

# Systems Biology of Circadian Rhythms: An Outlook

Luciano De Haro and Satchidananda Panda<sup>1</sup>

*The Salk Institute for Biological Studies, La Jolla, CA*

**Abstract** The circadian system in higher organisms temporally orchestrates rhythmic changes in a vast number of genes and gene products in different organs. Complex interactions between these components, both within and among cells, ultimately lead to rhythmic behavior and physiology. Identifying the plethora of circadian targets and mapping their interactions with one another is therefore essential to comprehend the molecular mechanisms of circadian regulation. The emergence of new technology for unbiased identification of biomolecules and for mapping interactions at the genome-wide scale is offering powerful tools to decipher the regulatory networks underpinning circadian rhythms. In this review, the authors discuss the potential application of these genome-wide approaches in the study of circadian rhythms.

**Key words** proteomics, genomics, metabolomics, systems biology, functional genomics

Systems biology may be defined as the systematic and comprehensive identification of components of a given biological system and the concomitant delineation of the functional interaction between these, so that the system's behavior may be understood and predicted. Systems biology can be applied at various levels of complexity. For example, the digestive and immune systems in mammals or the photosynthetic systems of plants may be described as interactions among different organs, cells, or molecular pathways and networks present within relevant cells. Hence, a systems approach integrates information from classical disciplines such as cell biology, physiology, and genetics with emerging fields of genomics, proteomics, and metabolomics (Figure 1).

Our current knowledge of the circadian oscillator at the molecular, cellular, and whole-organism levels, when combined with technological advances in the past two decades, has opened exciting avenues and approaches for the systematic study of the circadian

system. In this review, we discuss the potential of each of these approaches/technologies in (a) the identification of components of the circadian system and (b) mapping functional interaction among these components at cellular and molecular levels, with an emphasis on mammals.

## CELLULAR SYSTEMS BIOLOGY OF CIRCADIAN RHYTHM

For simplicity, the circadian system is broadly divided into 3 modules: (1) the entrainment pathway, (2) the core oscillator, and (3) the output. In unicellular organisms and in plant cells, all 3 modules are usually cell autonomous and, hence, the systems biology of their circadian oscillator largely involves understanding the ensemble of intracellular events. In higher animals, however, specialized sets of cells participate in entrainment, while the oscillating cells

1. To whom all correspondence should be addressed: Satchidananda Panda, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037; e-mail: satchin@salk.edu.

organize themselves into hierarchical orders of master and slave oscillators. Therefore, identifying the organs and cells that function in entrainment, master oscillator, and slave oscillators of higher animals, as well as understanding interactions among these cells, assumes critical significance.

During the past several decades, experiments in rodents have identified organs and cell types of the circadian system. Tissue perturbations by surgical ablation and transplantation experiments have established that the retina and retinohypothalamic tract (RHT) compose the key cellular infrastructure for light entrainment, with the SCN as the core oscillator, while the rest of the body and the brain represent the circadian output limb of the system. Mapping circadian interactions among these components has been greatly facilitated by the establishment of noninvasive luciferase and destabilized green fluorescent protein (GFP)-based reporter transgenic animals that allow real-time measurements of oscillator function in explant cultures (Welsh and Kay, 2005). Such temporal expression markers permit the systematic analysis of cellular oscillators under normal conditions and in response to various genetic, environmental, pharmacological, and surgical perturbations. These studies have revealed heterogeneity in the endogenous period length, phase, and the response of the oscillator in various organs—and even different cells of a given tissue—to perturbing agents (Yamaguchi et al., 2003; Yoo et al., 2004). For example, meticulous measurement of luciferase rhythms in the SCN has revealed that different subregions of the SCN exhibit diversity in oscillator amplitude, phase, and response to phase-resetting stimuli (Yamaguchi et al., 2003). The underlying causes for the observed heterogeneity or its functional significance are not clear at this point. It is often suggested that the differences in relative expression of individual oscillator components may result in functional diversity in oscillator function and that oscillators running at different phases may regulate distinct outputs. These observations and hypotheses have prompted efforts to identify oscillator cells in various organs, examine their gene expression profile, and generate resources that will facilitate probing the interactions between cells and organs relevant to the circadian system.

Temporal expression markers will continue to help identify new organs or cells with the functional oscillator. Integration of such circadian imaging data with the spatial expression pattern of the entire mouse genome would help us (a) understand whether expression of

a given gene or oscillator component correlates with the amplitude, period length, and robustness of the oscillator and (b) hypothesize the functional consequence of the oscillator on the cellular and organ physiology. The postgenomics era is facilitating the achievement of these goals by generating publicly available high-quality data and resources that molecularly define cell types in the mouse body. Two of these resources are the Allen Brain Atlas project and the Gene Expression Nervous System Atlas (GENSAT) transgenic mouse project.

The Allen Brain Atlas project has generated a virtual 3-dimensional expression map of more than 20,000 genes in the mouse brain by performing *in situ* hybridization of individual gene probes in serial brain sections of mice (<http://www.brainatlas.org/aba/>). The data and overlay tools they provide enable users to examine the expression profile of a given gene in the entire mouse brain and compare the expression profile of multiple genes in a specific brain section. In a different approach, genome sequence resources are now making it possible to generate thousands of bacterial artificial chromosome (BAC) transgenic mice expressing GFP-based reporter genes for individual genes (GENSAT) (Gong et al., 2003). These mice resources offer a plethora of cellular markers that can be used by generating double transgenic mice, which harbor both a spatial GFP marker to fluorescently tag a specific cell type and a luciferase marker to monitor oscillator function and, consequently, identify and isolate defined rhythmic cell types in different tissues. These mice resources also will be invaluable for the examination of how oscillator cells differentiate during development.

A related application of the composite spatiotemporal expression data is to generate cell-specific perturbation agents to uncover cellular interactions. Expression pattern data can be used to find cell-specific promoters that, in turn, can be used to generate important genetic resources, such as transgenic mice or lenti/adenoviruses, which specifically knock out gene function or ablate/electrically silence cells (Buch et al., 2005; Slimko et al., 2002). This approach, when applied to the targeted ablation/silencing of oscillator function in SCN subregions, will help determine how cells within the SCN communicate with each other and with their target brain regions. In summary, the use of imaging tools and available genetic resources will help us understand the cellular framework of the circadian system in mammals.

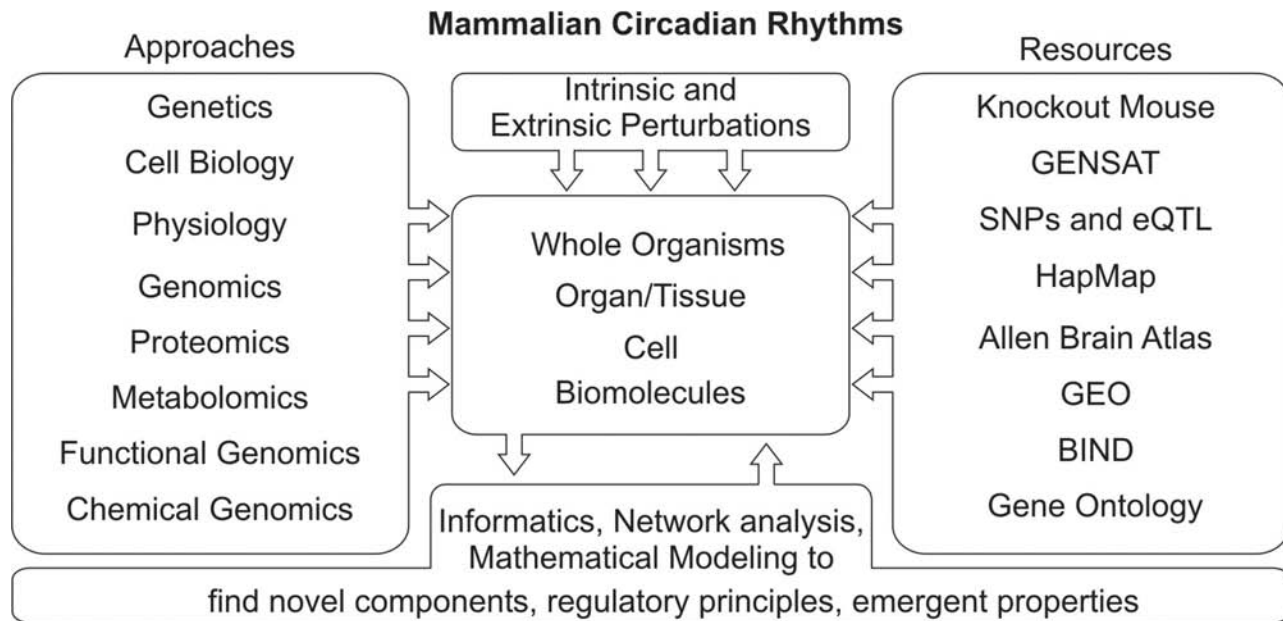


Figure 1. A schematic of systems approaches to circadian rhythms. Traditional and emerging genome-scale approaches (on the left) are applied at the whole-organism, organ, cellular, or molecular level to study the circadian system held either under the “normal” condition or under conditions of intrinsic perturbation (genetic perturbation, tissue ablation, etc.) or of extrinsic perturbation (photoperiod, feeding restriction, etc.). Publicly available resources, data sets, and gene annotations (on the right) are increasingly offering novel discovery tools, entry points, and candidate components of the circadian system. Data generated by these experiments are iteratively subjected to novel informatics approaches, network analysis, and modeling to find important regulatory nodes, discover the emergent property of the system, and predict the system’s behavior under various conditions. GEO, Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>); BIND, Biomolecular Interaction Network Database (<http://www.unleashedinformatics.com/index.php?pg=products&refer=bind>).

### GENETICS—THE TIDE HAS TURNED TO RIGOROUS PHENOTYPING

The classical forward genetic screen has been the single most powerful tool to conclusively identify critical components of the circadian oscillator, and its contribution in advancing the field of chronobiology cannot be overstated. Our knowledge of the molecular components of the core oscillator (Clock, Cycle/Bmal1, Per, Cry, Frq, WC, Kai, Toc, etc.) is the result of traditional genetic studies in *Drosophila*, *Neurospora*, and *Arabidopsis* and in rodents (Harmer et al., 2001). Genetic linkage analyses of segregating populations in naturally occurring murine populations (Shimomura et al., 2001) have indicated the participation of yet unidentified novel loci in the regulation of circadian behavior. New phenotypes and powerful genome-wide genetic analyses hold potential to identify novel oscillator components and important nodes of output regulation.

The power of genetics lies in its ability to both accurately measure a phenotype among a genetically diverse population and to score the genotypes of the

individuals at high resolution to eventually identify contributing loci by linkage analysis. Even in species whose genome has not been fully sequenced, data derived from genetic analyses have proven to be invaluable in identifying these loci when used in combination with ingenious genomics and synteny analysis. Perhaps the best example of this is provided by the cloning of the hamster *tau* locus (Lowrey et al., 2000).

Traditionally, altered period length has been a universal phenotypic criterion in genetic screens for novel oscillator components. However, with the cloning of several oscillator components in different organisms and the detailed characterization of overt rhythms in these mutants, it is now apparent that additional parameters—such as amplitude, activity/sleep consolidation, phase angle of entrainment, or surrogate phenotypes—can also yield important insights. For example, the sleep onset in humans can be described as the phase angle of entrainment. Genetic studies in human patients with an altered phase angle, such as patients with advanced sleep phase syndrome, have led to the identification of key

oscillator components as well as an enhanced understanding of the molecular bases for the observed phenotype (Toh et al., 2001).

Novel techniques to reproduce circadian phenotypes of the patient in cultured cells or use of surrogate markers may expedite genetic studies. Measurement of oscillator function in primary or in transformed cell lines using lentivirus-transduced Period2 promoter-driven luciferase (Per2:Luc) reporter may emerge as a novel assay of circadian function in humans (Brown et al., 2005a). Surrogate phenotypes may indicate abnormal oscillator function and, in addition, help advance our current understanding of the mechanism of circadian control of outputs, as has already been proven in the study of the plant circadian system. The plant oscillator regulates seedling growth and the time to flower. Thus, mutants exhibiting altered growth, as measured by hypocotyl length or altered flowering time, have led to the identification of oscillator components and elucidation of how the oscillator times important stages of growth and development (Doyle et al., 2002; Hazen et al., 2005; reviewed in McClung, 2006). Although similar success stories in mammals are rare, potential surrogate markers exist. One example is the use of the papillary light reflex (PLR) as a surrogate for the light entrainment pathway. Genetic analysis has shown that both the rod/cone and the melanopsin pathway contribute to circadian photoentrainment and that mutations in these pathways also affect PLR (reviewed in Fu et al., 2005). Similarly, increased activity in the open field during subjective daytime and altered insulin response potentially reflect abnormal oscillator function. Mice with diurnal behavior or that are completely arrhythmic would exhibit more daytime activity than their wild-type counterparts, which can be easily assessed by measuring mouse activity for a few minutes in the open field. Likewise, many known oscillator mutants exhibit altered metabolism and metabolic indicators, such as insulin response (Rudic et al., 2004). These surrogate phenotype-based measurements enable investigators to take single time-point samples, which significantly reduce the cost for phenotyping circadian abnormalities by traditional methods.

The sequenced genomes and the emergence of related technologies have made it faster and easier to perform genome-scale analysis of complex phenotypes. Efforts aimed at sequencing the genomes of several inbred model organism strains of plants, flies, and mice already provide rich information on natural polymorphisms and haplotype composition. For example, single-nucleotide polymorphisms

(SNPs)—up to thousands in mice and in *Arabidopsis* and more than a million in humans—are now publicly available that provide a SNP between any 2 genetically dissimilar individuals in almost every kilobase (Hinds et al., 2005; Pletcher et al., 2004; West et al., 2006). This rich haplotype information is exploited by several SNP detection technologies to generate a high-resolution linkage map in less than a week, enable rapid identification of natural or induced random mutations that lead to a given phenotype, find modifier loci affecting function of a given gene, and conduct complex quantitative trait loci (QTL) analysis in a rodent or human population. Once a locus is identified, gene sequence information and gene expression profiles of the defined genetic interval help to rapidly sequence candidate genes. For instance, mutants isolated by a random mutagenesis screen in mice that pervasively affect circadian outputs can be mapped to a genetic interval, and genes in this interval with significant expression in the SCN become an immediate candidate for further sequence analysis.

High-throughput genotyping technology allows high-density haplotype mapping of inbred strains or recombinant inbred lines. This information, when combined with inherent circadian phenotypes of the strains, will permit the *in silico* mapping of loci/genes that may modulate rhythmic outputs, as has recently been successfully implemented in identifying loci that modulate high-density lipoprotein phenotypes in mice and flowering time in *Arabidopsis* (Balasubramanian et al., 2006; Pletcher et al., 2004). Since the haplotype composition of the strains, or recombinant inbred lines once determined, remains unchanged, the power of the process lies in repeated measure of different phenotypes and performing linkage analysis with the fixed genotype information. A radical idea has also emerged to combine haplotype information with gene expression data to identify *cis*- and *trans*-acting regulatory elements in an approach termed *expression QTL analysis* (eQTL) (Chesler et al., 2004). The underlying principle is to substitute a quantitative measure of classical, physical phenotypes with molecular phenotypes such as the expression level of all genes in a given organ (Schadt et al., 2003). Once the “standard” gene expression levels in a tissue are determined, complex linkage analyses of tens of thousands of expression phenotypes with thousands of SNPs are performed to identify haplotype blocks that cosegregate with a high or low level of expression of a given gene. The process generates regulatory network maps of loci that affect the expression of individual genes.



For example, the expression profiles of SCN, liver, and skeletal muscle at a fixed time point from 30 to 40 strains can yield results that predict whether there are tissue-specific genetic networks that modulate expression of *Clock*, *Bmal1*, Cryptochromes, and Period genes or whether a single, global regulatory web exists across different tissues. Each additional layer of phenotypes, such as the metabolomic, proteomic, cellular, or behavioral phenotype of a given organ or organism, will enrich the phenotypic characterization of the strains. The resulting data sets can be continuously iterated to generate high-content regulatory genetic networks of varying complexity from gene expression to whole-organism behavior.

In addition, we are also beginning to witness the power of human genetic analysis. The focused effort by several genome sequencing centers and consortia has made it possible to sequence genes of interest from a large number of human individuals of different ethnic backgrounds to find natural sequence polymorphisms. Compilation of these sequence polymorphisms and analysis of human genetic variation through evolution, as part of the Hapmap project, are generating new powerful tools for genetic analysis of human phenotypic variation that was unimaginable only a decade ago (International HapMap Consortium, 2005). Use of such natural genetic polymorphisms, combined with chronotype assessment of human subjects, may soon shed light on the genetic basis of natural "chronopolymorphism" in humans.

While the approaches described above will enhance our ability for gene discovery in a forward genetic manner, the postgenome era is also beginning to generate powerful reverse genetics tools for functional characterization. For *Arabidopsis* and *Drosophila*, lines that carry loss-of-function mutations for nearly every known protein-coding gene are already available (Alonso et al., 2003; Bellen et al., 2004). In mice, the systematic knockout of every gene in the genome has yet to be achieved (Austin et al., 2004a), but a compilation of data from various distribution centers and international projects reveals that there is an immediate availability of ES cell lines, frozen embryos, or mice that carry a loss-of-function allele of close to 40% of protein-coding genes in the genome (Nord et al., 2006). Many of these publicly available mutants are important gene discovery tools to evaluate the role of genes whose transcripts oscillate in multiple organs, genes that are area-specifically expressed in organs such as the SCN and RHT, and genes that functionally interact with known oscillator components.

## GENOMICS

The term *genomics* generally refers to the global analysis of gene expression by the quantitative measurement of relative RNA abundance or of transcriptionally active regions of the genomic DNA. Current technology in RNA detection has been based on the selective amplification and labeling of poly(A)-containing RNA and hybridization to oligonucleotides or cDNA arrays. In parallel, selective amplification and labeling of chromatin immunoprecipitated (ChIP) DNA followed by hybridization to microarrays has been increasingly popular in the analysis of transcriptionally active DNA (Hanlon and Lieb, 2004).

The systematic analysis of gene expression in whole organisms and in individual tissue types has revealed many important properties of the circadian regulation of behavior and physiology. Circadian transcriptional regulation is both tissue and locus specific. In a given tissue, as many as 10% of the expressed genes exhibit circadian rhythm at the steady-state level, but a great majority of circadian regulation is tissue specific. Therefore, as the number of tissues examined for circadian gene expression increases, so may the number of transcripts that exhibit circadian oscillations. By different accounts and algorithms, up to 50% of the genome in mammals is under circadian regulation (Duffield, 2003; Lowrey and Takahashi, 2004; Oster et al., 2006; Ptitsyn et al., 2006; Sato et al., 2004).

Tissue samples used in most of the expression profiling studies are largely heterogeneous, representing several different cell types defined by function and molecular signatures. For example, there is growing evidence that the SCN is composed of a heterogeneous group of neurons that have differences in neurotransmitter content, afferent and efferent projections, and expression of oscillator genes (Antle and Silver, 2005) that may play specific roles in entrainment, rhythm generation, and output regulation. As methods for expression profiling of a limited material source and methods to specifically mark functionally distinct SCN neurons advance, we will be able to analyze the expression profile of the SCN subpopulation. A similar approach will also enable expression profiling of other critical cell types, such as oscillator neurons of *Drosophila* and melanopsin-containing retinal ganglion cells (RGCs) of mammals, and yield important clues about relevant signaling mechanisms.

The expression profiling efforts have also provided insight into the mechanism of the circadian regulation of a given physiological output. For example, several observed rhythms in physiology and metabolism can

now be explained by circadian transcriptional regulation of relevant pathways (Harmer et al., 2000; Panda et al., 2002). Conversely, many transcriptional outputs have indicated the existence of circadian regulation on novel pathways and processes that were not known to be under this type of temporal control. It is also important to note that only one third of rhythmically expressed genes have any associated meaningful functional annotations. As other fields progress, the annotated genome will continue to allow new discoveries about molecular bases of overt circadian rhythms.

Genomics approaches are also in use to find DNA elements that mediate rhythmic outputs. Circadian expression profiling, in combination with bioinformatics tools, has hinted at the molecular bases for circadian transcriptional regulation of target genes. A study by Ueda et al. (2005) uncovered possible regulation of several phase clusters of transcripts by Bmal1/Cry, Ror/Rev, or the D-element-binding protein (DBP) class of transcription factors. The underlying hypothesis in these analyses has been that transcripts that cycle with a common phase may harbor common *cis*-acting promoter elements that are targets of known transcription factors. A complementary strategy consists of comparing expression profiles of oscillator component mutants with that of the wild-type and combining these results with a bioinformatics approach to define targets of a specific transcriptional regulator. The application of this strategy to mice and flies has revealed direct targets of some of the oscillator components and has highlighted that only a small fraction of rhythmically expressed genes are directly regulated by the core oscillator components (Duffield, 2003). An increasingly popular method for identifying targets of a given transcription factor (TF) has been by performing a chromatin immunoprecipitation assay with antibodies raised against the TF in question, followed by hybridization of the coimmunoprecipitated DNA onto a microarray of promoter DNA (ChIP-on-chip) (Hanlon and Lieb, 2004). A combination of informatics, ChIP-on-chip, and the expression profile of mutants will continue to shed light on the transcriptional regulatory network generating rhythmic transcriptional outputs.

While the ongoing expression profiling experiments have largely focused on endogenous circadian regulation of outputs, the overall rhythms in behavior and physiology observed in the natural settings result from the interaction of the oscillator with the environment. Thus, the expression profiling of tissue samples

collected under controlled food and light conditions has also begun to resolve the molecular bases of interactions between intrinsic timing cues with extrinsic events (Kita et al., 2002). Iteration of such approach in wild-type mice, in entrainment-deficient mice, or under various conditions of light and feeding will help pinpoint the important molecular changes that correlate with phase adjustments in the SCN and in the periphery. By extending this type of analysis to different photoperiods, particularly in plants and photoperiodic animals, clues to photoperiodic control of physiology and behavior may be obtained.

As mentioned above, expression analyses so far have focused on mRNAs encoded by the nuclear genome. In addition, microRNA, noncoding RNAs, mitochondrial/chloroplast genome-derived RNAs, tRNAs, rhythms in splicing, and rhythms in chromatin modification may either play an important role in circadian rhythm maintenance or be subject to circadian regulation. New technologies and approaches that rely on hybridization of all RNAs (except the major ribosomal RNAs) to the array of oligonucleotides tiling the whole genome (tiling array) or sequencing-based detection of expressed genes are powerful tools in addressing these issues (Mockler et al., 2005).

## PROTEOMICS

Dynamic changes in the protein level, subcellular localization, posttranslational modification, and association with other biomolecules often define temporal changes in gene function. Therefore, the systematic quantification of proteins is emerging as a powerful discovery tool. The development of mass spectrometry (MS) methods, coupled with the availability of the entire genome sequence and therefore the predicted amino acid sequences of all possible proteins, has made the identification of specific proteins an efficient and faster endeavor.

Unlike genome-wide detection of DNA or mRNA, protein levels cannot be amplified prior to detection; hence, the initial sample size for protein extraction for proteomic detection has been relatively large. This feature currently limits our ability to carry out comprehensive protein identification experiments from small samples such as the SCN and its subregions. Furthermore, owing to the complexity of amino acid composition, their possible posttranslational modification, and secondary and tertiary protein structure, the complete extraction of all protein constituents

present in a given protocol has been almost impossible. Despite these prevailing limitations, advances in extraction methods and detection sensitivity are being achieved constantly. Proteomics approaches have tremendous potential and applications in the following areas: (1) quantitative protein identification in cellular lysate, subcellular fractions, and biological fluids; (2) posttranslational modifications of a given protein or of a subset of proteins; (3) protein-protein interactions; and (4) mapping of protein surfaces.

Highly sensitive liquid chromatography-based fractionation and MS-based detection technology have begun to make possible the comprehensive identification and quantification of a vast majority of protein constituents of any given tissue type. In turn, this has opened the possibility to test circadian fluctuation in proteins and variation in their subcellular localization. This kind of approach has begun to be applied to identify circadian fluctuations of protein levels in unicellular organisms such as *Chlamydomonas* (Wagner et al., 2004), in mouse liver (Reddy et al., 2006), and protein changes in the rodent SCN in response to light pulse (Fahrenkrug et al., 2005). mRNA levels of several proteins whose levels were found to change in a meaningful manner in these studies were surprisingly constant. This further illustrates the critical posttranslational circadian control mechanism that can be uncovered only by the proteomics approach.

Similarly important, the phosphorylation states of several proteins, including GSK3 $\beta$  and CREB, fluctuate either in circadian fashion or in response to entraining stimuli (Ginty et al., 1993; Iitaka et al., 2005). Identifying posttranslational changes in proteins, whose mRNA levels show only weak or no fluctuation, may help us understand how such alterations impose additional layers of circadian regulation. Such an approach may further uncover important transient posttranslational modifications during entrainment that may precede any detectable change in transcription. However, most posttranslationally modified proteins constitute only a small proportion of the total protein and are quite often transient. Therefore, their accurate identification and quantification largely depends on improvements in the purification of modified proteins from complex tissue samples and their subsequent detection by MS-based methodologies.

One important and immediate application of proteomics in circadian biology is to identify the interacting partners of oscillator components, which, as growing evidence seems to indicate, exist in large, time-variable complexes (Brown et al., 2005b). These complexes can

be purified from cytoplasm or nucleus at different circadian times, and their composition, stoichiometry, and the posttranslational modifications may be assessed.

Finally, an emerging application of proteomics is in the structural analysis of a given protein. Although X-ray crystallography and nuclear magnetic resonance (NMR) are ideal methods for structure determination, large proteins such as the core oscillator components are not easily amenable to high-level expression, purity, and sample preparation conditions for these methods. Therefore, a novel MS-based method that uses hydrogen/deuterium exchange analysis to probe protein surface exposure is gaining ground (Maier and Deinzer, 2005). The determination of the exposed residues and changes in exposed surface under different conditions or in the presence of interacting partners can map interaction modules and ligand-binding sites. In this approach, a protein is incubated in deuteriated water for the exchange of amide hydrogen atoms for those of deuterium. Then, the sample is subjected to proteolysis followed by mass spectrometry analysis. Flexible regions of the protein can then be pinpointed because they will exchange hydrogen for deuterium more than rigid, unchanging regions and, as a consequence, will exhibit a change in their mass-to-charge ratios. Thus, by monitoring the change in the mass-to-charge ratio spectrum of different peptide fragments, it will be possible to assess how different proteins and/or stimuli induce structural changes.

## METABOLOMICS

For simplicity of discussion, we refer to small molecular weight hormones (both peptide and nonpeptide), cofactors, and products or intermediates of metabolic pathways as part of the metabolome. Robust rhythms in hormones produced by the pituitary-adrenal, somatotrophic, and gonadotropic axes, as well as rhythm in prolactin, melatonin levels, and glucose metabolism in young healthy individuals, have been well documented (Becker, 2000). Altered or blunted rhythms of some of these factors have also been correlated with aging or diseased conditions. For instance, low-amplitude rhythms of melatonin, plasma prolactin, and growth hormone are observed in old individuals (van Coevorden et al., 1991), the early phase of some hormone rhythms has been recorded in severely depressed patients (Wehr et al., 1983), and a constitutively high plasma cortisol level is

a key feature of Cushing syndrome (in Becker, 2000). These observations have emphasized the rhythmic oscillation of hormones and metabolites in normal health. In addition, mRNA expression profiles have shown the circadian regulation of mRNA levels coding for many enzymes involved in different metabolic pathways in the liver, including those for cholesterol biogenesis and degradation pathways, energy metabolism, and metabolism of several cofactors such as heme, carotenoids, and cofactors for nuclear hormone receptors (Panda et al., 2002; Storch et al., 2002; Ueda et al., 2002). These small molecules and their intermediates, like classical hormones, often bind to their cognate receptors and mediate important biological functions. Therefore, the comprehensive identification of all small molecules found in plasma and in organs such as the liver, adipose tissue, and skeletal muscle in mice over the circadian cycle will produce reference circadian fluctuations in these factors, generate hypotheses about the biological significance of such fluctuations, and may potentially provide surrogate plasma biomarkers of circadian function in model organisms.

Despite the promise of metabolomics, the prevailing approach, technologies, and informatics needed for its implementation are merely in the early experimental stage (Nobeli and Thornton, 2006). While genomics and proteomics analyses enjoy the advantage of the relatively homogeneous nature of the target molecules, the metabolic constituents are extremely heterogeneous in solubility properties, charge distribution, and chemical nature. Furthermore, unlike the genome or the proteome, whose theoretical constituents can be deduced from the DNA sequence, the metabolic constituents cannot be similarly predicted. These challenges demand rigorous optimization of experimental conditions and close collaboration between chemoinformatics and bioinformatics expertise to expand the gamut of metabolites that can be reproducibly detected and identified.

A practical approach that can be applied to circadian biology in model organisms would involve metabolic fingerprinting or comprehensive quantification of a large number of known serum metabolites and hormones that are identified and quantified using existing methods. As the serum metabolites and hormones originate from different organs in the body, their measurements can serve as surrogate indicators of circadian function in various tissues. Thus, their levels in various genetic, photoperiod, chemical perturbation, and tissue ablation conditions

can offer a unique tool to assess circadian function and build models for systems-level interactions.

## INTERACTOME

Functional interactions between biomolecules often involve direct physical contacts. Therefore, the comprehensive mapping of all possible protein-protein associations has been an emerging approach to generating functional networks (Cusick et al., 2005). Genome-wide interaction screens, such as systematic yeast 2-hybrid assays or protein complex analysis with multidimensional protein identification technology (Mud-PIT), are being used to generate genome-wide interaction maps for model organisms as well as for humans; these maps are deposited in publicly accessible databases (Alfarano et al., 2005). However, information from interactome maps is useful only after careful filtering to ensure that interacting partners are also coexpressed in the same cell and are resident in the same subcellular compartments. Interactions also need wet lab reconfirmation in a cell type or conditions that closely match the physiological context. Such annotated and verified physical interactions from published literature are also being compiled in databases and used to enrich the "annotated interactome." Finally, physical interactions being determined in several circadian labs are also contributing high-quality data to the set. Adding levels of relevant information, such as cellular, subcellular, or the temporal expression profile of the interacting partners, will help build spatiotemporal molecular interaction maps.

## FUNCTIONAL GENOMICS

The availability of libraries of cDNAs, siRNA, and shRNAs; the development of methods for the massive parallel transfer of these agents into mammalian cells; and improvements in detection technologies now make it possible to systematically probe functional interaction between genes. In a typical functional genomics screen, the effect of overexpressing or "knocking down" individual genes on the properties of a gene of interest is typically measured to assign functional interaction. While large-scale genome-wide screens may not be within the reach of individual investigators, focused small libraries are also powerful. For example, a small focused library of 50 transcripts that cycle in 4 different tissues was used to find



transcriptional regulation of *Bmal1* by *Ror- $\alpha$*  (Sato et al., 2004). Similarly focused screens can potentially identify factors that modulate expression, posttranslational modification, subcellular trafficking, and degradation of specific components of the circadian system.

While these functional genomics strategies are already applied to elucidate interactions among components of a system, unbiased approaches to understanding the contribution of specific residues of a protein to its function are also emerging. Typically, libraries of gene mutants that harbor amino acid changes are generated by PCR-based random mutagenesis, in-frame transposon insertion, or mutagen treatment. Subsequently, the impact of these mutations on protein function may be probed in a transient transfection assay. These strategies have been successfully applied to probe functional elements within rodent *Clock* and *Bmal1* protein required for their normal function (Kiyohara et al., 2006; Sato et al., 2006).

### CHEMICAL GENOMICS

Chemical library screens for small molecules that target specific cellular components can be used to identify tools to modulate target function rapidly, reversibly, and conditionally in a dose-dependent manner with both temporal and quantitative control in a given experimental condition, something that cannot be achieved with the genetic perturbation approach. Indeed, several channel and G-protein-coupled receptor (GPCR) signaling modulators have been used to dissect the entrainment mechanism of the oscillator in higher animals (Meijer and Schwartz, 2003). In addition, the use of several small molecules and redox agents has helped researchers understand the function of the oscillator (Rutter et al., 2002; Yin et al., 2006). Nonetheless, the systematic effort to find novel modulators of oscillator components is just beginning. Advances in combinatorial chemistry have led to the development of large, diverse chemical libraries, which are now available to many researchers through the National Institutes of Health's (NIH's) chemical library screening centers and funding mechanisms (Austin et al., 2004b). In parallel, modern robotic technology and sensitive detection technology allow for assay miniaturization and massive screenings against libraries of hundreds of thousands of compounds in a timeframe of days. The discovery of a synthetic or natural small-molecule modulators of oscillator function in a cell-based screening

assay and further validation in the whole organism may ultimately lead to the discovery of candidate natural ligands and cofactors for several core oscillator components, as well as a novel therapeutic intervention strategy to manage circadian disorders in human patients.

### CONCLUDING REMARKS

The genome-wide, unbiased measurements of DNA, RNA, protein, and metabolites and the interactions among them generate such a large amount of data that our ability to reach meaningful conclusions is overwhelmed. Computational analysis of such large data sets has been making a notable contribution to chronobiology. For example, expression profiling experiments have successfully predicted key *cis*-acting regulatory elements in both plant and animal circadian oscillators (Harmer et al., 2000, Ueda et al., 2005). Analysis of data collected by different technologies or experimental conditions holds even greater promise for new discoveries. For example, integration of the circadian expression profile in metabolically important organs with similar expression profile data under different states of feeding, fasting, or redox state can produce new insight and provide novel entry points to understand circadian homeostasis of metabolism and to model how disruption of this homeostasis can enhance disposition to metabolic disorders. Similarly, compilation of proteomics data on oscillator components will generate a temporal profile of protein translocation, phosphorylation, and degradation, which can be integrated with the expression profile and target preference of mediators of protein modification to find novel oscillator components. To achieve these hefty potentials, new computational approaches and algorithms are becoming an integral part of every step of systems biology, from data acquisition and analysis to database construction and search algorithms and, finally, to network analysis and mathematical modeling. It is impossible to comprehensively discuss the important informatics tools and approaches that are currently in development for each step of the process, but we will allude to critical challenges and opportunities for the circadian community.

At the data acquisition and analysis end of the process, several technology-specific informatics and analysis tools are being developed by technology providers and by many freelance informatics labs. As

the technology improves or alternatives emerge, so do the associated algorithms. For example, the algorithms to extract gene expression data from Affymetrix arrays or from spotted arrays are going through several changes. Different algorithms produce slightly different values of the final expression data and the associated error, which in turn influence the downstream data processing and thresholds for defining rhythmic transcripts. As a result, the same experiment performed by different labs with the same algorithm can yield an overlapping but not necessarily identical set of rhythmic transcripts (Walker and Hogenesch, 2005). Although such discrepancy creates a sense of uncertainty about the quality of genome-scale data in general, several steps at the data generation and data use ends can vastly improve the “usefulness” of the approach for wet-lab researchers. For every large data set, an explicit assessment of false-positive and false-negative rates, as well as detailed elucidation of the experimental conditions and analysis parameters, will be a first step in helping the end-user to evaluate the data quality and apply his or her own criteria for subsequent follow-up.

Even if the genome-scale data generated are made freely available, our ability to analyze the resulting network and to generate mathematical models from this data set is currently in its infancy. Most of the prevailing modeling approaches often use only a few elements, and integration of both intercellular and intracellular pathways is rare. However, circadian biology has a long history of close interaction between modeling experts and biologists, so the area offers great opportunities for discovery. One feature of the circadian system that makes it attractive for modeling is the heterogeneity in its behavior in different systems environments, such as in different organs or different organisms. Such heterogeneity, along with perturbation experiments with external agents (light, temperature, etc.) or with intrinsic components (genetic or chemical perturbations), offers rich data sets that can be used to understand the robustness of the core oscillator, discover novel or emergent properties of the circadian system, and predict its behavior under different conditions that no other single approach can easily achieve.

## REFERENCES

Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, Boutilier K, Burgess E, et al. (2005) The Biomolecular Interaction Network

- Database and related tools 2005 update. *Nucleic Acids Res* 33:D418-D424.
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653-657.
- Antle MC and Silver R (2005) Orchestrating time: Arrangements of the brain circadian clock. *Trends Neurosci* 28: 145-151.
- Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, et al. (2004a) The knockout mouse project. *Nat Genet* 36: 921-924.
- Austin CP, Brady LS, Insel TR, and Collins FS (2004b) NIH molecular libraries initiative. *Science* 306:1138-1139.
- Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, and Weigel D (2006) The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat Genet* 38:711-715.
- Becker KL (2000) *Principles and Practice of Endocrinology and Metabolism*. Philadelphia (PA): Lippincott Williams & Wilkins.
- Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, Evans-Holm M, Hiesinger PR, Schulze KL, Rubin GM, et al. (2004) The BDGP gene disruption project: Single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* 167:761-781.
- Brown SA, Fleury-Olela F, Nagoshi E, Hauser C, Juge C, Meier CA, Chicheportiche R, Dayer JM, Albrecht U, and Schibler U (2005a) The period length of fibroblast circadian gene expression varies widely among human individuals. *PLoS Biol* 3:e338.
- Brown SA, Ripperger J, Kadener S, Fleury-Olela F, Vilbois F, Rosbash M, and Schibler U (2005b) PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science* 308:693-696.
- Buch T, Heppner FL, Tertilt C, Heinen TJ, Kremer M, Wunderlich FT, Jung S, and Waisman A (2005) A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nat Methods* 2: 419-426.
- Chesler EJ, Lu L, Wang J, Williams RW, and Manly KF (2004) WebQTL: Rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nat Neurosci* 7:485-486.
- Cusick ME, Klitgord N, Vidal M, and Hill DE (2005) Interactome: Gateway into systems biology. *Hum Mol Genet* 14(Spec No. 2):R171-R181.
- Doyle MR, Davis SJ, Bastow RM, McWatters HG, Kozma-Bognar L, Nagy F, Millar AJ, and Amasino RM (2002) The ELF4 gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature* 419:74-77.
- Duffield GE (2003) DNA microarray analyses of circadian timing: The genomic basis of biological time. *J Neuroendocrinol* 15:991-1002.
- Fahrenkrug J, Hannibal J, Honore B, and Vorum H (2005) Altered calmodulin response to light in the suprachiasmatic nucleus of PAC1 receptor knockout mice revealed by proteomic analysis. *J Mol Neurosci* 25:251-258.

- Fu Y, Liao HW, Do MT, and Yau KW (2005) Non-image-forming ocular photoreception in vertebrates. *Curr Opin Neurobiol* 15:415-422.
- Ginty DD, Kornhauser JM, Thompson MA, Bading H, Mayo KE, Takahashi JS, and Greenberg ME (1993) Regulation of CREB phosphorylation in the suprachiasmatic nucleus by light and a circadian clock. *Science* 260:238-241.
- Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, et al. (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* 425:917-925.
- Hanlon SE and Lieb JD (2004) Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr Opin Genet Dev* 14:697-705.
- Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, and Kay SA (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290:2110-2113.
- Harmer SL, Panda S, and Kay SA (2001) Molecular bases of circadian rhythms. *Annu Rev Cell Dev Biol* 17:215-253.
- Hazen SP, Schultz TF, Pruneda-Paz JL, Borevitz JO, Ecker JR, and Kay SA (2005) LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms. *Proc Natl Acad Sci USA* 102:10387-10392.
- Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, and Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* 307:1072-1079.
- Iitaka C, Miyazaki K, Akaike T, and Ishida N (2005) A role for glycogen synthase kinase-3beta in the mammalian circadian clock. *J Biol Chem* 280:29397-29402.
- International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*:437-1299-1320.
- Kita Y, Shiozawa M, Jin W, Majewski RR, Besharse JC, Greene AS, and Jacob HJ (2002) Implications of circadian gene expression in kidney, liver and the effects of fasting on pharmacogenomic studies. *Pharmacogenetics* 12:55-65.
- Kiyohara YB, Tagao S, Tamanini F, Morita A, Sugisawa Y, Yasuda M, Yamanaka I, Ueda HR, van der Horst GT, Kondo T, et al. (2006) The BMAL1 C terminus regulates the circadian transcription feedback loop. *Proc Natl Acad Sci USA* 103:10074-10079.
- Lowrey PL, Shimomura K, Antoch MP, Yamazaki S, Zemenides PD, Ralph MR, Menaker M, and Takahashi JS (2000) Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* 288:483-492.
- Lowrey PL and Takahashi JS (2004) Mammalian circadian biology: Elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet* 5:407-441.
- Maier CS and Deinzer ML (2005) Protein conformations, interactions, and H/D exchange. *Methods Enzymol* 402: 312-360.
- McClung CR (2006) Plant circadian rhythms. *Plant Cell* 18:792-803.
- Meijer JH and Schwartz WJ (2003) In search of the pathways for light-induced pacemaker resetting in the suprachiasmatic nucleus. *J Biol Rhythms* 18:235-249.
- Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, and Ecker JR (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85:1-15.
- Nobeli I and Thornton JM (2006) A bioinformatician's view of the metabolome. *Bioessays* 28:534-545.
- Nord AS, Chang PJ, Conklin BR, Cox AV, Harper CA, Hicks GG, Huang CC, Johns SJ, Kawamoto M, Liu S, et al. (2006) The International Gene Trap Consortium Website: A portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res* 34:D642-D648.
- Oster H, Damerow S, Kiessling S, Jakubcakova V, Abraham D, Tian J, Hoffmann MW, and Eichele G (2006) The circadian rhythm of glucocorticoids is regulated by a gating mechanism residing in the adrenal cortical clock. *Cell Metab* 4:163-173.
- Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, and Hogenesch JB (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109:307-320.
- Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, Lagler E, Korstanje R, Wang X, Nusskern D, Bogue MA, et al. (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* 2:e393.
- Ptitsyn AA, Zvonic S, Conrad SA, Scott LK, Mynatt RL, and Gimble JM (2006) Circadian clocks are resounding in peripheral tissues. *PLoS Comput Biol* 2:e16.
- Reddy AB, Karp NA, Maywood ES, Sage EA, Deery M, O'Neill JS, Wong GK, Chesham J, Odell M, Lilley KS, et al. (2006) Circadian orchestration of the hepatic proteome. *Curr Biol* 16:1107-1115.
- Rudic RD, McNamara P, Curtis AM, Boston RC, Panda S, Hogenesch JB, and Fitzgerald GA (2004) BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol* 2:e377.
- Rutter J, Reick M, and McKnight SL (2002) Metabolism and the control of circadian rhythms. *Annu Rev Biochem* 71:307-331.
- Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, Naik KA, FitzGerald GA, Kay SA, and Hogenesch JB (2004) A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* 43:527-537.
- Sato TK, Yamada RG, Ukai H, Baggs JE, Miraglia LJ, Kobayashi TJ, Welsh DK, Kay SA, Ueda HR, and Hogenesch JB (2006) Feedback repression is required for mammalian circadian clock function. *Nat Genet* 38: 312-319.
- Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297-302.
- Shimomura K, Low-Zeddies SS, King DP, Steeves TD, Whiteley A, Kushla J, Zemenides PD, Lin A, Vitaterna MH, Churchill GA, et al. (2001) Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice. *Genome Res* 11:959-980.
- Slimko EM, McKinney S, Anderson DJ, Davidson N, and Lester HA (2002) Selective electrical silencing of mammalian neurons in vitro by the use of invertebrate ligand-gated chloride channels. *J Neurosci* 22:7373-7379.

- Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, and Weitz CJ (2002) Extensive and divergent circadian gene expression in liver and heart. *Nature* 417: 78-83.
- Toh KL, Jones CR, He Y, Eide EJ, Hinz WA, Virshup DM, Ptacek LJ, and Fu YH (2001) An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* 291:1040-1043.
- Ueda HR, Chen W, Adachi A, Wakamatsu H, Hayashi S, Takasugi T, Nagano M, Nakahama K, Suzuki Y, Sugano S, et al. (2002) A transcription factor response element for gene expression during circadian night. *Nature* 418: 534-539.
- Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, Iino M, and Hashimoto S (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat Genet* 37:187-192.
- van Coevorden A, Mockel J, Laurent E, Kerkhofs M, L'Hermite-Baleriaux M, Decoster C, Neve P, and Van Cauter E (1991) Neuroendocrine rhythms and sleep in aging men. *Am J Physiol* 260:E651-E661.
- Wagner V, Fiedler M, Markert C, Hippler M, and Mittag M (2004) Functional proteomics of circadian expressed proteins from *Chlamydomonas reinhardtii*. *FEBS Lett* 559: 129-135.
- Walker JR and Hogenesch JB (2005) RNA profiling in circadian biology. *Methods Enzymol* 393:366-376.
- Wehr TA, Sack D, Rosenthal N, Duncan W, and Gillin JC (1983) Circadian rhythm disturbances in manic-depressive illness. *Fed Proc* 42:2809-2814.
- Welsh DK and Kay SA (2005) Bioluminescence imaging in living organisms. *Curr Opin Biotechnol* 16:73-78.
- West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA, and Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Res* 16:787-795.
- Yamaguchi S, Isejima H, Matsuo T, Okura R, Yagita K, Kobayashi M, and Okamura H (2003) Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science* 302:1408-1412.
- Yin L, Wang J, Klein PS, and Lazar MA (2006) Nuclear receptor Rev-erbalpha is a critical lithium-sensitive component of the circadian clock. *Science* 311:1002-1005.
- Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, Slepka SM, Hong HK, Oh WJ, Yoo OJ, et al. (2004) PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc Natl Acad Sci USA* 101:5339-5346.