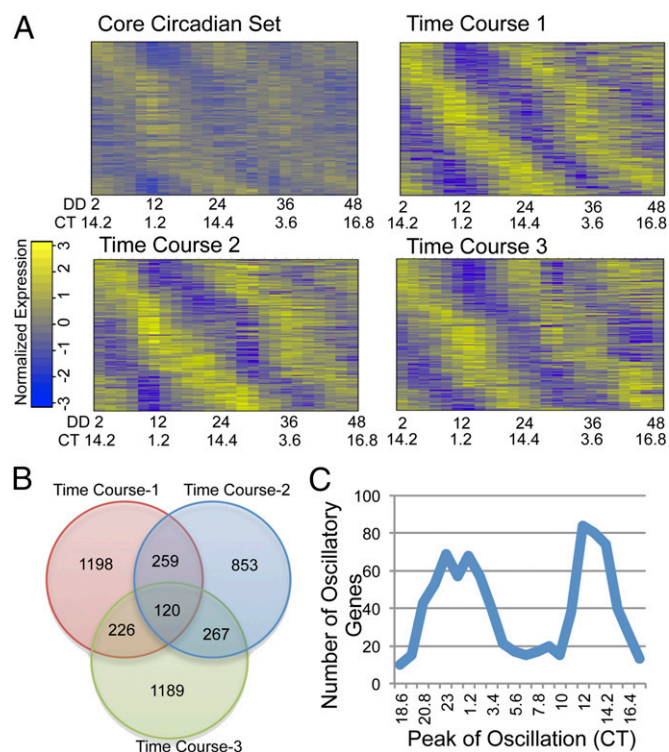


Fig. 1. *Neurospora* circadian mRNA-Seq. (A) Heat maps of the core 872 genes that are rhythmic in at least two of the three datasets using information from the average of the three datasets (core circadian set) as well as the three individual time courses (time courses 1, 2, and 3). CT, circadian time; DD, time after light to dark transfer. The color bar shows \log_2 of the normalized level of gene expression. (B) A Venn diagram of the genes found to be rhythmic in the three individual time courses and the overlap between the three sets. (C) A graph depicting the number of genes that are rhythmic at a given CT.

expression of mRNA, we tracked promoter activation of 296 genes known to be bound by WC-2 based on ChIP or known to be decidedly rhythmic based on other work (20, 21). We generated DNA constructs that replaced the coding sequence of the gene with an *N. crassa* codon optimized luciferase gene (22) for each of the 296 genes investigated (Fig. S2A), preserving the promoter and 500 bp of 3'-UTR. Promoter sizes varied between 1,500 and 10,000 bp and were based on the location of adjacent genes from the RNA-Seq data (www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html).

These cassettes were then transformed into the *cyclosporin-resistance 1* (*csr-1*) locus of wild type (FGSC2489) as previously described (23). For each cassette, at least three transformants were subjected to 12:12 dark:light (D:L) entrainment for 48 h before being released to DD and imaged by a CCD camera (Fig. S2B). Genes encoding each of the four core clock proteins were used as functional controls for the assay. As expected, *frq* was highly rhythmic under this analysis (Fig. 2) (10, 22, 24–26). Although not predicted from protein levels to be strongly rhythmic at the mRNA level (27), *wc-2*, *wc-1*, and *frh* showed promoter rhythms under our experimental conditions, with the rhythmic activation at the *frh* promoter less defined than the *wc-1* and *wc-2* promoters (10) (Fig. 2 and Fig. S3A).

In total, of the 296 genes tested in various media, 230 showed rhythms in luciferase reporter activity for at least the first several days in at least one medium (Dataset S2). Known *cgc*s (*cgc-1*, -2, -4, -6, -7, -8, -9, -12, -13, -14, and -15) displayed varying levels of rhythmicity (Fig. S3B), with some displaying extended and robust rhythms and others showing little to no persistence of rhythms beyond the first peak (compare *cgc-4* to *cgc-13*) (Fig. S3B). This

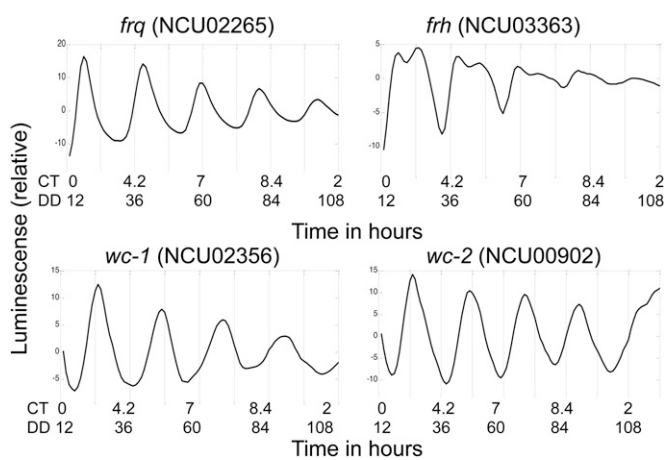


Fig. 2. Luciferase traces track rhythmic genes in the core clock. Luciferase traces acquired using the described format for the four core clock genes after detrending and rescaling.

suggested that rhythmic expression of even bona fide *ccgs* might not correlate with promoter activation rhythms and that post-transcriptional mechanisms may be at play. This analysis also identified rhythms in 75 low-abundance transcripts for which rhythmicity was not detected by RNA-Seq, a finding discussed below.

To a limited extent, we also examined the effect of growth medium on rhythmicity. While carrying out controls for other experiments, we noted that the amplitude of *ccg*-luciferase rhythms were often larger when 0.01 M quinic acid was added to a low-arginine, low-glucose media (QA media) (Dataset S3). To investigate this more systematically, 157 genes were investigated on QA media and race tube (RT) media. Fifty percent of the genes were rhythmic in both RT and QA media, whereas 35% were rhythmic only on QA, 5% were rhythmic only on RT, and 10% were rhythmic in neither. The lists of distinct gene sets from each independent category were subjected to Functional Category (FunCat) term analysis using FungiFun to determine whether the genes were enriched in a particular pathway with a value of $P < 0.05$ as compared with the entire *Neurospora* genome (28) (Fig. 3). FunCat classifies genes based on the pathway in which the gene acts instead of the chemical mechanism (29). Genes that were rhythmic in both media were enriched in core cellular functions including cell cycle, cell transport, and protein fate (Fig. 3A). However, genes that were only rhythmic in either QA or RT media were enriched in the metabolic FunCat term category (Fig. 3B and C). This interesting observation suggests that the subset of genes circadianly regulated by the core clock depends on environmental conditions (in this case growth medium). Although circadian control over core biological functions (e.g., cell cycle, cell transport, etc.) is required irrespective of the environment, circadian regulation is plastic enough to control gene expression hierarchy to best suit the need of the organism in a particular environment.

These data are consistent with a study of the effect of different nitrogen sources on the gene expression patterns in *Saccharomyces cerevisiae*, which revealed that close to 500 (~7.5%) genes are regulated differentially by different sources of nitrogen (30). Our observation in *N. crassa* suggests that the core clock exerts substantial control on at least a subset of these differentially regulated genes and hence forms a conduit for environmental sensing and circadian gene expression.

Genes Enriched in Circadian Transcription. To identify pathways that are enriched in circadian transcription, we subjected the 1,050 genes identified as rhythmic by RNA-Seq analysis as well as luciferase monitoring to FunCat categorization using FungiFun (28) (Dataset S4). The genes that were functionally annotated according to FunCat and the *Neurospora crassa* Genome Project were

subjected to an enrichment analysis that identified over-represented modulated categories compared with the entire *Neurospora* genome (28) (Datasets S5 and S6). Some of these categories, such as those representing metabolism, protein fate, protein synthesis, DNA processing, cellular transport, and the cell stress response, were found to be highly enriched in the FunCat analysis of the circadianly regulated genes (Fig. 4).

In the metabolic gene subset, genes that are involved in the regulation of carbohydrate, fatty acid, and vitamin metabolism are particularly enriched and are mostly up-regulated during the morning-to-day phase of the clock (CT 0–8). This is also the case for some of the other FunCat categories, including the stress response and cellular transport. It is likely that genes in these categories are up-regulated in anticipation of stress from heat as well as light that fungi face on a daily basis (31, 32) (Fig. S4A–D) (Dataset S7). Interestingly, categories such as protein synthesis and DNA processing reach their peak expression levels during the late circadian day and evening (CT 10–20) (Fig. S4E and F) (Dataset S7), suggesting that *Neurospora* builds up basic cell building material in the day to produce the more complicated resources at night.

To confirm this finding, we then subjected the data to FunCat term analysis in a time-specific manner. Rhythmic genes were divided into three categories by peak of expression (circadian night, circadian morning, and circadian afternoon/evening), and each of these gene lists was sorted by FunCat (Dataset S8). Metabolism was highly enriched in the late night and circadian morning, whereas protein synthesis was highly enriched in the circadian afternoon and early night (Fig. 5A). One of the more interesting findings in the FunCat terms relates to the genes involved in the production and processing of proteins. Although genes involved in ribosome biogenesis and translation peak in the circadian dusk to early evening, genes involved in downstream functions involving proteins, like those involved in protein folding, targeting, and modification, peak in the early circadian morning (Fig. 5B). Not surprisingly, genes like methionine synthase and deoxyhypusine hydroxylase (required for eIF5A) peak during the early circadian evening. We also evaluated our rhythmic gene set using the Kyoto Encyclopedia of Genes and Genomes or KEGG (33). Using this method, we found several enriched pathways. When we graphed the peaks of each of these pathways, we found that, similar to the protein pathway, the peaks of the pathways correlated well with the accumulation of biological materials (Fig. 5C and Fig. S5). For example, early in the circadian day, genes involved in isoprenoid metabolism peak, leading to the

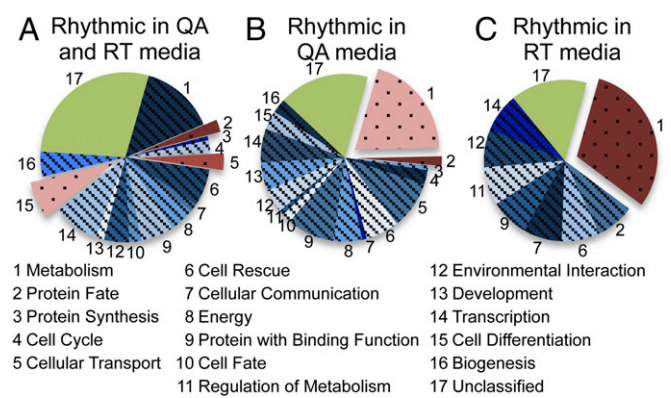


Fig. 3. Media changes affect which types of genes are regulated by the clock. FunCat term analysis of the genes found to be rhythmic at the activation level in (A) both QA and RT media; (B) QA media only; and (C) RT media only. In this and all subsequent figures displaying FunCat analysis, categories considered significant (values of $P < 0.05$) are pushed away from the pie chart and highlighted in red, blue categories have a value of $P > 0.05$, and green represents genes that are unclassified. The number next to each slice corresponds to the FunCat category.

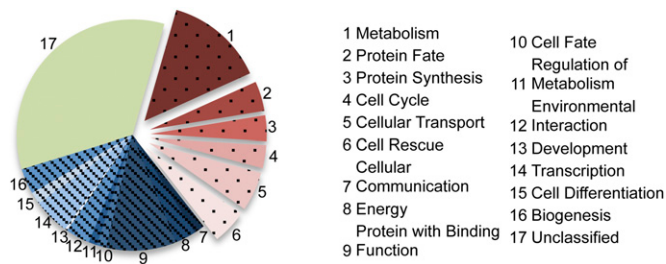


Fig. 4. FunCat term analysis highlights rhythmic gene enrichment in specific cellular processes. Breakdown of FunCat terms from all genes designated rhythmic in this dataset with a heat map of the specific genes.

development of protective materials including carotenoids (Fig. 5C). Genes involved in pyruvate metabolism (which develops precursors for fatty acid biosynthesis as well as providing energy) and fermentation peak first, followed by those involved in steroid biosynthesis (Fig. 5C and Fig. S5 A and B). Finally, late in the circadian day/early evening, there is a peak in fatty acid biosynthesis, using the building blocks that were created earlier in the circadian day (Fig. 5C and Fig. S5C). The intricate phase networking of the circadianly regulated genes shows the level of fine-tuning the clock provides the biological system.

Discriminative Regular Expression Motif Elicitation Analysis Identifies Motifs Correlated with Circadian Regulation. Using the 872 genes identified as being rhythmic based on RNA-Seq analyses, we used discriminative regular expression motif elicitation (DREME) to identify motifs that are significantly enriched among rhythmic genes (34). We analyzed 1,000 bp of the promoter of these 872 rhythmic genes and identified a total of 36 short (up to 8 bp in length) DNA motifs that were enriched in these sequences. To obtain a confident list of circadian-related motifs, we further examined the enrichment of these motifs in the promoters of rhythmic genes that displayed peak expression at different time windows of the circadian clock compared with all genes in the genome that contained these motifs. For example, for the motif STACASTA, we used the Find Individual Motif Occurrences (FIMO) algorithm to examine the motif's existence in the promoters of all genes. We found that the promoters of 2,823 genes out of the 9,729 gene promoters in the *Neurospora* genome (29%) contained this motif. We then looked at the number of genes from the rhythmic dataset that peaked in each specific time window; in the CT 20.8–1.2 window, there were 292 genes from the 872 gene rhythmic dataset with a peak of expression, and 113 of these 292 (39%) have the motif STACASTA. We found this to be a significantly enriched motif at this window based on Fisher's exact test: the probability that 113 genes contain this motif from among the 292 peaking at this time is significantly (P value of 0.0002) greater than the probability that 2,823 promoters from the genome contain this motif. Using this method, motifs were identified as circadian-related if they were significantly enriched ($P < 0.001$) in at least one specific time window.

Out of the 36 motifs, 4 were found to be significantly enriched in a circadian-related manner: STACASTA, GRCGGGA, GC-RCTAAC, and GVCAGCCA (Fig. 6A). We performed FunCat analysis on the genes identified from the 872 rhythmic genes that possessed these motifs in their promoters (Dataset S9). We found that two of the motifs were enriched in the same FunCat terms (metabolism, protein fate, protein synthesis, and cell cycle) that were generally associated with rhythmic gene expression (compare GRCGGGA and GVCAGCCA from Fig. 6B to Fig. 4). However, the other two motifs (STACASTA and GCRCTAAC) were preferentially enriched primarily in the area of “Proteins with binding function or cofactor,” suggesting that these motifs specifically affected this pathway.

By investigating the motif enrichment across different circadian time windows, this method not only identified with high

confidence several circadian-related motifs but also provided the particular time at which they might play regulatory functions. As shown in Fig. 6C, the motifs STACASTA and GRCGGGA are most significantly enriched in promoters of rhythmic genes with peak expression in the late night to early circadian morning. For motifs GCRCTAAC and GVCAGCCA, the opposite was true—most of the rhythmic genes associated with the motifs peaked in the circadian evening (Fig. 6C). This suggests that the active phases for these motifs are during these windows and the role that the motifs play in the transcription of the genes associated with these motifs occurs during this time.

Differences in Peak Times Between Genes Showing Rhythmic Promoter Activation and Rhythmic Steady-State mRNA Levels Suggests Posttranscriptional Regulation of Clock-Controlled mRNAs.

The genes chosen for analysis by luciferase were selected independently of the 872 rhythmic genes. Therefore, the overlap of rhythmic genes between the two sets affords an estimate of how often rhythmic transcription (as assessed by luciferase) leads to rhythmic RNA levels (as assessed by RNA-Seq). As noted above, a variety of different media were used for luciferase activity rhythm measurements; in all, 187 genes were examined by

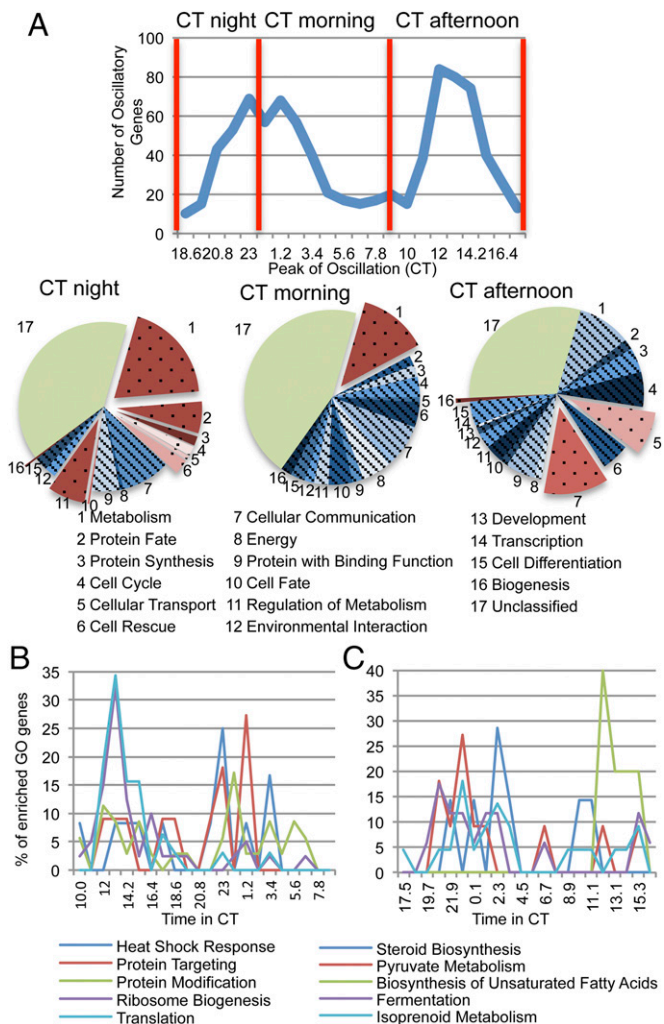


Fig. 5. FunCat and KEGG analysis highlight time-of-day-specific gene enrichment. (A) FunCat analysis of rhythmic genes peaking at specific circadian times. (B) The peak in expression of the genes in specific FunCat terms that are involved in protein production. (C) The peak in expression of genes in specific FunCat and KEGG analysis that are involved in metabolic function within the cell.

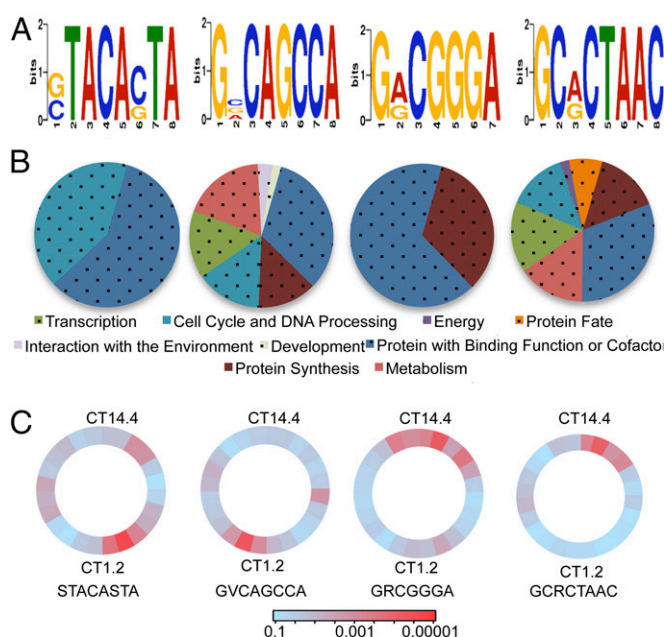


Fig. 6. DREME analysis highlights time-of-day-specific promoter elements. (A) The four significantly enriched promoter motifs identified from the rhythmic gene set. (B) FunCat term analysis of the genes in the rhythmic dataset containing the above sequence in their promoters. (C) Position weight motif plots of motifs from A, showing the *P* value of the significance of enrichment of a specific motif identified from the rhythmic genes as compared the same to motif in promoters over entire genome, in a 4-h sliding scale over 24 h. Thus, dark red corresponds to a value of *P* < 0.00001.

luciferase using a medium comparable to that used for the RNA-Seq experiments. Of these, 96 were rhythmic, 24 of which (25%) were rhythmic both by luciferase and by RNA-Seq, whereas 72 (75%) were rhythmic only by luciferase (Dataset S10). Although some of these are low-abundance transcripts for which the limits to consensus sequencing did not produce statistically different RPKM read numbers at different times of day, many probably reflect genes whose transcripts may be too stable to cycle on a circadian timescale. Conversely, an additional 11 genes that were identified as rhythmic by RNA-Seq failed to show rhythms by luciferase, again indicating that rhythmicity can be conferred on transcripts by regulation occurring after a gene is transcribed.

These conclusions were supported by independent analyses of individual genes by RT-PCR using the RNA samples used for the RNA-Seq analyses. The correlation between rhythmicity as detected by RT-PCR and RNA-Seq was very high: of the 29 candidate rhythmic genes (from RNA-Seq) analyzed by RT-PCR, rhythmicity was confirmed for 27 (95%) (Fig. S6). In many cases where rhythmicity was detected by luciferase but not by RNA-Seq or by RT-PCR, the transcripts were of low abundance. For instance, Fig. 7A shows a gene (NCU08141) that is plainly rhythmic by luciferase but appears to show ultradian cycling by RNA-Seq and by RT-PCR; however, the transcript abundance arising from this gene is quite low. For other genes, however, transcripts are quite abundant, and rhythmicity is plain in the RNA-Seq and RT-PCR analysis but is plainly muted by luciferase reporter assays (Fig. 7B). Such genes are strong candidates for regulation at levels following transcription. The existence of such regulation is further supported by differences in phase between luciferase assays that report promoter activity and sequencing/RT-PCR that report transcript abundance. As an example, Fig. 7C shows NCU06961 that is robustly rhythmic by both assays but displays a significant phase lag in abundance compared with transcription.

The WCC Acts as the Initiating TF for Gene Rhythmicity. Circadian regulation of WCC activity in the core clock is understood as

providing the principal means of circadian output from the clock, when the WCC regulates expression of genes whose products do not impact the core clock (1, 35). Thus, examining the spectrum of gene promoters bound by WC-2 (reflecting the WCC) and the time at which peak binding occurs affords a glimpse at the first step of output. To identify the circadian targets of WC-2, we performed a ChIP sequencing assay (ChIP-Seq) at regular intervals on cultures that were kept in the dark (after transfer from constant light) using antibody against WC-2. We then performed a RT-PCR using probes for the proximal light-regulated element (pLRE) and Clock box (C-box) (36, 37) motifs upstream of the *frq* gene to test whether our ChIP worked as expected. We found that occupation at the pLRE decreased following the light-to-dark transition and remained low for the remainder of the time points; however, there was still appreciable binding to the pLRE after even 8 h in darkness, establishing that the effects of the previous light cycle on gene expression can be long-lasting (Fig. 8A). This serves as a caution indicating that analyses of “circadian” gene expression must not rely solely on data from the first day in darkness. Also, as expected, occupation of the C-box changed rhythmically over the circadian day, peaking at about 16 h after light-to-dark transition [Fig. 8A (38)]. MACS software was used to identify regions in the genome where WC-2 binding was enriched compared with a WC-1 knockout strain in which the WCC does not bind DNA. A peak was identified at both the pLRE and C-box of *frq* (Dataset S11). We then compared these results from the ChIP-Seq data to the ChIP-RT-PCR and found that the reported peaks matched the RT-PCR data (Fig. 8A). When the WC-2 ChIP-Seq data were examined all together by MACS, WC-2 was found to bind at nearly 300 sites in the genome with a peak in the late circadian night similar to that found for the C-box (Fig. 8B). Comparing the genes that were bound by WC-2 in the ChIP experiment

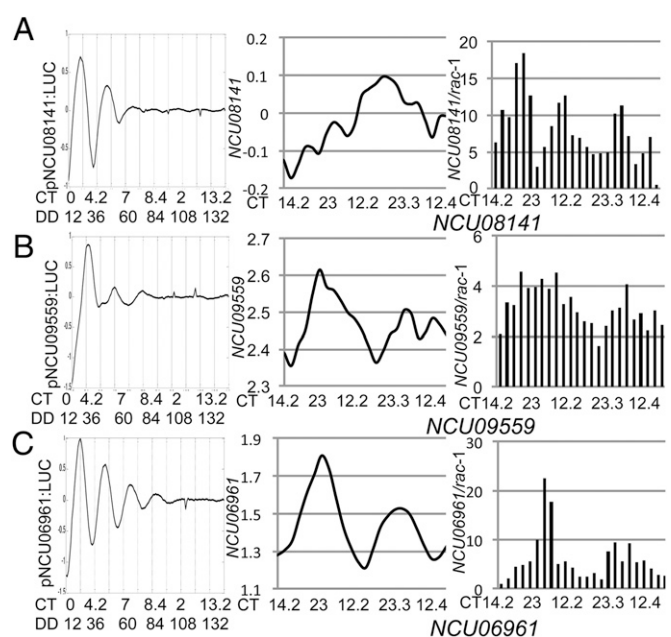


Fig. 7. RNA-Seq and real-time PCR correlate weakly with promoter activation. A comparison between luciferase traces, RNA-Seq data and RT-PCR analysis for three genes that exemplify cases where a gene is rhythmic by Luc only (A), by RNA-Seq only (B), and by both methods (C). (A) NCU08141 showed rhythms by luciferase but not by RNA-Seq or RT-PCR analysis, most likely due to low transcript abundance. (B) NCU09559 demonstrated rhythms by RNA-Seq and RT-PCR but not by luciferase, most likely due to post-transcriptional regulation. (C) NCU06961 exhibited rhythms with all methods but with a lag between peak of activation and mRNA levels, potentially from posttranscriptional mechanisms. Luciferase data are detrended and rescaled and RPKM values are shown in log scale.

to a list of predicted TFs in *Neurospora*, we found that 8% of predicted TFs were bound by WC-2.

To determine whether the WCC is the initiator of rhythmicity but not necessarily responsible for setting the phase of *cggs*, we compared the genes that have WC-2 peaks to genes that are rhythmic in the RNA-Seq dataset. Of the 872 genes that were rhythmic from the RNA-Seq, 36 of them had major binding sites for WC-2. We then mapped the peak of binding by WC-2 to the peak of mRNA levels for each gene. It was evident that, although the peak binding time was in all cases near to CT0 for WC-2, the peak mRNA levels were delayed to phases spanning the circadian cycle (Fig. 8C). We compared the ChIP-Seq data to the RNA-Seq data for a few of the genes and noted that this trend was confirmed by visual observation of specific genes (Fig. S7A–D).

Discussion

Circadian regulation of gene expression is the point at which the circadian oscillator begins to impact the biology of the organism. For this reason, a detailed description of the circadian transcriptome is key to understanding the physiology of *Neurospora*, and indeed elucidation of the circadian regulation of the transcriptome has thus been a topic of interest in all of the key circadian model organisms (4, 6, 10, 11, 16, 39–43). Temporal control over the expression of key metabolic/functional genes has always been a prediction for the role of an “anticipatory” core clock in any organism, and *cggs* have been identified at the genome-wide level through the use of ESTs, microarrays, and RNA-Seq. In fungi and animals, the clock generates rhythmic activation of a heterodimeric TF joined by interacting PAS domains, and the first step in output is the action of this transcription factor to activate genes whose products do not directly impact the oscillator. Studies in animal

systems have used various techniques, beginning with ChIP-Seq, to report binding of the PAS heterodimer that initiates transcription and Nascent-Seq (e.g., refs. 6 and 16) to distinguish transcript rhythmicity conferred by transcription from rhythmicity later conferred through posttranscriptional regulation (reviewed in ref. 44). Recent years have also seen increased sophistication in data analysis software and in realization of the importance of extended time courses and dense (2 h or less) sampling resolution for the identification of *cggs* (e.g., refs. 13 and 18). We aimed here for a definitive dataset, and the data provided from our studies are unique in using both true circadian (nonentrained) conditions and high temporal resolution sampling over two cycles.

mRNAs displaying rhythmic abundance were identified using three independent 48-h time courses with a 2-h resolution using RNA-Seq, the most thorough analysis of a rhythmic transcriptome in any organism to date. Because Nascent-Seq requires the isolation of nuclei, difficult across a circadian time course at this resolution in *N. crassa*, we have used a high-throughput promoter-luc strategy to correlate promoter activation with rhythmicity at the mRNA level. Although this approach lacks the depth of transcriptome analysis afforded by Nascent-Seq, its advantages include much finer detail of the transcriptional dynamics of individual genes, better estimates of phase, estimates of rhythm damping and persistence not afforded by Nascent-Seq except at inordinate expense, and creation of reporters that can be reused for subsequent dissection of the regulatory network.

We find that there is a significant discordance between the promoter activation of a gene and rhythmicity of its transcript at the mRNA level; i.e., we have identified a number of cases where promoters are rhythmically activated but the corresponding mRNAs do not display rhythmic abundance and vice versa (Fig. 7) (see also ref. 16). This discordance is also strikingly similar to that found in a similar analysis completed in mouse liver (45). Our analysis of the individual genes using RT-PCR showed good correlation with the RNA-Seq data (95% of the genes rhythmic by RNA-Seq are rhythmic by RT-PCR), suggesting that there is indeed a level of posttranscriptional regulation between gene activation and rhythmic abundance of mRNA (Fig. 7A and B). Moreover, in some cases, we find significant phase differences between the peak of gene activation and rhythmic mRNA levels, which suggests that accumulation of mRNA is not always solely governed by the activation of the corresponding gene.

Several possible mechanisms may underlie this discordance outside of artifactual issues, including specific RNA stability and/or differential RNA processing. One can also envision a situation where RNA from a nonrhythmically driven gene could have different stability at different times of the day owing to its binding to stabilizing/destabilizing RNA-binding proteins (RBPs). This time-specific binding might induce robust rhythms at the mRNA level. In fact, the strategy of time-of-day-dependent differential transcript stability exists in *N. crassa* to generate robust *frq* rhythms (46). Similarly, *CCR-LIKE (CCL)* and *SENESCENCE ASSOCIATED GENE 1* transcripts in *Arabidopsis* have time-of-day-specific changes in stability, and the changes in *CCL* mRNA stability are under bona fide clock control, i.e., they persist under free-running conditions (47). Other examples include RBPs like HuR that binds to and stabilizes cyclin D1 transcript as opposed to AUF1 that binds to and destabilizes cyclin D1 transcript—both processes required for a highly coordinated and regulated cell cycle progression (48).

Our RNA-Seq data also allowed us to correlate the phase of WCC binding (WC-2 ChIP) with the phase of expression of its target genes. It is quite obvious from our analysis that, although WCC binding is very discrete, the phase of expression for its target genes is heterogeneous. A similar scenario was observed when the phase of BMAL1-CLOCK binding and the phase of expression of its target genes were compared in mouse livers (16). However, this contrasts with the observation in flies where both binding of CLK:CYC to its target genes and the expression of its targets were found to be in the same phase (49). The

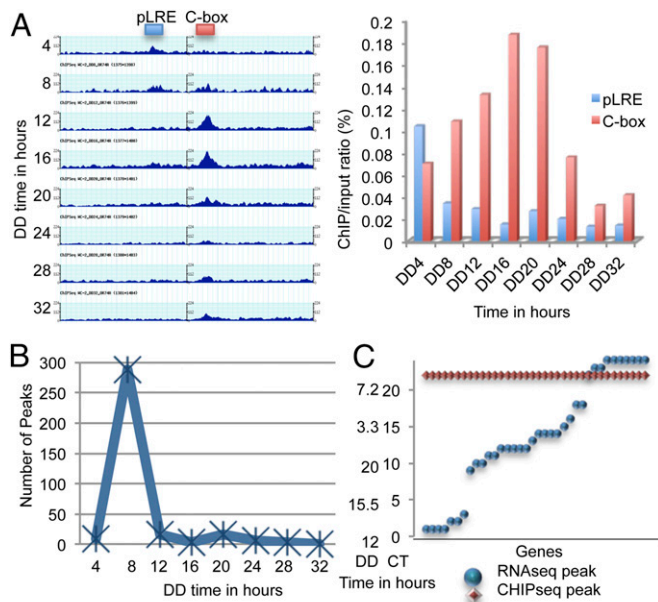


Fig. 8. In WC-2 target genes, the delay between WC-2 binding and mRNA levels suggests that the WCC acts as a pioneer TF. (A) Left panel shows results of ChIP-Seq of WC-2 revealing sites bound within the *frq* promoter. Here, *frq* is transcribed right to left with the pLRE close to the transcription start site and the C-box more distal. Plots for ChIP-Seq data from *gbrowse* (60) are shown for each time in darkness from DD4 to DD32. All are plotted as FPKM on the same vertical scale. Right panel shows the RT-PCR of DNA used for ChIP-Seq at both the pLRE and C-box of *frq* demonstrating peak binding of the C-box at DD16 as in the left panel. (B) The number of WC-2 peaks identified by MACS plotted against circadian time shows the highest number of peaks identified as bound by WC-2 is at DD8. (C) Distribution of the peak phase of WC-2 DNA binding (in red) and the peak phase for mRNA levels of the same gene (in blue) for the direct targets of WC-2.

absence of a fixed phase of target gene expression despite a discrete phase of WCC binding suggests that, for genes that show delayed expression, there is a requirement for additional factors (TFs, chromatin remodelers) to set the phase of expression for these genes. More recently, it has been shown that BMAL1-CLOCK behaves like a pioneer-like TF by modulating rhythmic opening of chromatin through chromatin modification, which allows other TFs to bind adjacent to BMAL1-CLOCK (50) and modulate gene expression. Recent research from our group has shown that WCC recruits the SWI-SNF complex (51) and modulates rhythmic opening of chromatin at the *frq* locus. It is thus possible that WCC similarly binds to various targets in the genome as a pioneer-like TF and promotes remodeling of the chromatin so as to allow other TFs to bind. This could potentially explain the heterogeneity in the phase of peak expression of its target genes.

Setting aside the exact means of circadian control, a major outcome of this study is the picture it affords of the circadian control of physiology in *Neurospora*. Concomitant functional and pathway categorization of our rhythmic genes using FunCat and KEGG analysis, respectively, revealed broad categories (e.g., metabolism, protein fate, DNA processing, etc.) as well as sub-categories/pathways that were enriched and could be studied for temporal relevance in the circadian biology of the organism. For example, genes involved in ribosome biogenesis and protein assembly peak late in the circadian day compared with genes involved in protein folding, targeting, and modification (Fig. 5B), similar to what is seen in mouse liver (52). Logically, this makes sense as one can envision the peak for genes required for translational activity and ribosome assembly to precede the peak for protein folding/targeting/modification as the latter events can be “phase-concentrated” after a burst of protein translation, thereby making the process more efficient (Fig. 5B). Indeed, most major categories of cellular metabolism display some rhythmic control (Fig. 4). When querying specific pathways for rhythmicity, catabolic pathways such as pyruvate metabolism peak earlier in the day than those involved in anabolic pathways like steroid biosynthesis and fatty acid biosynthesis (Fig. 5C and Fig. S5). The enrichment of metabolic genes is not altogether surprising as many metabolic genes have been shown to be circadianly controlled in several other studies (e.g., ref. 53). As a broad generalization, the transcriptome data suggest that daytime activities involve catabolism, energy production, and assembly of precursors, whereas nighttime activities favor biosynthesis of cellular components and growth.

Moreover, it is now highly regarded that the clock is not just a static master controller that sets up anticipatory metabolic gene expression but can also serve as a conduit between the environment and the transcriptome. In other words, the subset of genes that is circadianly expressed can be altered depending on the environment and the clock can coordinate this alteration. Although it is tough to test this from RNA-Seq data (all biological sets grown in one media), use of high-throughput promoter-luc assays clearly shows that subsets of the genes were rhythmically activated in specific media (Fig. 3). These subsets were again enriched in metabolic FunCat categories suggesting the core-clock indeed coordinates, undoubtedly with the help of other factors (TFs, metabolite-binding proteins, etc.), the gene expression landscape as a function of the environment. More interestingly, rhythmicity of genes involved in core cellular functions (cell cycle, cell transport, etc.) appears not to be dependent on the media (environment), suggesting a more stringent and environmentally refractory circadian control over these genes.

The DREME analysis of rhythmic genes (872) in our RNA-Seq dataset also allowed us to identify four motifs that were significantly enriched (STACASTA, GRCGGGA, GCRCTAAC, and GVCAGCCA) in the promoters of rhythmic genes compared with the rest of the genome. For two of the motifs (STACASTA and GRCGGGA), we find that time of day (CTs) at which there

is significant enrichment in the promoters of rhythmic genes is coincident with the time of day (morning) where a large fraction of rhythmic genes shows peak mRNA expression in the rhythmic dataset. This appears not to be the case with the other two motifs (GCRCTAAC and GVCAGCCA) where the time of significant enrichment is in the early evening and does not coincide with the time where most genes in the rhythmic dataset show peak mRNA levels. This analysis does not allow us to conclude what role a particular motif is playing during its active period, whether it is an activator or repressor, only that the impact on gene expression for each motif is likely during that circadian time window.

Approximately 10% of the *N. crassa* transcriptome is reproducibly rhythmic at the mRNA level under normal growth, and potentially the clock may influence 40% of the genome. Undeniably, circadian control has a major impact on the biology of *Neurospora*: Much of metabolism is clock-controlled, and in broad outline, daytime activities involve catabolism, energy production, and assembly of precursors, whereas nighttime activities favor biosynthesis of cellular components and growth. Through use of a high-throughput assay using promoter-luc fusions, we show that rhythmic gene activation in and of itself is insufficient to dictate the transcript cycling at the mRNA level. This indicates that posttranscriptional regulation plays a major role in control of cycling transcripts and recommends further analysis of how such posttranscriptional regulators involved are themselves regulated (circadian expression, circadian activity, circadian binding, etc.). Finally, WCC binding alone is insufficient to set the phase of circadian transcription in *N. crassa*, indicating that additional factors are needed to determine the phase at which promoters of cycling genes are activated.

Materials and Methods

Strains. A protocol similar to that described (54) was used in all transformants. The *Neurospora* strain FGSC2489 was used for all analyses. mRNA was extracted for the circadian time course as previously described, with minor exceptions (21). RT-PCR was performed on the same mRNA extracted for use in the library construction as previously described, with minor changes (21) using primers (Dataset S12). Camera recordings and data analysis were as described (22, 25), with minor modifications (SI Text).

Library Construction and Mapping. Libraries were prepared as per the manufacturers instructions and libraries were analyzed using Illumina/Solexa sequencing (SI Text). The resulting raw and processed data was deposited to National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) under accession number SRP046458.

Data and FunCat Analysis. RPKM values were subjected to further normalization and each individual time course was subjected to JTK cycle analysis (18). Any gene with a false-discovery rate of $Q < 0.05$ was selected as a rhythmic gene. DREME analysis (34) was used to identify significantly enriched motifs (SI Text). Functional annotation of the differently expressed genes was performed according to Functional Categories (55) and *N. crassa* Genome Database (56). FunCat term and KEGG enrichment was performed using the FungiFun (28) and a value of $P < 0.05$ was used as a cutoff for significantly enriched terms.

ChIP-Seq. *Neurospora* tissue was cross-linked with 1% paraformaldehyde for 15 min and quenched with 0.125 M glycine for an additional 5 min for samples ranging from DD4 to DD32. ChIP was performed as described previously (38, 57) on 500 mg of tissue with WC-2 antibody with some modifications (58). Indexed sequencing libraries were prepared as described (59) (SI Text). Raw and processed data were submitted to the NCBI SRA under accession number SRP045821.

ACKNOWLEDGMENTS. We thank the Fungal Genetics Stock Center at the University of Missouri for *Neurospora* strains. This work is supported by NIH-National Institute of General Medical Sciences Grants GM083336 (to J.J.L.), GM34985 and GM068087 (to J.C.D.), and DE-AC02-05CH11231 (US Department of Energy).

1. Baker CL, Loros JJ, Dunlap JC (2012) The circadian clock of *Neurospora crassa*. *FEMS Microbiol Rev* 36(1):95–110.

2. Liu Y, Bell-Pedersen D (2006) Circadian rhythms in *Neurospora crassa* and other filamentous fungi. *Eukaryot Cell* 5(8):1184–1193.

3. Diernfellner AC, Schafmeier T (2011) Phosphorylations: Making the *Neurospora crassa* circadian clock tick. *FEBS Lett* 585(10):1461–1466.
4. Hughes ME, Grant GR, Paquin C, Qian J, Nitabach MN (2012) Deep sequencing the circadian and diurnal transcriptome of *Drosophila* brain. *Genome Res* 22(7):1266–1281.
5. Lowrey PL, Takahashi JS (2004) Mammalian circadian biology: Elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet* 5:407–441.
6. Rodriguez J, et al. (2013) Nascent-Seq analysis of *Drosophila* cycling gene expression. *Proc Natl Acad Sci USA* 110(4):E275–E284.
7. Vollmers C, et al. (2009) Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc Natl Acad Sci USA* 106(50):21453–21458.
8. Masri S, Sassone-Corsi P (2013) The circadian clock: A framework linking metabolism, epigenetics and neuronal function. *Nat Rev Neurosci* 14(1):69–75.
9. Nowrousian M, Duffield GE, Loros JJ, Dunlap JC (2003) The frequency gene is required for temperature-dependent regulation of many clock-controlled genes in *Neurospora crassa*. *Genetics* 164(3):923–933.
10. Dong W, et al. (2008) Systems biology of the clock in *Neurospora crassa*. *PLoS One* 3(8):e3105.
11. Zhu H, et al. (2001) Analysis of expressed sequence tags from two starvation, time-of-day-specific libraries of *Neurospora crassa* reveals novel clock-controlled genes. *Genetics* 157(3):1057–1065.
12. Correa A, et al. (2003) Multiple oscillators regulate circadian gene expression in *Neurospora*. *Proc Natl Acad Sci USA* 100(23):13597–13602.
13. Hughes ME, et al. (2009) Harmonics of circadian gene transcription in mammals. *PLoS Genet* 5(4):e1000442.
14. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9(1):e78644.
15. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63.
16. Menet JS, Rodriguez J, Abruzzi KC, Rosbash M (2012) Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife* 1:e00011.
17. Dunlap JC, Loros JJ (2004) The *Neurospora* circadian system. *J Biol Rhythms* 19(5):414–424.
18. Hughes ME, Hogenesch JB, Kornacker K (2010) JTK_CYCLE: An efficient non-parametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms* 25(5):372–380.
19. Bell-Pedersen D, Shinohara ML, Loros JJ, Dunlap JC (1996) Circadian clock-controlled genes isolated from *Neurospora crassa* are late night- to early morning-specific. *Proc Natl Acad Sci USA* 93(23):13096–13101.
20. Smith KM, et al. (2010) Transcription factors in light and circadian clock signaling networks revealed by genome-wide mapping of direct targets for *Neurospora* white collar complex. *Eukaryot Cell* 9(10):1549–1556.
21. Chen CH, Ringelberg CS, Gross RH, Dunlap JC, Loros JJ (2009) Genome-wide analysis of light-inducible responses reveals hierarchical light signalling in *Neurospora*. *EMBO J* 28(8):1029–1042.
22. Gooch VD, et al. (2008) Fully codon-optimized luciferase uncovers novel temperature characteristics of the *Neurospora* clock. *Eukaryot Cell* 7(1):28–37.
23. Bardiya N, Shiu PK (2007) Cyclosporin A-resistance based gene placement system for *Neurospora crassa*. *Fungal Genet Biol* 44(5):307–314.
24. Morgan LW, Greene AV, Bell-Pedersen D (2003) Circadian and light-induced expression of luciferase in *Neurospora crassa*. *Fungal Genet Biol* 38(3):327–332.
25. Larrondo LF, Loros JJ, Dunlap JC (2012) High-resolution spatiotemporal analysis of gene expression in real time: In vivo analysis of circadian rhythms in *Neurospora crassa* using a FREQUENCY-luciferase translational reporter. *Fungal Genet Biol* 49(9):681–683.
26. Gooch VD, et al. (2014) A kinetic study of the effects of light on circadian rhythmicity of the frq promoter of *Neurospora crassa*. *J Biol Rhythms* 29(1):38–48.
27. Lee K, Loros JJ, Dunlap JC (2000) Interconnected feedback loops in the *Neurospora* circadian system. *Science* 289(5476):107–110.
28. Priebe S, Linde J, Albrecht D, Guthke R, Brakhage AA (2011) FungiFun: A web-based application for functional categorization of fungal genes and proteins. *Fungal Genet Biol* 48(4):353–358.
29. Ruepp A, et al. (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res* 32(18):5539–5545.
30. Godard P, et al. (2007) Effect of 21 different nitrogen sources on global gene expression in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 27(8):3065–3086.
31. Shinohara ML, Correa A, Bell-Pedersen D, Dunlap JC, Loros JJ (2002) *Neurospora* clock-controlled gene 9 (ccg-9) encodes trehalose synthase: Circadian regulation of stress responses and development. *Eukaryot Cell* 1(1):33–43.
32. Bahn YS, et al. (2007) Sensing the environment: Lessons from fungi. *Nat Rev Microbiol* 5(1):57–69.
33. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28(1):27–30.
34. Bailey TL (2011) DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27(12):1653–1659.
35. Fuller KK, Hurley JM, Loros J, Dunlap J (2014) Photobiology and circadian clocks in *Neurospora*. *The Mycota*, ed Nowrousian M (Springer, Heidelberg), Vol XIII, pp 121–148.
36. Froehlich AC, Liu Y, Loros JJ, Dunlap JC (2002) White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science* 297(5582):815–819.
37. Froehlich AC, Loros JJ, Dunlap JC (2003) Rhythmic binding of a WHITE COLLAR-containing complex to the frequency promoter is inhibited by FREQUENCY. *Proc Natl Acad Sci USA* 100(10):5914–5919.
38. Belden WJ, Lewis ZA, Selker EU, Loros JJ, Dunlap JC (2011) CHD1 remodels chromatin and influences transient DNA methylation at the clock gene frequency. *PLoS Genet* 7(7):e1002166.
39. Akhtar CA, et al. (2002) Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr Biol* 12(7):540–550.
40. Duffield GE, et al. (2002) Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells. *Curr Biol* 12(7):551–557.
41. Harmer SL, et al. (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290(5499):2110–2113.
42. Rodriguez J, Menet JS, Rosbash M (2012) Nascent-seq indicates widespread cotranscriptional RNA editing in *Drosophila*. *Mol Cell* 47(1):27–37.
43. Storch KF, et al. (2002) Extensive and divergent circadian gene expression in liver and heart. *Nature* 417(6884):78–83.
44. Beckwith EJ, Yanovsky MJ (2014) Circadian regulation of gene expression: At the crossroads of transcriptional and post-transcriptional regulatory networks. *Curr Opin Genet Dev* 27:35–42.
45. Koike N, et al. (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338(6105):349–354.
46. Guo J, Cheng P, Yuan H, Liu Y (2009) The exosome regulates circadian gene expression in a posttranscriptional negative feedback loop. *Cell* 138(6):1236–1246.
47. Lidder P, Gutiérrez RA, Salomé PA, McClung CR, Green PJ (2005) Circadian control of messenger RNA stability. Association with a sequence-specific messenger RNA decay pathway. *Plant Physiol* 138(4):2374–2385.
48. Lal A, et al. (2004) Concurrent versus individual binding of HuR and AUF1 to common labile target mRNAs. *EMBO J* 23(15):3092–3102.
49. Abruzzi KC, et al. (2011) *Drosophila* CLOCK target gene characterization: Implications for circadian tissue-specific gene expression. *Genes Dev* 25(22):2374–2386.
50. Menet JS, Pescatore S, Rosbash M (2014) CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev* 28(1):8–13.
51. Wang B, Kettenbach AN, Gerber SA, Loros JJ, Dunlap JC (2014) *Neurospora* WC-1 recruits SWI/SNF to remodel frequency and initiate a circadian cycle. *PLoS Genet* 10(9):e1004599.
52. Joffe C, et al. (2013) The circadian clock coordinates ribosome biogenesis. *PLoS Biol* 11(1):e1001455.
53. Sancar G, Sancar C, Brunner M (2012) Metabolic compensation of the *Neurospora* clock by a glucose-dependent feedback of the circadian repressor CSP1 on the core oscillator. *Genes Dev* 26(21):2435–2442.
54. Colot HV, et al. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc Natl Acad Sci USA* 103(27):10352–10357.
55. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
56. Galagan JE, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422(6934):859–868.
57. Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12(16):1360–1367.
58. Denault DL, Loros JJ, Dunlap JC (2001) WC-2 mediates WC-1-FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. *EMBO J* 20(1-2):109–117.
59. Connolly LR, Smith KM, Freitag M (2013) The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. *PLoS Genet* 9(10):e1003916.
60. Stein LD, et al. (2002) The generic genome browser: A building block for a model organism system database. *Genome Res* 12(10):1599–1610.

Supporting Information

Hurley et al. 10.1073/pnas.1418963111

SI Text

SI Materials and Methods

Primer Design and Synthesis of Promoter Cassettes and Strains.

Primer3 was customized for the *Neurospora* genome by Peter Andrews, Institute for Quantitative Biomedical Sciences, Geisel School of Medicine, Hanover, NH, to retrieve regions adjacent to each start/stop codon of each gene, with the program searching around 1,500-bp region upstream (5') of the translational start codon and around 500-bp region downstream (3') of gene stop codon. For some genes, a gap range of no greater than 100 bp from start/stop codon was added in order for the selection of appropriate sequences. For 5' fragments bigger than 1,500 bp, the 5f (forward) primer was designed by hand. 5f+5r primers and 3f+3r primers PCR synthesized 5' and 3' flanks, respectively, using WT gDNA as template. Flanking primers had additional sequences for 1,000 bp of cyclophilin locus (5f and 3r) and the luciferase (5r and 3f) cassette (1). All *Neurospora* strains were tested on a luminometer (Turner TD-20e) for signal higher than 0.5, indicating luminescence.

Time Course Preparation. Tissue was prepared for rhythmic time-series analysis as previously described (2). Total RNA was extracted by adding 1 mL of TRIZOL reagent (Invitrogen) to the frozen tissue and then pulverizing it using a homogenizer (TissueLyzer; Qiagen) containing stainless beads. The tissue was separated from the supernatant by rapid centrifugation (16,000 × g for 15 min). The RNA was further purified and DNase treated using the RNeasy Mini Kit from Qiagen. The RNA was quantified using the Quant-It RiboGreen Kit from Invitrogen according to the manufacturer's protocol.

Library Construction for RNA-Seq. Plate-based sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system, where purified mRNA was converted into cDNA library templates of known strand origin for sequencing on the Illumina Sequencer platform. The first step was the purification of mRNA in the total RNA sample using Invitrogen's DynaBeads mRNA purification kit, which uses magnetic beads with bound poly-T oligos to select and purify poly-A-containing mRNA molecules. Following purification, the mRNA was chemically fragmented with 10× fragmentation solution (Ambion) at 70 °C for 3 min to generate fragments ranging in size between 250 and 300 bp. The fragmented RNA was then purified using AMPure SPRI beads (Agencourt) using a ratio of 160:100 beads volume to RNA. The RNA fragments were then synthesized into first-strand cDNA using SuperScript II Reverse Transcriptase (Invitrogen) and random hexamer primers (MBI Fermentas) with Actinomycin D in the master-mix to further ensure strand specificity by preventing spurious DNA dependent synthesis during the first-strand cDNA synthesis. First-strand synthesis thermocycler conditions were incubation at 42 °C for 50 min and then inactivation at 70 °C for 10 min. This was followed by another purification using AMPure SPRI beads at a ratio of 140/100 beads volume to cDNA. Next, second-strand cDNA was synthesized using DNA Polymerase I (Invitrogen), RNase H (Invitrogen), and a nucleotide mix containing dUTP (Roche) that incorporates dUTP in place of dTTP, which ensures strand specificity, in the second strand to generate ds cDNA. Second-strand synthesis thermocycler conditions were incubation at 16 °C for 60 min. The ds cDNA fragments were purified using AMPure SPRI beads at a ratio of 75/100 beads volume to cDNA and a second purification using a ratio of 140/100. Using a DNA sample prep kit designed

for Illumina sequencing, the purified ds cDNA fragments were then end-repaired to be blunt-ended, phosphorylated for the addition of a single A base tail, and finally ligation of the Illumina adapters, which includes unique barcode sequences that allow for the multiplexing of the sample libraries in sequencing. To ensure the strand specificity of the cDNA library templates, AmpErase uracil *N*-glycosylase (UNG) (Applied Biosystems) was added to the ds cDNA library fragments and they were incubated at 37 °C for 15 min to cleave and degrade the strand containing dUTP. The ss cDNA was then enriched using 10 cycles of PCR with Illumina TruSeq primers and purified using AMPure SPRI beads at a ratio of 90/100 beads volume to ds cDNA to create the final cDNA library ready for sequencing. The libraries were created in plate format (rather than tube) using a robot. Once libraries were created then they were quantified, pooled, and run on the Hi-seq instrument. The amount of starting RNA material used for each library was 5 µg. Sequencing libraries were diluted to the specifications required by the Illumina/Solexa sequencing machine in use. Manuals and protocols that describe these steps are available from Illumina/Solexa. The *N. crassa* RNA-Seq data are displayed on JGI MycoCosm portal [genome.jgi.doe.gov/Ncrasss (3)] and have been deposited to NCBI SRA under accession number SRP046458.

Data Analysis. Reads from each RNA-Seq sample were aligned to the reference transcriptome using BWA (parameters used: seed length = 25; maximum hits = 1). The reads were first trimmed to 36 bp, to ensure that reads with a low quality 3' end will still align. Finally, using the alignment, the number of reads that aligned to each gene was counted, for each sample. Only uniquely mapping reads were counted. The counts were then normalized using reads per kilobase of exon model per million mapped reads (RPKM) normalization. This normalizes both by the number of mapped reads for each library, and by the length for each transcript in the reference file. In other words:

$$\text{RPKM} = \text{raw_count} * (1,000/\text{gene_length}) \\ * (1,000,000/\text{num_mapped_reads}).$$

The data were subjected to quantile normalization to force all expression profiles to have the exactly same distribution and were then log transformed to eliminate outlying points. To generate a list of rhythmically expressed genes, after quantile normalization and log transformation, time points that did not pass quality control were replaced as follows: Time Course 1—DD30 was replaced by DD30 from Time Course 2; Time Course 2—DD28 was replaced by DD28 from Time Course 1; Time Course 3—DD44, DD46, and DD48 were replaced by DD44, DD46, and DD48 from Time Course 2. The data were then smoothed by the following formula: replace expression at t by averaging $t - 1, t, t + 1$; the exceptions were DD2 (averaging DD2, DD4, and DD44); DD48 (averaging DD46, DD48, and DD6) to correct for the CT time. After the data were processed, each individual time course was subjected to JTK cycle analysis (4). Any gene with a false-discovery rate of $Q < 0.05$ was selected as a rhythmic gene and Venn diagram analysis was used to determine which genes were rhythmic in at least two of the three datasets. To generate heat maps, the three datasets were averaged, and then this single time course was smoothed by the above method and subjected to JTK cycle analysis (4).

DREME analysis (5) was used to identify motifs that were significantly enriched in the promoters of the rhythmic genes compared with the rest of the genome. Briefly, analysis of 1,000 bp of the

promoters of each of the 872 genes identified 36 potential motifs. To test for significant enrichment of these motifs in our rhythmic gene set, we first used a Find Individual Motif Occurrences (FIMO) algorithm (6) to find genes in the entire genome that contained these motifs in their promoters. Typically an individual motif was found in several hundred promoters. Then, at each CT, the number of genes from the rhythmic dataset peaking at that time ± 2 h, and that contained a particular motif, was mapped on circular scale. A motif was considered enriched when the number of genes from among the 872, both peaking within a 4-h window and containing the motif, was significantly greater than the number that would be expected from a random distribution of the motif-containing genes in the entire genome. Enrichment of rhythmic genes whose promoters contain a specific motif peaking at an explicit time compared with all of the genes whose promoters contain that motif was examined and a position weight motifs (PWM) probability value was assigned. For every motif, the enrichment was examined at all 24 time points. Then PWMs with minimum P values of <0.001 (out of 24 time points) were selected and a plot was drawn to show the significance of enrichment at each time point.

Real-Time CCD Recording. Liquid nitrogen or electronically cooled cameras from Roper Scientific (VersArray 1300 B/LN and PIXIS 1020) were used to follow luminescence. Camera runs were done at 25 °C with cultures grown in 96-well plates using black opaque tubes. Liquid suspensions of conidia were plated onto race tube (RT) medium (0.1% glucose, 0.17% arginine), maltose medium (0.1% maltose, 0.17% arginine), and/or QA medium (0.03% glucose, 0.05% arginine, 0.01 M QA, pH 5.75) containing 25 μ M luciferin (Dataset S3) and covered with a Breathe-Easy strip (USA Scientific). Plates were then subjected to 48 h of 12:12 dark:light cycle conditions at 25 °C before transferring the plates to constant-dark conditions in the camera setup. Signals were accumulated for either 5 or 10 min every hour, depending on the intensity of the luciferase signal. A customized Excel Macro to process the images was developed (7). Raw data arising from each time series were then processed using custom-written software to detrend the data and then to normalize amplitudes.

Validation of Gene Expression by Real-Time PCR. A total of 1,500 ng from each sample of the RNA extracted as described above was used to prepare cDNA using the SuperScript III First-Strand synthesis kit (Invitrogen) followed by real-time PCR using the Fast SYBR green master mix kit in an ABI 7500 real-time cyler (Applied Biosystems). Primer sequences for the various genes tested were generated using an in-house primer designing soft-

ware. The primer combinations used are listed in Dataset S12, and the final concentration of primers in the reaction mix was 0.5 μ M. The following cycling parameters were used: step 1: 95 °C for 5 min; and step 2: 95 °C for 10 s, 60 °C for 30 s for 40 cycles. Ct values were calculated using software provided by the instrument manufacturer. *rac-1* gene was used as a control noncycling gene for the determination of relative mRNA levels for all genes tested. The relative mRNA levels for each time point were calculated using at least two out of three biological replicates in the case where one of the replicates differed from either of the other two by more than threefold.

ChIP Assay to Identify WC-2 Binding Sites. Conidia from the wild-type strain (OR74A) was germinated and grown in 200 mL of Bird medium at 25 °C in the light, before appropriately timed transfer to dark to get samples ranging from DD4 to DD32—representing a 32-h time course with a 4-h resolution. The *Neurospora* mycelium was cross-linked in the light/dark with 1% para-formaldehyde for 15 min and quenched with 0.125 M glycine for an additional 5 min. ChIP was performed as described previously (8, 9) on 500 mg of tissue with WC-2 antibody with some modifications (10). To improve the sonication, the tissue was probe-sonicated for 8 s at 30% amplitude for three times, and then sonicated with a Biorupter for four times with 30-s on/off cycles for 5 min at high power to get an average chromatin size of 100–500 bp. The sheared chromatin was obtained by centrifuging at 11,000 $\times g$ for 5 min and incubated with protein A dynabeads and WC-2 antibody overnight at 4 °C with constant rotation. The beads were washed once with wash buffer 1 (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 150 mM NaCl), once with wash buffer 2 (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.1% Triton-X-100, 0.1% SDS, 500 mM NaCl), once with wash buffer 3 (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1% Nonidet P-40, 1% sodium deoxycholate, 0.25 M LiCl), once with wash buffer 4 (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and eluted with 0.1 M sodium bicarbonate and 1% SDS, and the cross-link was reversed by incubation at 65 °C for 8 h. The DNA was recovered by Qiagen PCR purification kit. To verify the quality of the ChIP time course, RT-PCR was performed with the purified DNA using the following primers: *frq*-PLRE-F, 5'-atacttgtaggccgctccccatc-3'; *frq*-PLRE-R, 5'-gctgtcatctctcagcattttgtc-3'; *frq*-CBOX-F, 5'-caattttgcagcgtcatcggctc-3'; and *frq*-CBOX-R, 5'-tcaagtcagctcgtaccacatc-3' using the cycling conditions described before relative enrichment of pLRE and C-box of the *frq* promoter region in the samples was determined as a percentage of input.

1. Goch VD, et al. (2008) Fully codon-optimized luciferase uncovers novel temperature characteristics of the *Neurospora* clock. *Eukaryot Cell* 7(1):28–37.
2. Loros JJ, Denome SA, Dunlap JC (1989) Molecular cloning of genes under control of the circadian clock in *Neurospora*. *Science* 243(4889):385–388.
3. Grigoriev IV, et al. (2014) MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res* 42(Database issue):D699–D704.
4. Hughes ME, Hogenesch JB, Kornacker K (2010) JTK_CYCLE: An efficient non-parametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms* 25(5):372–380.
5. Bailey TL (2011) DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27(12):1653–1659.
6. Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.
7. Larrondo LF, Loros JJ, Dunlap JC (2012) High-resolution spatiotemporal analysis of gene expression in real time: In vivo analysis of circadian rhythms in *Neurospora crassa* using a FREQUENCY-luciferase translational reporter. *Fungal Genet Biol* 49(9):681–683.
8. Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12(16):1360–1367.
9. Belden WJ, Lewis ZA, Selker EU, Loros JJ, Dunlap JC (2011) CHD1 remodels chromatin and influences transient DNA methylation at the clock gene frequency. *PLoS Genet* 7(7):e1002166.
10. Denault DL, Loros JJ, Dunlap JC (2001) WC-2 mediates WC-1-FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. *EMBO J* 20(1-2):109–117.

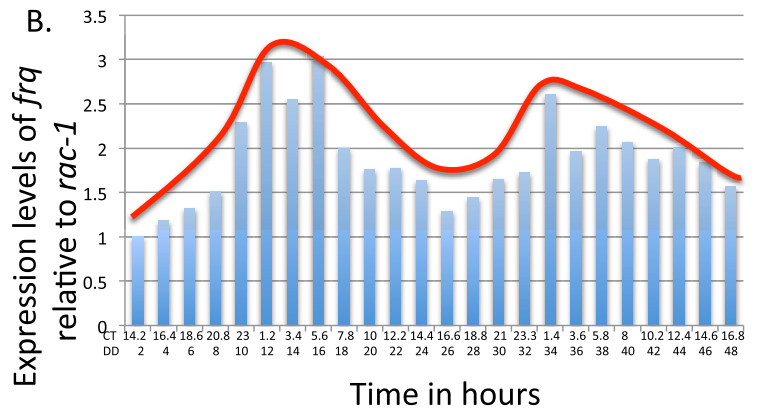
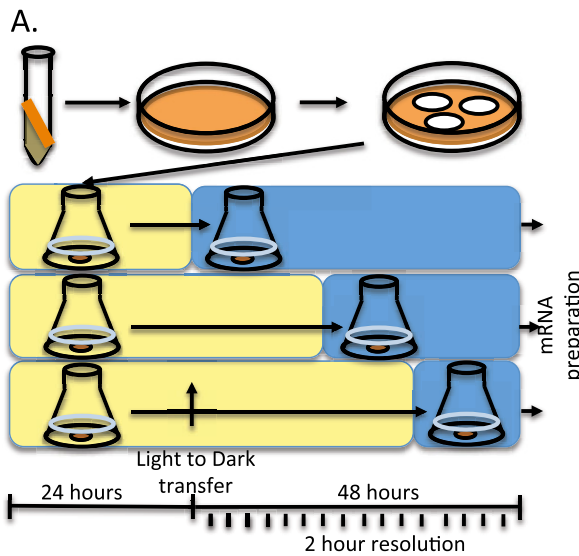


Fig. S1. *Neurospora* time course displays circadian rhythms. (A) A diagrammatic explanation of the harvest protocol used to obtain the three time courses. Briefly, conidia were inoculated into Bird medium, grown 24 h, and then discs were cut from the resultant tissue, transferred to individual flasks, and kept in light and dark for times appropriate to yield a 48-h time course. The age of all cultures was the same ± 4 h. (B) Representative RT-PCR of *frq* mRNA normalized to *rac-1* from one of the three circadian time courses, described with both CT and DD time points.

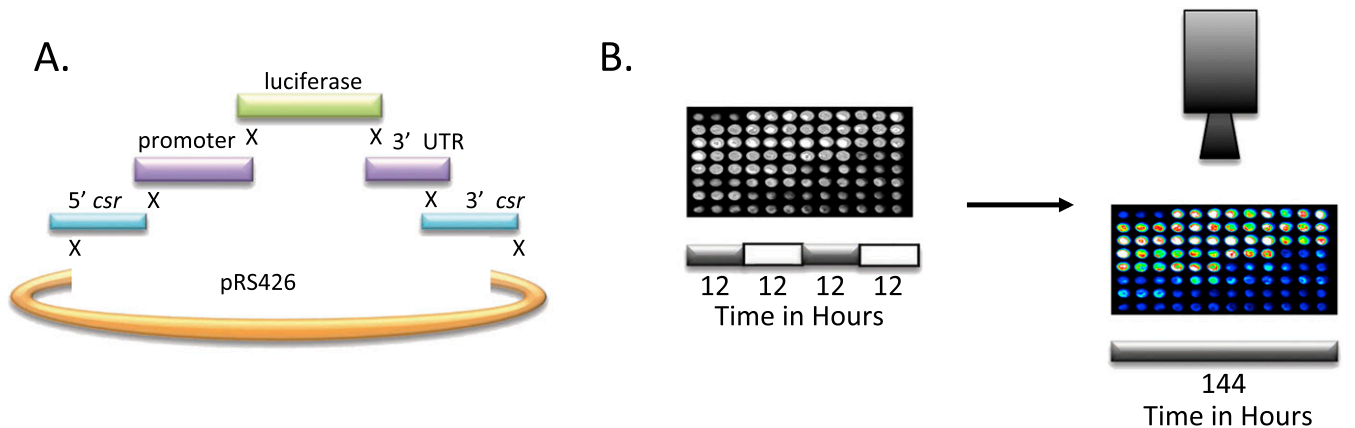


Fig. S2. Setup of strains for luciferase imaging. (A) Schematic of the method used to construct the luciferase reporters used to track the promoter activity oscillations. (B) Description of the data acquisition of each strain using a 96-well format.

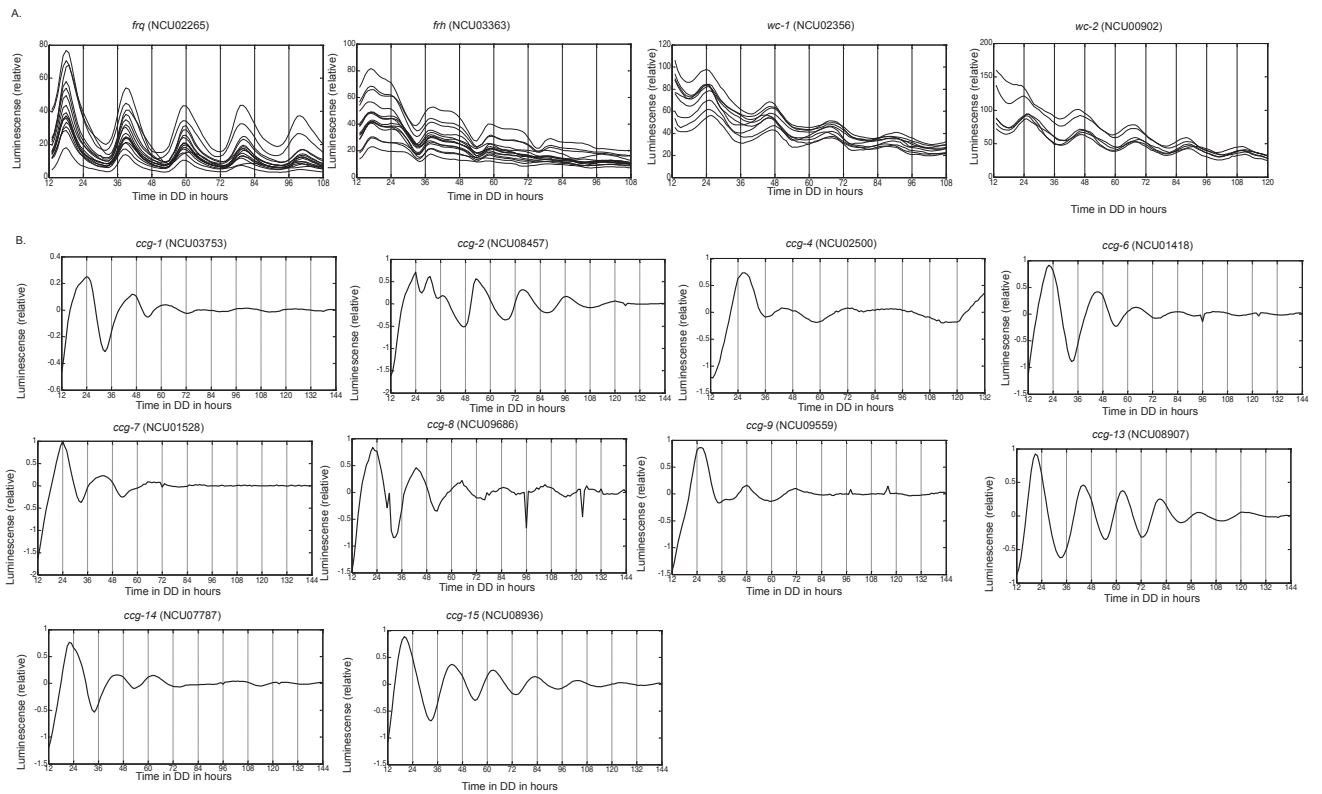


Fig. S3. Luciferase traces of *ccgs* and core clock genes. (A) Luciferase traces for three technical replicates for each of three biological replicates of the four core clock genes without detrending and rescaling. (B) Luciferase traces acquired for all of the published *ccgs* after detrending and rescaling.

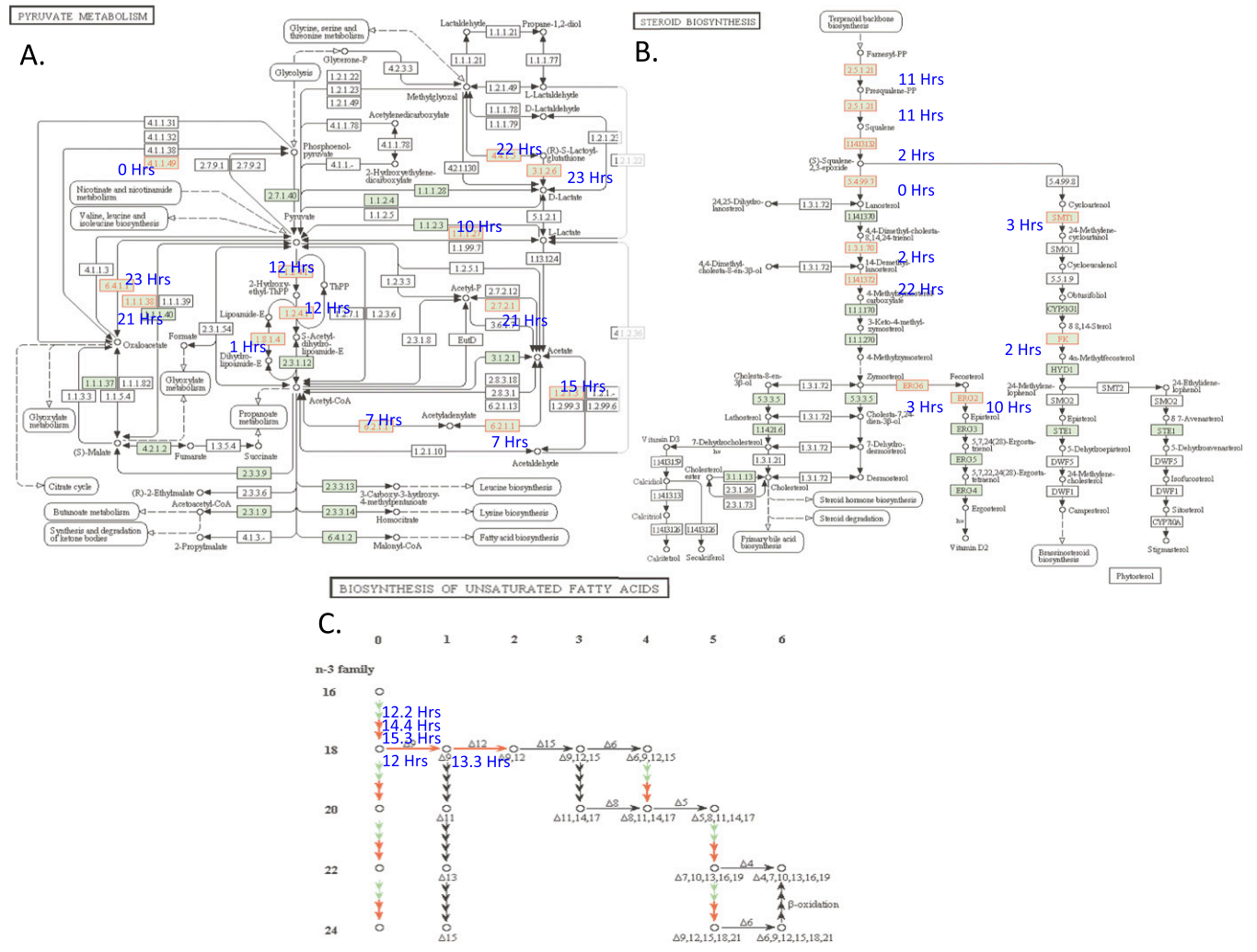


Fig. S5. Identification of rhythmic genes in the KEGG pathway. (A–C) Diagrammatic representation of the KEGG pathways for, pyruvate metabolism, steroid biosynthesis, and biosynthesis of unsaturated fatty acids [modified from the Kyoto Encyclopedia of Genes and Genomes website (1)]. Each gene identified as rhythmic in our analysis and playing a role in the pathway has been highlighted in red in the pathway map and its peak expression time is noted in blue next to it.

1. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28(1):27–30.

Gene	Rhythmic by RNA-Seq	Rhythmic by RT-PCR	RNA-Seq correlates with RT-PCR	Rhythmic by Luc	RNA-Seq correlates with luciferase
NCU00285	+	+	✓	+	✓
NCU00289	-	-	✓	+	X
NCU00499	-	-	✓	+	X
NCU00582	-	-	✓	+	X
NCU00855	-	-	✓	+	X
NCU01528 (<i>ccg-7</i>)	-	-	✓	+	X
NCU01810	-	-	✓	+	X
NCU01898	-	-	✓	+	X
NCU02084	-	-	✓	-	✓
NCU02193	-	-	✓	+	X
NCU02238	-	-	✓	+	X
NCU02265	+	+	✓	+	✓
NCU03362	-	-	✓	+	X
NCU03753 (<i>ccg-1</i>)	+	+	✓	-	X
NCU04041	-	-	✓	-	✓
NCU04923	+	-	X	+	✓
NCU06799	+	-	X	+	✓
NCU06961	+	+	✓	+	✓
NCU07139	-	-	✓	+	X
NCU07787 (<i>ccg-14</i>)	-	-	✓	-	✓
NCU07817	+	+	✓	+	✓
NCU08141	-	-	✓	+	X
NCU08457 (<i>ccg-2</i>)	+	+	✓	-	X
NCU08769	+	+	✓	+	✓
NCU08791	-	-	✓	+	X
NCU08907 (<i>ccg-13</i>)	+	+	✓	+	✓
NCU09559 (<i>ccg-9</i>)	+	+	✓	-	X
NCU09686 (<i>ccg-8</i>)	-	-	✓	-	✓
NCU09976	+	+	✓	+	✓

Red	Rhythmic in both
Green	Rhythmic in RNA-Seq
Blue	Rhythmic in Luc only
White	Rhythmic in neither

Fig. S6. Weak correlation between RNA-Seq and promoter activation. A list of all of the genes investigated by RNA-Seq, luciferase, and RT-PCR. Genes in red are rhythmic by RNA-Seq and luciferase, in green by RNA-Seq only, and in blue by Luc only and white by neither method.

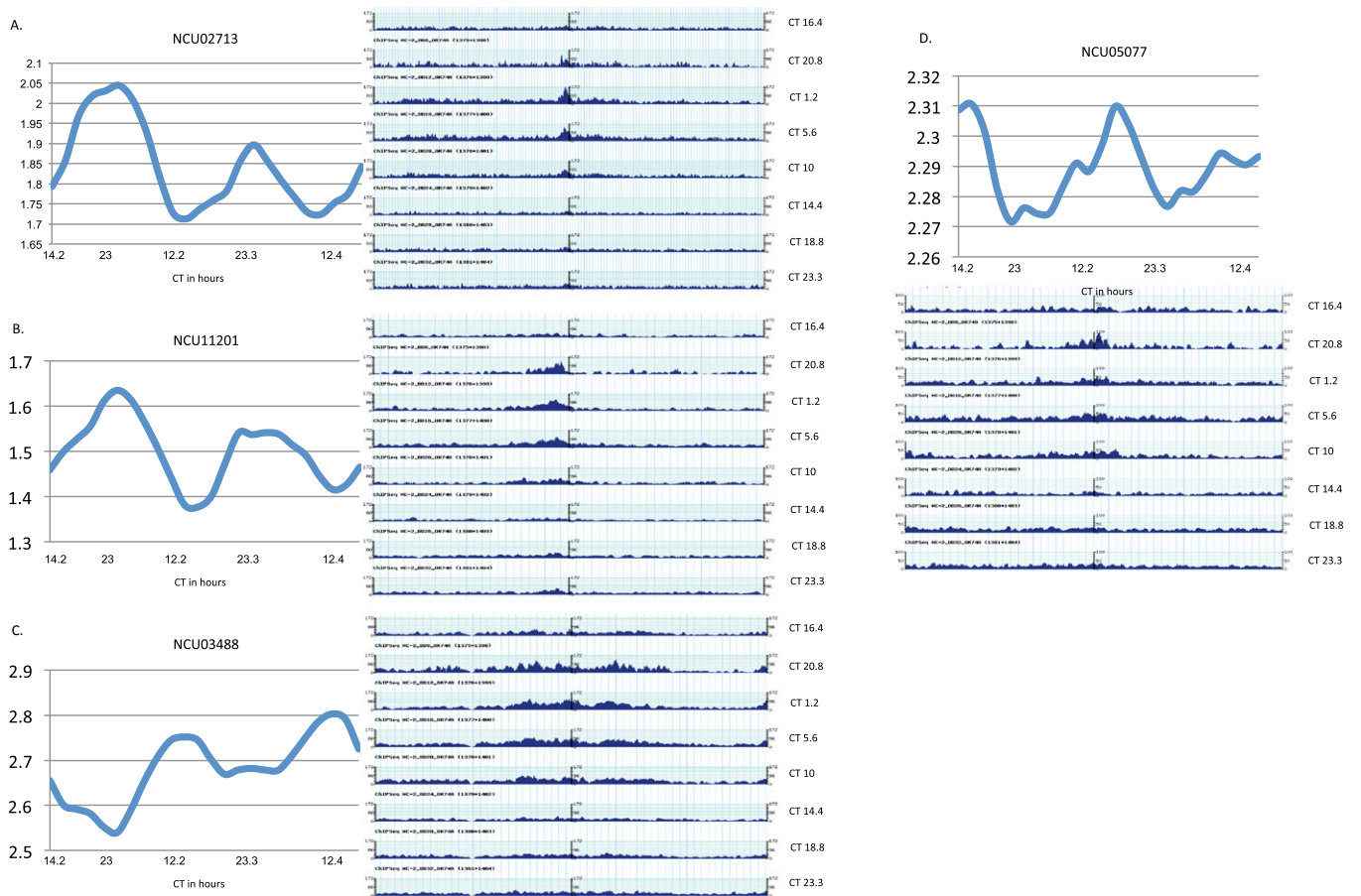


Fig. S7. Comparison of mRNA levels to ChIP-Seq data. (A–D) A comparison between the RNA-Seq data and ChIP-Seq data for four genes identified from each dataset. A and B show peaks in mRNA levels that correspond to the peak in WC-2 binding, whereas C and D demonstrate that the WC-2 binding peak does not always correlate with the peak of mRNA levels.

Dataset S1. Genes identified as rhythmic by JTK analysis

[Dataset S1](#)

Dataset S2. Genes identified as rhythmic by luciferase trace

[Dataset S2](#)

Dataset S3. Media used in luciferase assays

[Dataset S3](#)

Dataset S4. Genes identified as rhythmic by all methods

[Dataset S4](#)

Dataset S5. FunCat term category *P* values

[Dataset S5](#)

Dataset S6. Breakdown of each gene and the significance in each FunCat category

[Dataset S6](#)

Dataset S7. Genes in significantly enriched FunCat term categories for RNA-Seq and luciferase rhythmic genes

[Dataset S7](#)

Dataset S8. Rhythmic genes in each time-of-day-specific circadian category

[Dataset S8](#)

Dataset S9. Genes in significantly enriched FunCat term categories for DREME motifs

[Dataset S9](#)

Dataset S10. Genes that differ between the RNA-Seq and luciferase data

[Dataset S10](#)

Dataset S11. Genes with peaks identified in WC-2 circadian chip

[Dataset S11](#)

Dataset S12. Primers used in real-time PCR

[Dataset S12](#)