Boolean function network analysis of time course liver transcriptome data to reveal novel circadian transcriptional regulators in mammals

Maria Simak[a,b,c], Henry Horng-Shing Lu[b,*], Jinn-Moon Yang[c]

[a]Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC; [b]Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, ROC; [c]Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan, ROC

Abstract

Background: Many biological processes in mammals are subject to circadian control at the molecular level. Disruption of circadian rhythms has been demonstrated to be associated with a wide range of diseases, such as diabetes mellitus, mental disorders, and cancer. Although the core circadian genes are well established, there are multiple reports of novel peripheral circadian regulators. The goal of this study was to provide a comprehensive computational analysis to identify novel potential circadian transcriptional regulators.

Methods: To fulfill the aforementioned goal, we applied a Boolean function network method to analyze the microarray time course mouse and rat liver datasets available in the literature. The inferred direct pairwise relations were further investigated using the functional annotation tool. This approach generated a list of transcription factors (TFs) and cofactors, which were associated with significantly enriched circadian gene ontology (GO) categories.

Results: As a result, we identified 93 transcriptional circadian regulators in mouse and 95 transcriptional circadian regulators in rat. Of these, 19 regulators in mouse and 21 regulators in rat were known, whereas the rest were novel. Furthermore, we validated novel circadian TFs with bioinformatics databases, previous large-scale circadian studies, and related small-scale studies. Moreover, according to predictions inferred from ChIP-Seq experiments reported in the database, 40 of our candidate circadian regulators were confirmed to have circadian genes as direct regulatory targets. In addition, we annotated candidate circadian regulators with disorders that were often associated with disruptions of circadian rhythm in the literature.

Conclusion: In summary, our computational analysis, which was followed by an extensive verification by means of a literature review, can contribute to translational study from endocrinology to cancer research and provide insights for future investigation.

Keywords: Chronobiology discipline; Circadian rhythm; Computational biology

1. INTRODUCTION

A vast number of processes in mammals, including transcription of certain genes, are rhythmic, with a period of oscillation of 24 or 12 hours; therefore, they are called circadian or diurnal. Although the circadian rhythms of organisms are endogenous and are controlled by the central body pacemaker located in the suprachiasmatic nucleus (SCN) of the hypothalamus, they are responsive to external signals, such as light, temperature, and glucose level. Sets of genes as well as their amplitude and phases of oscillation in different tissues vary. The liver has the largest proportion of rhythmically expressed genes,[1] which implies a strong interconnection between metabolism and the sleep–wake cycle. Several studies have indicated a need for further investigation of molecular mechanisms of the circadian clock. For instance, an imbalance of circadian rhythms in shift workers has been associated with diabetes mellitus,[2] severe circadian disruptions were found to be connected to schizophrenia,[3] and the risk of breast cancer was found to be elevated in women working night shifts.[4] To date, extensive research has been done in an attempt to identify important circadian genes and their relations with each other. The most notable microarray transcriptome data were obtained by Almon et al (liver and skeletal muscle of rats),[5,6] Atwood et al (mouse liver),[7] Bailey et al (rat pineal gland),[8] and Zhang et al (12 mouse organs).[1] Vollmers et al used deep sequencing technology for analyses of mouse liver transcriptome and epigenome and identified the circadian oscillating genes.[9] Yoshitane et al conducted analysis of high-throughput ChIP-Seq data to reveal *Clock*-controlled E-boxes across the genome.[10] In the most recent study, Wang et al used an integrative approach of multiple experimental techniques on mouse liver transcriptome and proteome to provide a list of transcription factors (TFs) and cofactors involved in circadian regulation.[11] Westermark and Herzel also integrated expression data of mouse liver and TF binding information to identify the TFs that oscillate with a period of 12 hours.[12] Laing et al proposed

*Address correspondence: Dr. Henry Horng-Shing Lu, Institute of Statistics, National Chiao Tung University, 1001, Daxue Road, Hsinchu 300, Taiwan, ROC. E-mail address: hslu@stat.nctu.edu.tw (H.H.-S. Lu).

Conflicts of interest: The authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

Journal of Chinese Medical Association. (2019) 82: 872-880.

Received May 3, 2019; accepted July 14, 2019.

doi: 10.1097/JCMA.0000000000000180.

biomarkers for human circadian phase obtained from blood transcriptome.[13] Korenčič et al demonstrated that the expression of *Clock*-controlled genes in the mouse liver and adrenal gland has surprisingly little overlap.[14] The study by Relógio et al proposed a list of circadian oscillators associated with deregulation of the circadian clock in skin and colorectal cancer.[15] We used these high-throughput studies as references to validate the results obtained using the Boolean function network (BFN) method on publicly available transcriptome datasets of mouse and rat livers.[5,16] BFN is a two-step procedure that integrates the hidden Markov model, likelihood ratio test, and Boolean functions to infer direct pairwise relations between genes from time course expression data. Previously, we successfully applied BFN to the yeast time course data and proved its advantages over other methods in identifying direct pairwise regulatory relations between genes.[17] For instance, it achieves higher accuracy in reverse engineering of a gene regulatory network, it differentiates direct and indirect relations, it is computationally efficient on large datasets, and it assigns direction, Boolean function, and time delay to the link.

2. METHODS

The process of BFN inference is illustrated in Figure 1.

2.1. Datasets

We used publicly available datasets from a gene expression omnibus genomics data repository. Dataset GSE8988 contained expression data of genes in rat liver over the course of 24 hours and comprised three biological replicates and 15 923 microarray probes at 18 time points.[5] Dataset GSE11923 was an expression array data of mouse liver genes over 48 hours that comprised 45 101 probes at 48 time points, which were pooled across three to five biological replicates.[16]

2.2. Preprocessing

The preprocessing part consisted of following four steps: removing probes from analysis that had no corresponding gene IDs in gene annotation tool, filtering probes with low variance and values, assigning a list of TFs and cofactors as sources (targets can be both TFs and genes not involved in transcriptional activity), and undertaking empirical cumulative distribution function (ECDF) discretization.

We applied the database for annotation, visualization and integrated discovery (DAVID) conversion tool for mapping affymetrix probes to standard gene names.[18] This step removed 1760 probes from the rat dataset and 5697 probes from the mouse dataset.

As the second step of preprocessing, analysis of variance (ANOVA) with a significance level of 0.05 was applied to the rat dataset, reducing the number of probes to 2164 genes. For the mouse dataset, we consecutively filtered out 25% of probes having low absolute values and 25% of probes having low variance, leaving 25 252 genes for further analysis. The significance level of ANOVA for the rat dataset and quartiles for the mouse dataset were chosen so that the known core circadian genes were not discarded. We checked for probes corresponding to genes *Clock*, *Arntl*, *Per1*, *Per2*, *Per3*, *Cry1*, *Cry2*, *Nr1d1*, *Nr1d2*, and *Rora* to be present in the reduced datasets.

The list of TFs and cofactors for the third step of preprocessing was obtained from AnimalTFDB 2.0[19] and included 1223 Ensemble IDs for rat and 1882 Ensemble IDs for mouse, corresponding to 183 and 2374 probes for rat and mouse in the reduced datasets, respectively.

Discretization with an ECDF was applied for each gene that transformed expression values to continuous values lying within the [0, 1] interval, which was required for input in Test 1 and Test 2. Prior to the discretization of values, biological replicates in the rat dataset were averaged.

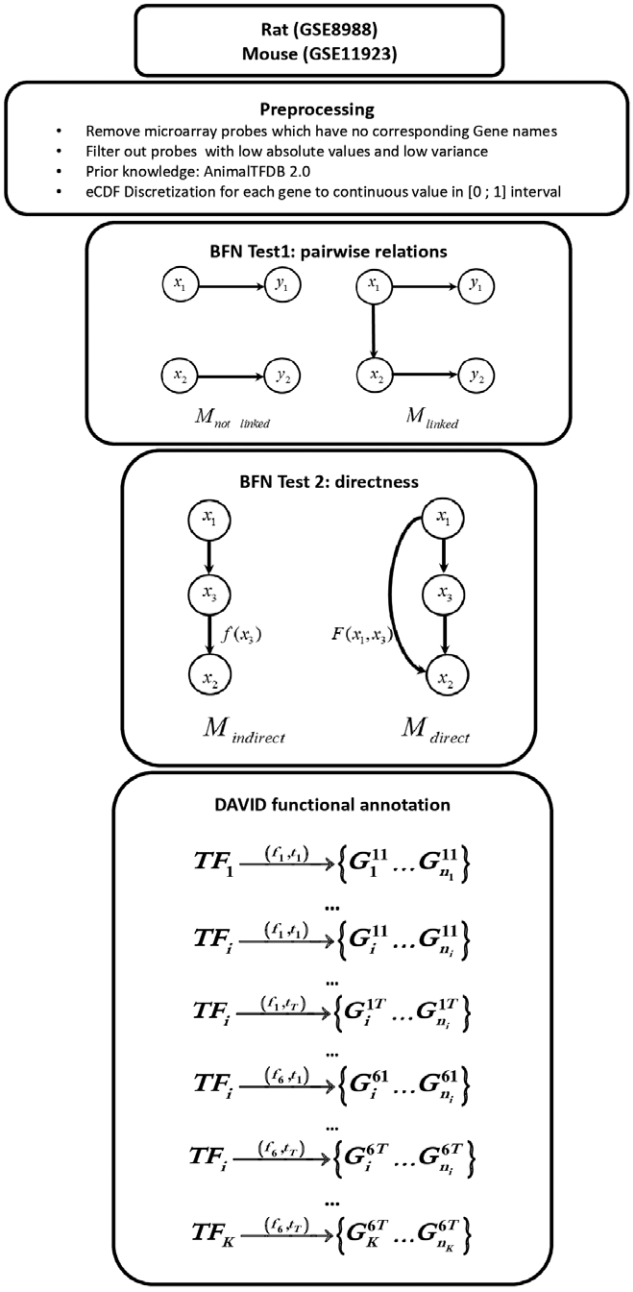

Fig. 1. Steps of BFN analysis. BFN, Boolean function network.

2.3. BFN Test 1 and Test 2

We applied BFN procedures after preprocessing to identify pairwise relations between genes (Test 1) and consecutively filtered out indirect links among those (Test 2). Both tests were based on the comparison of two concurrent models, namely linked vs not linked in Test 1 and direct vs indirect in Test 2. Models were based on the assumption that true gene expressions, which were unaffected by other genes and measurement errors at a given time point, were not known (x_1 and x_2 at Figure 1 for Test 1), but realization of these variables (y_1 and y_2) in the form of discretized expression values was observed. In each test, we calculated

the likelihoods across time points for both concurrent models and inferred Boolean functions and time delays that maximize this difference. On the basis of likelihood ratios, we made conclusions as to whether pairs of genes were linked or not linked (Test 1) and had direct or indirect links (Test 2). Details of BFN procedures can be found in our previous manuscript.[17] The output of both tests depended on the following three parameters: the significance level of Test 1, the significance level of Test 2, and the maximum allowed time delay between genes. For the rat dataset, we used the following BFN settings: time delay range from 1 to 5, $p_1 = 0.05$ as the significance of Test 1 and $p_2 = 0.05$ as the significance of Test 2. For the mouse dataset, the settings were as follows: time delay range from 1 to 8, $p_1 = 0.0001$ as the significance of Test 1, and $p_2 = 0.05$ as the significance of Test 2. The maximal allowed time delays were decided based on the number of time points in the dataset. Limiting the number of time delays serves to avoid unnecessary computations, because with increase in time delay between genes the likelihood of the genes being related decreases in general, as there are fewer points for comparison. The thresholds for p values of Test 1 were chosen based on the number of links that we wanted to use for further functional analysis (ie, $p_1 = 0.0001$ in Test 1 for mouse produced 276 767 links and $p_1 = 0.00005$ resulted in only 2944 links). Thus, with a threshold of $p_1 = 0.00005$, many important relations were missed.

2.4. Functional annotation

The BFN method produced pairs of genes having regulatory relations with assigned Boolean functions and time delays. Thus, every TF or TF cofactor had sets of related target genes, which were found to be regulated by this TF and had the same Boolean function and time delay; these were thus pooled together. These resulting sets were further directed to the functional annotation tool, DAVID.[19] TFs associated with significantly enriched circadian gene ontology (GO) categories were labeled as our candidate circadian TFs and were arranged based on the literature and database scores. In total, we analyzed 557 gene groups of rat and 531 gene groups of mouse datasets that resulted in the identification of 158 rat and 159 mouse gene sets, having significant "circadian rhythm." These sets correspond to 93 unique TFs of mouse and 95 unique TFs of rat.

The source code of BFN, the results of BFN Test 1 and Test 2, and the results of the functional annotation of gene sets for both mouse and rat datasets can be found in our repository: https://github.com/BooleanFunctionNetwork/CIRCADIAN.

3. RESULTS

3.1. Candidate circadian regulators: Validation with databases and circadian studies

To validate and enhance the results of computational analysis described in the previous section, we carried out a reference search of databases, published high-throughput circadian studies, and small-scale studies for each individual candidate circadian TF. This enabled us to compile a list of mammalian circadian regulators arranged according to the current knowledge of their biological significance in circadian processes. The complete list of 165 genes with detailed reference information can be found in our repository in Supplementary Table 5, http://links.lww.com/JCMA/A32. Table 1 is a fragment of the full list that presents the top 20 novel circadian regulators.

Table 1

Top 20 novel circadian regulators inferred using the BFN method

Gene Name	RGD	MGI	NCBI Human	NCBI Mouse	NCBI Rat	Evidence	CGDB	CGDB Species	CircaDB Human	CircaDB Mouse	Literature Score	Dataset	Reference
Klf15			+	+		Reactome[a]	+	*Mus musculus, Rattus norvegicus*	+	+	2	Mouse/rat	20, 21
Rbpms			+	+		WikiPathways (Zambon)[b]	+	*Mus musculus*	+	+	1	Mouse/rat	22
Btg1			+	+		WikiPathways (Zambon)	+	*Gallus gallus*	+	+	0	Mouse/rat	23
Sumo3			+	+		WikiPathways (Zambon)	+	*Rattus norvegicus, Mus musculus*		+	0	Rat	24, 25
Tob1			+	+		WikiPathways (Zambon)	+	*Mus musculus*	+	+	1	Rat	
Tbl1xr1			+			Reactome	+	*Mus musculus*	+	+	1	Rat	26
Nr1h3			+			TAS,[c] WikiPathways (Fehrhart)[d]	+	*Mus musculus*		+	1	Rat	27
Ncoa1			+			Reactome	+	*Mus musculus*	+	+	0	Mouse	
Ncoa6			+			Reactome	+	*Mus musculus*	+		0	Rat	
Esr1							+	*Mus musculus, Rattus norvegicus, Homo sapiens*	+	+	6	Mouse	28, 29, 30
Litaf							+	*Homo sapiens*	+	+	5	Mouse/rat	31
Nr3c2							+	*Mus musculus, Homo sapiens*	+	+	5	Mouse	32
Tsc22d3							+	*Mus musculus, Rattus norvegicus, Homo sapiens*	+	+	4	Mouse	33, 34, 35
Pnrc1							+	*Mus musculus*	+	+	3	Mouse/rat	
Klf13							+	*Mus musculus*	+	+	3	Mouse/Rat	
Foxo3							+	*Mus musculus, Rattus norvegicus*	+	+	3	Mouse	36
Nr2f2							+	*Mus musculus*	+	+	3	Mouse	37
Ppard							+	*Mus musculus*	+	+	3	Rat	38, 39
Smad4							+	*Mus musculus*	+	+	3	Mouse	40
Ctnnb1							+	*Mus musculus, Homo sapiens*		+	3	Mouse	41

[a]R-HSA-1368108 (BMAL1:CLOCK, NPAS2 activates circadian gene expression) Reactome pathway.

[b]"Exercise-induced circadian regulation" WikiPathway (provided by Zambon et al, Genome Biology 2003).

[c]Traceable author statement (TAS).

[d]"Circadian rhythm related genes" WikiPathway (provided by Fehrhart et al).

BFN = Boolean function network; CGDB = circadian gene database; CircaDB = circadian expression profiles database; NCBI = National Center for Biotechnology Information; MGI = mouse genome informatics; RGD = rat genome database.

Columns 2–4 of Table 1 indicate the presence of annotation of candidate circadian TFs in one of the following databases: rat genome database (RGD; information available for both rat and mouse species),[42] mouse genome informatics (MGI; for mouse only),[43] and National Center for Biotechnology Information (NCBI) database (human, mouse, and rat).[44] RGD, MGI, and NCBI databases have multiple sources of information; therefore, the source of evidence is mentioned in column 7. The next two columns (columns 8 and 9) indicate validation using the circadian gene database (CGDB).[45] This database contains both potentially oscillatory and experimentally validated (by real-time polymerase chain reaction, Northern blot, and in situ hybridization) oscillatory transcripts of multiple species; we searched only among experimentally validated ones and reported species in which this gene transcript oscillated. The next two columns (columns 10 and 11) are validation by circadian expression profiles database (CircaDB), the database of circadian transcriptional rhythms from 13 human, and 12 mouse organ systems.[46] Annotations of RGD, MGI, and NCBI of circadian genes were much more discriminative than the ones of CGDB or CircaDB. Only 23.6% of our candidate TFs had circadian annotation by any of RGD, MGI, or NCBI, whereas the CGDB confirmed 93.3% of our circadian candidates (if orthologs were counted; that is, the candidate coming from the rat dataset was experimentally validated in mouse). Analogously, CircaDB Human and CircaDB Mouse validated 58.8% and 80% of our results, respectively. This was the reason for placing TFs confirmed by any of MGI, RGD, or NCBI databases at the top of our table. Column 12 of Table 1 indicates the number of matches with previous high-throughput circadian or diurnal studies that have been listed in the introduction.[1,5–15] The results of intersection of our candidate circadian regulators list with circadian regulators from each of the 12 previous studies can be found in Supplementary Table 6, http://links.lww.com/JCMA/A33, and Table 7, http://links.lww.com/JCMA/A334, for mouse and rat, respectively. Moreover, in Supplementary Table 8, http://links.lww.com/JCMA/A35, and Table 9, http://links.lww.com/JCMA/A36, we have provided original lists of circadian genes from each study for mouse and rat, respectively, adjusted according to the knowledge about TFs (AnimalTFDB 2.0) because most of the studies (except those by Wang et al and Westermark and Herzel) did not focus on TFs only but also considered all rhythmic or circadian-controlled genes. Column 13 of Table 1 indicates the dataset from which the result was inferred, and TFs that were identified as circadian in both datasets are highlighted in bold font. Column 14 of Table 1 cites the small-scale studies, which

Fig. 2. Seven major groups of circadian TFs and TF cofactors inferred with BFN analysis. Group 1: genes validated by RGD, MGI, and NCBI databases to be circadian. Group 2: genes validated by NCBI (either for human, mouse, or rat). Groups 3, 4, and 5: genes supported by at least one of CGDB, CircaDB Mouse, or CircaDB Human databases which have literature scores of at least 3, 2, 1, respectively. Group 6: genes supported by at least one of CGDB, CircaDB Mouse, or CircaDB Human databases and not supported by any of the high-throughput circadian studies. Group 7: genes that are neither supported by databases or previous high-throughput circadian studies. The genes found to be circadian in both mouse and rat datasets are highlighted in bold. BFN, Boolean function network; CircaDB, circadian expression profiles database; CGDB, circadian gene database; MGI, mouse genome informatics; NCBI, National Center for Biotechnology Information; RGD, rat genome database; TF, transcription factor.

explain the mechanism of involvement of every particular gene in circadian regulation, wherever available.

The results of our study along with seven main categories of candidate circadian TFs and cofactors are summarized in Figure 2. The first group comprises TFs, which were confirmed to be circadian by all databases. A total of 31 genes were confirmed to be circadian, which included 11 from the mouse, 12 from the rat, and nine from both datasets. In this article, we have focused on the novel circadian TFs.

Group 2 comprises nine genes, which were annotated as circadian by NCBI for at least one of the species, either human, mouse, or rat. These nine genes, which are at the top of our novel candidate circadian regulators, are described in detail in Table 1. Although they are most likely to be circadian, inconsistency between their annotations across databases was observed. MGI and RGD do not annotate these genes as circadian because they only provide relevant GO categories, whereas NCBI also includes annotation from the pathway databases. While all nine genes from Group 2 besides NCBI were also validated by CircaDB and CGDB databases (at least as orthologs), some of the TFs in this group had more literature support. For instance, Kruppel-like factor 15 (*Klf15*) was not only supported by high-throughput studies by Almon et al and Yoshitane et al[6,10] but is involved in circadian control of bile acid synthesis in mice and transcriptomic oscillations in the mouse heart according to our literature review.[20,21] Another gene that was inferred in the rat and mouse dataset was an RNA-binding protein with multiple splicing (*Rbpms*) that is uniquely expressed in retinal ganglion cells, which innervate SCN and mediate circadian responses in multiple species, including rat.[22] Similarly, BTG antiproliferation factor 1 (*Btg1*) was recently demonstrated to be a component of the circadian immune system in burn trauma in rats.[23] According to our references, sumoylation of BMAL1 by Small Ubiquitin-like Modifier 3 (*Sumo3*) is an essential posttranslational modification required for the circadian cycle in multiple species, including human, mouse, and rat.[24,25] Although no relevant supporting study for circadian mechanism of Transducin Beta-like 1 X-linked Receptor 1 (*Tbl1xr1*) in mammals was found, it was reported to be part of the circadian mechanism in ectotherms.[26] Wada et al revealed functional interplay of circadian ROR-alpha and liver X receptor (*Nr1h3*) in mouse lipid homeostasis.[27] Our results suggested that *Nr1h3* has circadian activity in rats as well.

Groups 3, 4, and 5 comprise TFs supported by at least one of the CGDB, CircaDB Human, or CircaDB Mouse databases and have literature scores of at least 3, 2, and 1, respectively (Fig. 1). Group 6 includes TFs that were only verified by at least one of the CGDB, CircaDB Human, or CircaDB Mouse databases and were not supported by any previous high-throughput circadian studies, which were used as references for this study. Group 7 includes genes that were not validated by either databases or high-throughput studies. In general, this classification into groups reflects the degree of confidence whether each TF is circadian. However, there were TFs that stood out based on individual studies, such as Estrogen receptor 1 (*Esr1*), lipopolysaccharide-induced tumor necrosis factor factor *(Litaf*), Nuclear receptor subfamily 3 Group C member 2 (*Nr3c2*), TSC22 Domain Family Member 3 (*Tsc22d3*), *Foxo3*, *Nr2f2*, proliferator-activated receptor delta (*Ppard*), SMAD family member 4 (*Smad4*), and Catenin Beta 1 (*Ctnnb1*) of Group 3 (Fig. 2).

Esr1 had the highest literature score because it was identified as circadian in several high-throughput studies, namely Zhang et al,[1] Yoshitane et al,[10] Wang et al,[11] Westermark and Herzel,[12] Laing et al,[13] and Korenčič et al[14] Specifically, the study by Zhang et al named it as a circadian drug target, Yoshitane et al demonstrated that it was directly controlled by CLOCK-BMAL1 complex, Wang et al identified it as a TF that controls the diurnal rhythm of genes in the mouse liver, Westermark and Herzel suggested that together with other genes it can create 12-hour oscillations in controlled genes, Laing et al recognized it as being rhythmic in phase with melatonin during sleep, and Korenčič et al observed it oscillating in both the liver and adrenaline gland, but in different phases. Moreover, in small-scale studies, it was demonstrated to modulate circadian rhythms in adult female mice and circadian systemic circulation in male rats[28,29] and to be involved in circadian regulation of breast cancer proliferation in humans.[30] *Litaf*, which also has high literature support and has been found in both rat and mouse datasets, is a key mediator of the inflammatory cytokine response to lipopolysaccharides. *Litaf* is also known as a sleep deprivation biomarker;[31] therefore, it may have a role in establishing a connection between immunity and circadian activity. *Nr3c2* encodes the mineralocorticoid receptor, which contributes to control of blood pressure and cardiac function through regulation of sodium transport in renal tissue. Gumz et al demonstrated the circadian nature of this process.[32] Gene *Tsc22d3* is one of the universal oscillators and is translated into GILZ protein. *Tsc22d3* was found to be related to many circadian processes in different tissues, such as kidney homeostasis,[33] metabolism in adipose tissue,[34] and age-related memory formation in hippocampus.[35] Chaves et al demonstrated that PI3K-FOXO3 signaling was required for circadian rhythmicity in the liver through regulation of the *Clock* gene.[36] Nuclear receptor (NR) *Nr2f2* encodes COUP-TF II protein, and its reduction in *Clock*-mutant mouse causes accumulation of adipose tissue, and consequently, the overweight phenotype.[37] *Ppard* modulates circadian utilization of lipids by muscles[38] and was demonstrated to be altered in pregnant woman with gestational diabetes mellitus.[39] *Smad4*, a part of Smad2/3:Smad4 complex, plays a role in interactions between the circadian clock and TGF-β signaling in zebrafish larvae.[40] Gene *Ctnnb1* encodes β-catenin, which potentially acts as an oncogene in colorectal cancer and is *Clock* controlled.[41] This and previous high-throughput circadian studies supporting *Ctnnb1* may encourage further investigation of its role in circadian regulation. The aforementioned genes are included in the top 20 circadian candidates listed in Table 1. We also included TFs *Pnrc1* and *Klf13* in Table 1 because of their relatively high literature scores and presence in results of both mouse and rat datasets, despite the fact that no convincing small-scale study was found that would explain their involvement in circadian regulation.

In Group 4 of Figure 2, we would like to highlight *Mitf*, *Dnmt3b*, *Hes6*, and *Rarb* genes. The literature review established the following circadian associations: *Mitf* with melanin synthesis,[47] *Dnmt3b* with feeding regulated DNA methylation in liver,[48] *Hes6* with cholesterol homeostasis,[49] and *Rarb* with vitamin A modulated hippocampal rhythms.[50]

In Group 5, genes *Elk3*, *Gtf2ird1*, *Mef2a*, *Ppargc1b* stood out based on the evidence of circadian activity found in small-scale studies. In these studies, the circadian rhythms were associated with psychosis and schizophrenia (*Elk3*),[51] decreased locomotor activity and Williams syndrome (*Gtf2ird1*),[52] Leydig cells homeostasis (*Mef2a*),[53] and mitochondrial oxidative energy metabolism (*Ppargc1b*).[54] In Group 6, the genes associated with circadian processes are *Nlk* (phosphorylates core circadian genes in drosophila),[55] *Nfkbia* (inflammatory response induced by weight loss changes),[56] and *Pdlim1* (cardiac hypertrophy).[57]

The genes in Group 7 were neither supported by databases nor by studies that were used as references for this study. However, we would like to highlight the *Dlx5* gene because it was found to be associated with self-reported "morningness" in a genome-wide association study,[58] and alterations in DNA methylation of this gene in shift workers has been associated with increased risk of cancer development.

As an observation across groups, there were categories of TFs and TF cofactors that were overrepresented in our results, such as the KLF family of TFs (*Klf1*, *Klf9*, *Klf10*, *Klf13*, *Klf15*, and *Klf16*); some of them, such as *Klf10*, *Klf9*, and *Klf15*, have an established function in circadian regulation of mammals.

Although other genes, such as *Klf1*, *Klf13*, and *Klf16*, have been less studied, they have a potential role in the circadian process because they possess E-boxes in their promoter region, which are binding sites for the BMAL1-CLOCK circadian complex. Another large group of TFs found in our results was NRs (*Nr1d1*, *Nr1d2*, *Nr2f6*, *Nrip1*, *Nr2f2*, *Nr3c2*, *Nr1i2*, *Nr1h3*, and *Nr4a2*). They are hormone-sensitive transcriptional regulators involved in many biological processes, such as development, energy metabolism, reproduction, inflammation, and tissue homeostasis. There are growing number of studies that have demonstrated that NRs are regulated by the clock genes and that they modulate circadian activity.[59] Here, we identified novel circadian regulators among NRs, namely *Nr3c2*, *Nr2f2*, *Nr1i2*, *Nr1h3*, and *Nr4a2*. In addition, we found many metabolically associated transcriptional regulators and co-regulators in our results (ie, *Klf15*, *Nr1h3*, *Ppard*, *Nr2f2*, *Tsc22d3*, *Pparg1b*, and *Hes6*), which may serve as links that couple metabolic signals from peripheral oscillators to the master molecular clock of SCN. Notably, we had TFs overlapping between two datasets (selected in bold in Figure 2 and Table 1). Although some of them, such as *Pnrc1*, *Arl2bp*, *Irf6, Eno1*, *Pdlim1*, *Nfkbia*, *Tdp2*, *Carhsp1*, and *Sfmbt1*, do not have a definitive known role in circadian processes, they have been verified by CGDB and CircaDB databases and also by some high-throughput circadian studies to be rhythmic; therefore, they deserve further detailed investigation.

3.2. Validation of candidate circadian regulators with ChIP-Seq experimental data

ChIP–ChIP and its recent successor ChIP-Seq are experimental techniques that enable the identification of potential binding sites of a gene of interest. Several databases are devoted to ChIP-Seq experimental information; these include ENCODE, ChIPBase, GTRD, and ChIP-Atlas.[60] For our study, we chose ChIP-Atlas because it covers the largest number of experiments (over 96 000) and TFs (over 700 of human and 500 of mouse). We searched the complete list of 165 candidate circadian regulators that were obtained through BFN analysis of the microarray data to identify their potential binding targets. Target genes were accepted if the peak-call intervals of a given protein overlapped with a transcription start site ± 5 kb. The details of circadian targets of specific TFs can be found in Supplementary Table 10, http://links.lww.com/JCMA/A37, in our repository. Table 2 summarizes the results of our ChIP-Atlas search. The first column refers to the group number that was assigned earlier to the TFs based on the literature evidence. The gene names in the second column written in capital letters indicate that circadian targets of these genes were found in humans only, otherwise either from mouse or both mouse and human. Currently, information for rat species in ChIP-Atlas is scarce; thus, we limited our search to mouse and human.

Table 2 lists 40 TFs that have more support for classifying them as direct circadian regulators than the rest of the candidates. However, remaining candidates were not discarded because additional experimental evidence may, and is likely to, appear. Moreover, Yoshitane et al indicated the importance of indirect transcriptional and posttranscriptional regulators because there are many rhythmic genes that do not have CLOCK-binding cites.[10] Thus, our novel candidate circadian regulators *Klf13, Ppard, Litaf, Tsc22d3, Gtf2a2, Nfx1, Lpin2, Gtf2ird1, Zkscan17, Pqbp1, Tfdp2, Irf6, Trib3, Nab1*, and *Skil* that are directly controlled by *Clock* (according to Yoshitane et al) can be indirect regulators for other genes. Notably, genes *Nr3c2, Esr1, Klf15, Dnmt3b, Ldb1, Erf,* and *Tle1* belonged to both categories: directly controlled by *Clock* and have direct core circadian targets.

Table 2

TFs with predicted binding sites near core circadian genes inferred from ChIP-Seq experimental data

Group	Genes
2	*Nr1h3, KLF15, TBL1XR1, NCOA1*
3	*Esr1, Foxo3, Smad4, Ctnnb1, Irf1, Mafb, NR3C2, NR2F2*
4	*Mitf, Mxi1, Tle1, Irf2, Xbp1, Nfia, Rarb, Erf, Smarcc1, DNMT3B, GTF2I, RB1*
5	*Gmnn, Ldb1, Tead1, Aff4, ELK3, MEF2A, TRIM24, KLF1*
6	*Sfmbt1, Sap130, Cdk8, Cbfb, Cebpg, Jarid2, Thap11, Aebp2*

TFs = transcription factors.

3.3. Candidate circadian regulators and associated disorders

Information regarding diseases that were previously associated with circadian rhythm disturbances of our candidate circadian regulators is presented in Table 3.

Interestingly, some of the genes have been overrepresented within one cluster of diseases, such as *Ppard* in metabolic disorders and *Nr3c2* in cardiovascular diseases, whereas other genes, such as *Esr1*, can be seen to be associated with almost all disease categories.

4. DISCUSSION

In this study, we developed a systematic approach to discover candidate circadian transcriptional regulators. We applied the BFN method to infer direct pairwise relations between genes along with corresponding Boolean functions and time delays from available transcriptome time course datasets of mouse and rat livers. To reduce the number of false positive links, we integrated prior knowledge regarding TFs and cofactors in mouse and rat. As a result of the BFN algorithm, we identified target genes for each transcriptional regulator associated with the optimal Boolean functions and time delays. The target genes were divided into sets according to these attributes, and the GO enrichment analysis was conducted. Finally, we identified a TF as a circadian candidate if the associated target gene groups had at least one group revealing significant GO annotation related to circadian process. Our analysis discovered 93 transcriptional circadian regulators in mouse and 95 transcriptional circadian regulators in rat, 23 of which were common for both datasets. Although the computational approach used in this study was different from the typical analysis of rhythmicity of gene expression profiles, its results were in accordance with the previous research. This confirmation provided a foundation of confidence regarding candidate circadian regulators that were discovered through this study. Compared with the other methods of reverse engineering the gene regulatory network, which are based on correlation or mutual information, the proposed BFN method assigns Boolean functions and time delays to each relation. Thus, we were able to define more specific groups of related target genes, which further established the basis of biological investigation with functional annotation. The additional strength of this article is an extensive validation of our results with 12 previous circadian studies, three integrative bioinformatics databases, two circadian databases, and the related small-scale circadian studies. Furthermore, we integrated our results with the available ChIP-Seq data, and the investigation revealed that 40 circadian TF candidates discovered in this study had potential binding sites near core circadian genes. The limitation of the current study is that we restricted the scope of our research with transcriptional regulators only, excluding interactions on proteome level. Moreover, more time course datasets from different species and tissues can be analyzed in future studies. Meanwhile, we have highlighted some of the strong candidate circadian regulators and co-regulators that have potential application in molecular medicine.

Table 3

Candidate circadian regulators and associated disorders

Metabolic Disorders			
Diabetes	**Obesity**	**Fatty Liver**	**Insulin Resistance**
Ncoa6	Nr1h3	Nr1h3	Trib3
Esr1	Ncoa1	Ppard	Ppard
Litaf	Esr1	Xbp1	Xbp1
Foxo3	Tsc22d3	Nr1i2	Rb1
Ppard	Foxo3		Ppargc1b
Smad4	Ppard	**Hyperinsulinism**	Insr
Dnmt3b	Nr1i2	Tsc22d3	
Nr1i2	Rb1	Insr	
Mef2a	Aff4		
Nfkbia	Ppargc1b		
Ppargc1b	Hmga2		
Insr			
Ran			

Cardiovascular Disorders			
Aortic aneurysm	**Cardiomyopathy**	**Hypertension**	**Myocardial Infarction**
Klf15	Nr1h3	Esr1	Esr1
	Ncoa6	Nr3c2	Nr3c2
Aortic disease	Nr3c2	Nr2f2	Smad4
Smad4	Nr2f2	Smad4	Mef2a
	Ppard	Eno1	Nfkbia
Congestive heart failure	Ctnnb1	Nfkbia	Npm1
Nr3c2	Mef2a		**Coronary artery disease**
	Insr		
Calr	Npm1		Esr1
			Mef2a

Mood/Mental Disorders			
Schizophrenia	**Major Depressive Disorder**	**Major Affective Disorder**	**Anxiety Disorder**
Btg1	Pawr	Xbp1	Dnmt3b
Litaf			
Klf13			
Dnmt3b			
Gtf2i			
Nr4a2			
Gtf2ird1			
Trak1			
Erf			
Baz1b			

Cancer			
Breast Cancer	**Breast Neoplasms**	**Hepatocellular Carcinoma**	**Renal Cell Carcinoma**
Ncoa6	Ncoa1	Esr1	Smad4
Esr1	Esr1	Smad4	Ctnnb1
Smad4	Smad4	Ctnnb1	Mitf
Insr	Ctnnb1	Eno1	Pbrm1
Apex1	Dnmtr3b	Gmnn	Rb1
Breast adenocarcinoma	Rarb	Trim24	Insr
Mxi1	Rb1	Dnmt3b	**Pancreatic cancer**
Breast carcinoma	Eno1	Irf2	Ppard
Ncoa1	Elk3	Rb1	Smad4
Esr1	Nfkbia	Nfkbia	Ctnnb1
Apex1	Ppargc1b	Apex1	Pbrm1
			Rb1
			Nfkbia

Neurological Disorders		
Migraine	**Alzheimer's Disease**	**Stroke**
Esr1	Esr1	Ctnnb1
	Foxo3	Irf1
	Ctnnb1	
	Eno1	
	Insr	
	Pawr	

ACKNOWLEDGMENTS

This research was funded by the Ministry of Science and Technology in Taiwan, grant no. 107-2118-M-009-006-MY3.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data related to this article can be found at http://doi.org/10.1097/JCMA.0000000000000180.

REFERENCES

1. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc Natl Acad Sci U S A* 2014;**111**:16219–24.
2. Gan Y, Yang C, Tong X, Sun H, Cong Y, Yin X, et al. Shift work and diabetes mellitus: a meta-analysis of observational studies. *Occup Environ Med* 2015;**72**:72–8.
3. Wulff K, Dijk DJ, Middleton B, Foster RG, Joyce EM. Sleep and circadian rhythm disruption in schizophrenia. *Br J Psychiatry* 2012;**200**:308–16.
4. Knutsson A, Alfredsson L, Karlsson B, Akerstedt T, Fransson EI, Westerholm P, et al. Breast cancer among shift workers: results of the WOLF longitudinal cohort study. *Scand J Work Environ Health* 2013;**39**:170–7.
5. Almon RR, Yang E, Lai W, Androulakis IP, DuBois DC, Jusko WJ. Circadian variations in rat liver gene expression: relationships to drug actions. *J Pharmacol Exp Ther* 2008;**326**:700–16.
6. Almon RR, Yang E, Lai W, Androulakis IP, Ghimbovschi S, Hoffman EP, et al. Relationships between circadian rhythms and modulation of gene expression by glucocorticoids in skeletal muscle. *Am J Physiol Regul Integr Comp Physiol* 2008;**295**:R1031–47.
7. Atwood A, DeConde R, Wang SS, Mockler TC, Sabir JS, Ideker T, et al. Cell-autonomous circadian clock of hepatocytes drives rhythms in transcription and polyamine synthesis. *Proc Natl Acad Sci U S A* 2011;**108**:18560–5.
8. Bailey MJ, Coon SL, Carter DA, Humphries A, Kim JS, Shi Q, et al. Night/day changes in pineal expression of >600 genes: central role of adrenergic/camp signaling. *J Biol Chem* 2009;**284**:7606–22.
9. Vollmers C, Schmitz RJ, Nathanson J, Yeo G, Ecker JR, Panda S. Circadian oscillations of protein-coding and regulatory rnas in a highly dynamic mammalian liver epigenome. *Cell Metab* 2012;**16**:833–45.
10. Yoshitane H, Ozaki H, Terajima H, Du NH, Suzuki Y, Fujimori T, et al. CLOCK-controlled polyphonic regulation of circadian rhythms through canonical and noncanonical E-boxes. *Mol Cell Biol* 2014;**34**:1776–87.
11. Wang Y, Song L, Liu M, Ge R, Zhou Q, Liu W, et al. A proteomics landscape of circadian clock in mouse liver. *Nat Commun* 2018;**9**:1553.
12. Westermark PO, Herzel H. Mechanism for 12 hr rhythm generation by the circadian clock. *Cell Rep* 2013;**3**:1228–38.
13. Laing EE, Möller-Levet CS, Poh N, Santhi N, Archer SN, Dijk DJ. Blood transcriptome based biomarkers for human circadian phase. *Elife* 2017;**6**:e20214.
14. Korenčič A, Košir R, Bordyugov G, Lehmann R, Rozman D, Herzel H. Timing of circadian genes in mammalian tissues. *Sci Rep* 2014;**4**:5782.
15. Relógio A, Thomas P, Medina-Pérez P, Reischl S, Bervoets S, Gloc E, et al. Ras-mediated deregulation of the circadian clock in cancer. *Plos Genet* 2014;**10**:e1004338.
16. Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, et al. Harmonics of circadian gene transcription in mammals. *Plos Genet* 2009;**5**:e1000442.
17. Simak M, Yeang CH, Lu HH. Exploring candidate biological functions by Boolean function networks for saccharomyces cerevisiae. *Plos One* 2017;**12**:e0185475.
18. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44–57.
19. Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, et al. Animaltfdb 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res* 2015;**43**(Database issue):D76–81.
20. Han S, Han SS, Zhang R, Jain R, Shi H, Zhang L, et al. Circadian control of bile acid synthesis by a KLF15-fgf15 axis. *Nat Commun* 2015;**6**:7231.
21. Zhang L, Prosdocimo DA, Bai X, Fu C, Zhang R, Campbell F, et al. KLF15 establishes the landscape of diurnal expression in the heart. *Cell Rep* 2015;**13**:2368–75.
22. Rodriguez AR, de Sevilla Müller LP, Brecha NC. The RNA binding protein RBPMS is a selective marker of ganglion cells in the mammalian retina. *J Comp Neurol* 2014;**522**:1411–43.
23. Rao R, Yang Q, Orman MA, Berthiaume F, Ierapetritou MG, Androulakis IP. Burn trauma disrupts circadian rhythms in rat liver. *Int J Burns Trauma* 2016;**6**:12–25.
24. Lee J, Lee Y, Lee MJ, Park E, Kang SH, Chung CH, et al. Dual modification of BMAL1 by SUMO2/3 and ubiquitin promotes circadian activation of the CLOCK/BMAL1 complex. *Mol Cell Biol* 2008;**28**:6056–65.
25. Lee Y, Chun SK, Kim K. Sumoylation controls CLOCK-BMAL1-mediated clock resetting via CBP recruitment in nuclear transcriptional foci. *Biochim Biophys Acta* 2015;**1853**(10 Pt A):2697–708.
26. Prokkola JM, Nikinmaa M, Lewis M, Anttila K, Kanerva M, Ikkala K, et al. Cold temperature represses daily rhythms in the liver transcriptome of a stenothermal teleost under decreasing day length. *Journal of Experimental Biology* 2018;**221**:jeb.170670.
27. Wada T, Kang HS, Angers M, Gong H, Bhatia S, Khadem S, et al. Identification of oxysterol 7alpha-hydroxylase (cyp7b1) as a novel retinoid-related orphan receptor alpha (roralpha) (NR1F1) target gene and a functional cross-talk between roralpha and liver X receptor (NR1H3). *Mol Pharmacol* 2008;**73**:891–9.
28. Blattner MS, Mahoney MM. Estrogen receptor 1 modulates circadian rhythms in adult female mice. *Chronobiol Int* 2014;**31**:637–44.
29. Kobayashi H, Yoshida S, Sun YJ, Shirasawa N, Naito A. Gastric 17β-estradiol in portal vein and liver esr1 make a circadian rhythm in systemic circulation in male rats. *Endocrine* 2016;**53**:565–73.
30. Xiao L, Chang AK, Zang MX, Bi H, Li S, Wang M, et al. Induction of the CLOCK gene by E2-erα signaling promotes the proliferation of breast cancer cells. *Plos One* 2014;**9**:e95878.
31. Uyhelji HA, Kupfer DM, White VL, Jackson ML, Van Dongen HPA, Burian DM. Exploring gene expression biomarker candidates for neurobehavioral impairment from total sleep deprivation. *BMC Genomics* 2018;**19**:341.
32. Gumz ML, Stow LR, Lynch IJ, Greenlee MM, Rudin A, Cain BD, et al. The circadian clock protein Period 1 regulates expression of the renal epithelial sodium channel in mice. *J Clin Invest* 2009;**119**:2423–34.
33. Zuber AM, Centeno G, Pradervand S, Nikolaeva S, Maquelin L, Cardinaux L, et al. Molecular clock is involved in predictive circadian adjustment of renal function. *Proc Natl Acad Sci U S A* 2009;**106**:16523–8.
34. Gimble JM, Ptitsyn AA, Goh BC, Hebert T, Yu G, Wu X, et al. Delta sleep-inducing peptide and glucocorticoid-induced leucine zipper: potential links between circadian mechanisms and obesity? *Obes Rev* 2009;**10** (**Suppl 2**):46–51.
35. Kwapis JL, Alaghband Y, Kramár EA, López AJ, Vogel Ciernia A, White AO, et al. Epigenetic regulation of the circadian gene Per1 contributes to age-related changes in hippocampal memory. *Nat Commun* 2018;**9**:3323.
36. Chaves I, van der Horst GT, Schellevis R, Nijman RM, Koerkamp MG, Holstege FC, et al. Insulin-FOXO3 signaling modulates circadian rhythms via regulation of clock transcription. *Curr Biol* 2014;**24**:1248–55.
37. Heckman M. *The circadian clock and Nr2f2 (COUP-TF) regulation of adipogenesis*. PhD dissertation, The University of Rochester. 2009.
38. Liu S, Brown JD, Stanya KJ, Homan E, Leidl M, Inouye K, et al. A diurnal serum lipid integrates hepatic lipogenesis and peripheral fatty acid utilization. *Nature* 2013;**502**:550–4.
39. Pappa KI, Gazouli M, Anastasiou E, Iliodromiti Z, Antsaklis A, Anagnou NP. Circadian clock gene expression is impaired in gestational diabetes mellitus. *Gynecol Endocrinol* 2013;**29**:331–5.
40. Sloin HE, Ruggiero G, Rubinstein A, Smadja Storz S, Foulkes NS, Gothilf Y. Interactions between the circadian clock and TGF-β signaling pathway in zebrafish. *Plos One* 2018;**13**:e0199777.
41. Mazzoccoli G, Vinciguerra M, Papa G, Piepoli A. Circadian clock circuitry in colorectal cancer. *World J Gastroenterol* 2014;**20**:4197–207.
42. Shimoyama M, De Pons J, Hayman GT, Laulederkind SJ, Liu W, Nigam R, et al. The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res* 2015;**43**:D743–50.
43. Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ, the Mouse Genome Database Group. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res* 2018;**46**:D836–42.
44. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, et al. The NCBI BioSystems database. *Nucleic Acids Res* 2010;**38**:D492–6.
45. Li S, Shui K, Zhang Y, Lv Y, Deng W, Ullah S, et al. CGDB: a database of circadian genes in eukaryotes. *Nucleic Acids Res* 2017;**45**(D1):D397–403.

46. Pizarro A, Hayer K, Lahens NF, Hogenesch JB. Circadb: a database of mammalian circadian gene expression profiles. *Nucleic Acids Res* 2013;**41**(Database issue):D1009–13.
47. Hardman JA, Tobin DJ, Haslam IS, Farjo N, Farjo B, Al-Nuaimi Y, et al. The peripheral clock regulates human pigmentation. *J Invest Dermatol* 2015;**135**:1053–64.
48. Maekawa F, Shimba S, Takumi S, Sano T, Suzuki T, Bao J, et al. Diurnal expression of dnmt3b mrna in mouse liver is regulated by feeding and hepatic clockwork. *Epigenetics* 2012;**7**:1046–56.
49. Lee YJ, Han DH, Pak YK, Cho SH. Circadian regulation of low density lipoprotein receptor promoter activity by CLOCK/BMAL1, hes1 and hes6. *Exp Mol Med* 2012;**44**:642–52.
50. Navigatore-Fonzo LS, Golini RL, Ponce IT, Delgado SM, Plateo-Pignatari MG, Gimenez MS, et al. Retinoic acid receptors move in time with the clock in the hippocampus. Effect of a vitamin-A-deficient diet. *J Nutr Biochem* 2013;**24**:859–67.
51. Breen MS, Uhlmann A, Nday CM, Glatt SJ, Mitt M, Metsalpu A, et al. Candidate gene networks and blood biomarkers of methamphetamine-associated psychosis: an integrative RNA-sequencing report. *Transl Psychiatry* 2016;**6**:e802.
52. Schneider T, Skitt Z, Liu Y, Deacon RM, Flint J, Karmiloff-Smith A, et al. Anxious, hypoactive phenotype combined with motor deficits in Gtf2ird1 null mouse model relevant to Williams syndrome. *Behav Brain Res* 2012;**233**:458–73.
53. Di-Luoffo M, Daems C, Bergeron F, Tremblay JJ. Novel targets for the transcription factors MEF2 in MA-10 Leydig cells. *Biol Reprod* 2015;**93**:9.
54. Sonoda J, Mehl IR, Chong LW, Nofsinger RR, Evans RM. PGC-1beta controls mitochondrial metabolism to modulate circadian activity, adaptive thermogenesis, and hepatic steatosis. *Proc Natl Acad Sci U S A* 2007;**104**:5223–8.
55. Chiu JC, Ko HW, Edery I. NEMO/NLK phosphorylates PERIOD to initiate a time-delay phosphorylation circuit that sets circadian clock speed. *Cell* 2011;**145**:357–70.
56. Pivovarova O, Gögebakan Ö, Sucher S, Groth J, Murahovschi V, Kessler K, et al. Regulation of the clock gene expression in human adipose tissue by weight loss. *Int J Obes (Lond)* 2016;**40**:899–906.
57. Podobed PS, Kirby GM, Martino TA. *Circadian proteomics and its unique advantage for discovery of biomarkers of heart disease*. Proteomics. Human Diseases and Protein Functions. 2012
58. Hu Y, Shmygelska A, Tran D, Eriksson N, Tung JY, Hinds DA. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat Commun* 2016;**7**:10448.
59. Zhao X, Cho H, Yu RT, Atkins AR, Downes M, Evans RM. Nuclear receptors rock around the clock. *EMBO Rep* 2014;**15**:518–28.
60. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* 2018;**19**: e46255.