High-resolution Time Course Analysis of Gene Expression from Pituitary

M. HUGHES,*† L. DEHARO,‡† S.R. PULIVARTHY,‡ J. GU,* K. HAYES,* S. PANDA,‡ AND J.B. HOGENESCH*

*Department of Pharmacology, Institute for Translational Medicine and Therapeutics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104; ‡Regulatory Biology, Salk Institute for Biological Studies, La Jolla, California 92186

In both the suprachiasmatic nucleus (SCN) and peripheral tissues, the circadian oscillator drives rhythmic transcription of downstream target genes. Recently, a number of studies have used DNA microarrays to systematically identify oscillating transcripts in plants, fruit flies, rats, and mice. These studies have identified several dozen to many hundred rhythmically expressed genes by sampling tissues every 4 hours for 1, 2, or more days. To extend this work, we have performed DNA microarray analysis on RNA derived from the mouse pituitary sampled every hour for 2 days. COSOPT and Fisher’s G-test were used at a false-discovery rate of less than 5% to identify more than 250 genes in the pituitary that oscillate with a 24-hour period length. We found that increasing the frequency of sampling across the circadian day dramatically increased the statistical power of both COSOPT and Fisher’s G-test, resulting in considerably more high-confidence identifications of rhythmic transcripts than previously described. Finally, to extend the utility of these data sets, a Web-based resource has been constructed (at http://wasabi.itmat.upenn.edu/circa/mouse) that is freely available to the research community.

INTRODUCTION

The central circadian oscillator in mammals is located in a small cluster of neurons in the hypothalamus called the suprachiasmatic nucleus (SCN) (Stratmann and Schibler 2006). These neurons have an endogenous circadian clock that oscillates with a period of approximately 24 hours (Ko and Takahashi 2006). In addition, these neurons receive information from retinal ganglion cells, permitting them to entrain the clock to environmental light cues (Cermakian and Sassone-Corsi 2002). Peripheral tissues have endogenous circadian clocks as well and may be entrained by a number of different stimuli (for review, see Stratmann and Schibler 2006). However, these peripheral oscillators are subordinate to the central circadian clock in the SCN, which is responsible for integrating environmental input and orchestrating the biological rhythms of the entire organism (Schibler and Sassone-Corsi 2002).

Consequently, many complex physiologies in an organism show regular oscillations during the course of a single day. For example, rhythms of sleep and arousal are controlled by the circadian clock, as well as rhythms in food consumption, blood pressure, body temperature, and metabolism (for review, see Hastings et al. 2003; Curtis and Fitzgerald 2006). These rhythms have direct and indirect implications for human health. The metabolism and efficacy of drugs depend heavily on the time of day they are administered (Lis et al. 2003; Antoch et al. 2005; Halberg et al. 2006). Moreover, disruptions of the circadian clock have been shown to increase susceptibility to cancer, heart disease, and metabolic disorders while directly causing serious sleep disorders and influencing mental illness (Klerman 2005; Curtis and Fitzgerald 2006; Halberg et al. 2006; Levi and Schibler 2007).

Despite considerable effort, the link between the molecular oscillations of the core circadian clock and rhythms of organismal physiology remains poorly understood (Hastings et al. 2003). As a first step toward bridging this gap, many groups, including ours, have used DNA microarrays to systematically identify genes that oscillate during the course of a signal day (for examples of this work, see Harmer et al. 2000; Claridge-Chang et al. 2001; McDonald and Rosbash 2001; Akhtar et al. 2002; Ceriani et al. 2002; Duffield et al. 2002; Lin et al. 2002; Panda et al. 2002; Storch et al. 2002; Ueda et al. 2002; Kornmann et al. 2007). Typically, these studies have analyzed RNA samples isolated every 4 hours over the course of 2 or more days and used curve-fitting algorithms, Fourier analysis, or autocorrelation tests to identify rhythmic transcripts (Table 1). Using these methodologies, several hundred genes have been found under circadian regulations in mouse tissues and rat fibroblasts, as well as in plants (Arabidopsis) and fruit flies (Drosophila) (Table 1). These data and analyses

| Table 1. Previous Circadian Microarray Studies Have Predominately Used 4-hour Time Resolution |
|---|---|---|
| Study resolution | Year | Tissue | Time (hour) |
| Harmer et al. | 2000 | Arabidopsis | 4 |
| Claridge-Chang et al. | 2001 | fly heads | 4 |
| McDonald and Rosbash | 2001 | fly heads | 4 |
| Ceriani et al. | 2002 | fly heads and bodies | 4 |
| Lin et al. | 2002 | fly heads | 4 |
| Akhtar et al. | 2002 | mouse liver | 4 |
| Storch et al. | 2002 | mouse liver and heart | 4 |
| Panda et al. | 2002 | mouse SCN and liver | 4 |
| Ueda et al. | 2002 | mouse SCN and liver | 4 |
| Duffield et al. | 2002 | Rat-1 fibroblasts | 4 |
| Kornmann et al. | 2007 | mouse liver | 4 |

†These authors contributed equally to this work.
represented an important first step in elucidating the link between molecular oscillations of the core circadian clock and downstream physiology.

Several controversies, however, have emerged as a consequence of this work (Etter and Ramaswami 2002; Duffield 2003; Hayes et al. 2005). For example, most transcriptional profiling studies were able to identify many of the core components of the circadian clock, although these studies missed a number of bona fide oscillating genes and clock components (Harmer et al. 2000; Ceriani et al. 2002; Panda et al. 2002). This suggests that the conventional design of these experiments may have been insufficiently powerful to identify every significant circadian gene, as well as illustrated a traditional weakness of chip-based analysis: sensitivity. Similarly, there was surprisingly small overlap between the rhythmically expressed genes identified in different studies (Etter and Ramaswami 2002). In part, this discrepancy can be explained by the use of different experimental designs, strains of organisms, array platforms, and condensation algorithms, as well as the analytic methods used to detect cycling genes (Walker and Hogenesch 2005). These examples illustrate the high false-positive and high false-negative rates in these original studies, and taken as a whole, they suggest that additional work is necessary to precisely define the subset of the transcriptome regulated by the circadian clock.

TRANSCRIPTIONAL PROFILING AT HIGH TEMPORAL RESOLUTION

To address this issue, a second-generation microarray analysis was performed to identify genes whose transcripts are under circadian regulation. Wild-type mice were entrained to a 12 hours light:12 hours dark (12:12 L:D) schedule before release into constant darkness conditions. Tissue samples from the pituitary of four mice were collected every hour for 2 full days, and RNA samples were analyzed using Affymetrix microarrays. To identify rhythmically expressed genes, these data were subsequently analyzed using COSOPT, a method that relies on curve fitting and permutation analysis, and Fisher’s G-test, based on the Fourier transform. Following the initial analyses, the false-discovery rate (FDR) was calculated for all transcripts using both methods in R, a high-level programming language used for statistical computing.

Using this analysis, we found considerable evidence for robust circadian rhythms (Fig. 1). These included confidently detecting the circadian oscillations of known core clock transcripts. For example, two components of the core oscillator, Bmal1 and Clock (Ko and Takahashi 2006), as well as many other components of the circadian clock, including Rev-erbα, Rorα, Per2, Cry1, and Cry2 were among the genes with the lowest p values by both statistical tests (Fig. 1 and data not shown). Importantly, the phase relationship among these components in both the liver and pituitary was in excellent agreement with previous expectations and was resolved to a preciseness that was not possible in previous studies.

Prompted by earlier work on meta-analysis of circadian gene expression data from DNA arrays, we sought to identify a cohort of oscillating transcripts using two distinct statistical methods: COSOPT and Fisher’s G-test. Each identified several hundred rhythmic transcripts in pituitary samples at highly restrictive FDRs (Table 2). Importantly, there was considerable overlap between the transcripts identified by both tests, increasing the confidence that these genes are bona fide outputs of the circadian clock. All told, 274 transcripts were identified as rhythmically expressed by both COSOPT and Fisher’s G-test at an FDR of 5%; when the FDR is reduced to 1%, 120 genes were found to be rhythmic by both algorithms (Table 2). As expected, the vast majority of these genes show period lengths of approximately 24 hours (Fig. 2).

Table 2. Cycling transcripts in Pituitary Identified at Different False-discovery Rates

<table>
<thead>
<tr>
<th>FDR</th>
<th>COSOPT</th>
<th>Fisher’s G-test</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.05</td>
<td>334</td>
<td>1152</td>
<td>274</td>
</tr>
<tr>
<td>&lt;0.01</td>
<td>131</td>
<td>316</td>
<td>120</td>
</tr>
</tbody>
</table>

Figure 1. Cycling transcripts detected by microarray analysis in the pituitary. Tissue samples from the pituitary were collected from wild-type mice every hour for 2 days. RNA was purified from these samples and profiled on Affymetrix microarrays. Bmal1 (left) and Per2 (right) show robust circadian oscillations through the entire time course.

Figure 2. COSOPT and Fisher’s G-test predominantly detect rhythmic transcripts with periods of approximately 24 hours. In the pituitary, several hundred rhythmic transcripts were detected at a false-discovery rate of <0.05.
Taken as a whole, these data suggest that more than 100 genes in the mouse pituitary oscillate in a circadian manner. Interestingly, the phases of these circadian genes show an approximately even distribution throughout a single day (Fig. 3), indicating that the output of the clock is operating nearly equally at every hour in these tissues.

**STATISTICAL ANALYSIS**

The design of the original studies by our group and others was somewhat arbitrarily derived from northern and western blot analysis studies predominantly aimed at the study of one or a few genes. The inherent weakness in this approach for DNA array analysis is the multiple testing problem; instead of one or a few genes being analyzed, tens of thousands were analyzed, which greatly amplified the likelihood of both false-positive and false-negative errors (type I and type II errors). The availability of the high-resolution time course experiment enabled empirical statistical analyses to determine the optimal time frequency of sampling. By analyzing expression data from 48 time points across 2 complete days, the present study successfully identified several hundred rhythmically expressed genes in the mouse pituitary (Table 2). These data indicate that the power of statistical tests to determine rhythmicity is dramatically improved through modest increases in sampling frequency. One of the key questions was the number of time points in a 48-hour time course necessary to robustly detect circadian transcription. In an era of tight research funding and relatively expensive DNA arrays, the best balance between sensitivity, selectivity, and cost would provide a tremendous benefit in the detection of cycling genes.

To evaluate this issue, we performed a number of simulations to assess the success of COSOPT and Fisher’s G-test in identifying rhythmic transcripts at different sampling densities. To this end, we randomly selected time points from our data set at 1-, 2-, 3-, and 4-hour intervals and employed COSOPT and Fisher’s G-test to identify oscillating genes. We discovered that there is a considerable statistical advantage in increasing the frequency of sampling (Fig. 4). At a wide range of FDR (0.01 to 0.4), both COSOPT and Fisher’s G-test show approximately logarithmic improvements in detecting rhythmic transcripts as sampling density increases (Fig. 4). At low sampling resolutions (less than or equal to every 3 hours), Fisher’s G-test performs considerably better than COSOPT, particularly at relatively high FDR (compare graphs in Fig. 4). Interestingly, both algorithms perform similar efficacies at high sampling resolutions, which is consistent with our expectation that sufficiently high-quality data sets render their downstream analysis insensitive to the particular algorithm used. These data indicate that the detection of transcripts under circadian control using conventional algorithms with a low rate of false discovery necessitates a sampling density of at least once every 2 hours over 2 full days and best balances cost and data quality.

**BIOINFORMATICS RESOURCES**

Several software resources were developed in the course this work and are publicly available to facilitate circadian gene expression analysis. COSOPT requires an input data step that is transcript-centric, meaning that the number of input files rises linearly with the number of genes being analyzed. To streamline this process, we developed a Perl script that automatically converts microarray data (in *.txt or *.csv format) into the files necessary for COSOPT. In addition, we have written an R script that utilizes the GeneTS package, calculates the average periodogram of the data, and finds period length of transcription.
The identification of novel circadian genes in the pituitary will enable further studies on the molecular mechanism of peripheral oscillators as well as additional work on the output of the circadian clock. To facilitate the distribution of this data set, we have created a Web-based interface available at http://wasabi.itmat.upenn.edu/circa/mouse (Fig. 5). This interface, affectionately known as “Wasabi,” is written in Ruby on Rails, implemented on a Linux server, and permits the user to search for the transcriptional profile of any gene in either the liver or pituitary (Fig. 5, top). At the same time, $q$ values (representing the false-discovery rate) and $p$ values (representing the probability that a transcript was identified as rhythmic by chance alone) from COSOPT and Fisher’s G-test, as well as phase and period length, can be used to filter queries (Fig. 5, top).

Figure 5 shows an example of a query submitted to Wasabi. The user is searching for the transcriptional profile of $Per2$ in the liver, filtered by COSOPT $q$ value ($<0.05$) and period length ($>22$ hours). The resulting profile is shown in Figure 5 (bottom left); the transcriptional profile of $Per2$ is plotted as intensity (expression level) versus time. Additionally, information on the statistical analysis of this transcript, as well as external Web links to information on $Per2$, and its genomic locus are available within this interface (Fig. 5, bottom right). By streamlining the storage and retrieval of gene expression data, we anticipate Wasabi will become an important resource for the circadian community, analogous to SymAtlas, a Web-based resource describing the tissue-specific expression of most mammalian genes (Su et al. 2004).

ANALYSIS OF CYCLING TRANSCRIPTS

As previously discussed, a considerable gap exists in our understanding between the molecular mechanisms of...
the circadian clock and its role in regulating rhythmic physiology (Etter and Ramaswami 2002; Duffield 2003; Hastings et al. 2003). High-throughput gene expression analysis enabled by DNA arrays and informatics tools can help to bridge this gap. In its simplest form, this analysis can consist of identifying a rate-limiting enzyme in a biochemical pathway as a target of the clock, inferring that the pathway is clock-regulated, and testing the hypothesis. Even more power can be generated by analyzing groups (rather than single) of transcripts for their coherent action in cellular pathways. For example, we can use informatics technologies to infer the physiological function of subsets of rhythmically expressed genes. We have used Ingenuity pathway analysis (IPA) to identify networks of interrelated circadian genes in the pituitary (Fig. 6). As expected, the highest confidence network identified (14 genes, \( p < 10^{-8} \)) focused on the canonical circadian clock (Fig. 6, top). In this network, BMAL1 / ARNTL and Clock form a pair of central nodes that link together familiar components of the clock, including Per, Cry, and Rev-erb genes. The similarity of this network to conventional models of clock mechanics suggests that IPA is a useful tool for identifying functionally related genes.

Interestingly, another significant network identified by IPA \( (p < 10^{-4}) \) seems to pivot around genes involved in aldehyde metabolism (Fig. 6, bottom). Of the 17 aldehyde dehydrogenase genes in mice, 5 show significant rhythmicity in the pituitary (Fig. 6, bottom). Previous work has established the role of the hypothalamic-pituitary-adrenal (HPA) axis in modulating alcohol sensitivity (Gianoulakis 1998). Moreover, considerable efforts have been made to document circadian variation in the organismal response to alcohol (Wasielewski and Holloway 2001). However, to the best of our knowledge, no study has addressed the role of ALDH (aldehyde dehydrogenase) genes in the pituitary let alone their role in modulating circadian rhythms of alcohol metabolism. Moreover, several members of the ALDH family have been shown to have important roles in the metabolism of GABA (\( \gamma \)-amino-\( \beta \)-butyric acid) (Vasiliou et al. 2004), an inhibitory neurotransmitter known to have a role in susceptibility to alcoholism (Morrow et al. 2006). On the basis of this ingenuity network analysis, we suggest that the circadian regulation of ALDH genes in the pituitary may be an important component of the organismal response to alcohol. Thus, the combined use of both informatics tools and the Wasabi database may prove to be a valuable strategy for generating testable hypotheses in subsequent studies of circadian clock output and its relationship to clock-regulated physiology.

**CONCLUSIONS**

Recently, more than 70 studies have used microarray analysis to identify genes whose transcription is under circadian control. To extend on this work, we have analyzed RNA samples isolated from the pituitary of wild-type mice every hour over the course of 2 complete days. By increasing the frequency of sampling to once every hour, more than 250 genes in the pituitary were identified as circadianly regulated. Simulations based on this data set indicate that the power of conventional algorithms for detecting rhythmic transcripts is dramatically improved by increasing the sampling density of transcriptional profiling studies. Consequently, we recommend that future studies sample tissues at least once every 2 hours for 48 hours as the best balance in managing costs and robustly identifying circadian genes. Software tools to analyze circadian data sets and the results of this transcriptional profiling study have been made available to the circadian community in a fully searchable Web resource: Wasabi.
AKNOWLEDGMENTS

We thank members of the Hogenesch and Panda laboratories for helpful comments through the execution of these experiments and the preparation of the manuscript. The microarray facilities at Scripps Research Institute and the University of Pennsylvania were indispensable for the collection and analysis of these data. J.B.H. is supported by the National Institute of Neurological Disease and Stroke (1R01NS054794), and the National Institute of Mental Health (P50 MH074924-01, awarded to Joseph S. Takahashi, Northwestern University). S.P. is supported by the Pew Scholars Program in Biomedical Science and the Whitehall Foundation.

REFERENCES

High-resolution Time Course Analysis of Gene Expression from Pituitary

M. Hughes, L. DeHaro, S. R. Pulivarthy, et al.

Cold Spring Harb Symp Quant Biol 2007 72: 381-386

References

This article cites 31 articles, 5 of which can be accessed free at:
http://symposium.cshlp.org/content/72/381.full.html#ref-list-1

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.