Analysis and synthesis of high-amplitude *Cis*-elements in the mammalian circadian clock

Yuichi Kumaki*[†‡], Maki Ukai-Tadenuma*, Ken-ichiro D. Uno[§], Junko Nishio[§], Koh-hei Masumoto*[¶], Mamoru Nagano[¶], Takashi Komori[†], Yasufumi Shigeyoshi[¶], John B. Hogenesch[‖], and Hiroki R. Ueda*[§]**

*Laboratory for Systems Biology and [§]Functional Genomics Unit, Center for Developmental Biology, RIKEN, 2-2-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan; [†]INTEC Systems Institute, Inc., 1-3-3 Shinsuna, Koto-ku, Tokyo 136-0075, Japan; [¶]Department of Anatomy and Neurobiology, Kinki University School of Medicine, 377-2 Ohno-Higashi, Osaka-Sayama, Osaka 589-8511, Japan; and [‖]Institute for Translational Medicine and Therapeutics and the Department of Pharmacology, University of Pennsylvania School of Medicine, 810 Biomedical Research Building II/III, 421 Curie Boulevard, Philadelphia, PA 19104-6160

Edited by Joseph S. Takahashi, Northwestern University, Evanston, IL, and approved August 15, 2008 (received for review March 16, 2008)

Mammalian circadian clocks consist of regulatory loops mediated by Clock/Bmal1-binding elements, DBP/E4BP4 binding elements, and RevErbA/ROR binding elements. As a step toward system-level understanding of the dynamic transcriptional regulation of the oscillator, we constructed and used a mammalian promoter/enhancer database (http://promoter.cdb.riken.jp/) with computational models of the Clock/Bmal1-binding elements, DBP/E4BP4 binding elements, and RevErbA/ROR binding elements to predict new targets of the clock and subsequently validated these targets at the level of the cell and organism. We further demonstrated the predictive nature of these models by generating and testing synthetic regulatory elements that do not occur in nature and showed that these elements produced high-amplitude circadian gene regulation. Biochemical experiments to characterize these synthetic elements revealed the importance of the affinity balance between transactivators and transrepressors in generating high-amplitude circadian transcriptional output. These results highlight the power of comparative genomics approaches for system-level identification and knowledge-based design of dynamic regulatory circuits.

comparative genomics | promoter and enhancer database | synthetic biology | systems biology | transcription

The rapidly expanding number of sequenced mammalian genomes (1–3), annotated and cloned full-length cDNAs (4–6), transcriptional starts sites (TSSs) (7–9) and transcription factor binding sites (TFBSs) (10–12) has provided new opportunities to unravel the control of dynamic transcriptional programs. Comparative genomics approaches applying these resources have been used to identify target genes of specific biological pathways. These efforts used consensus sequence searches (13, 14), positional weight matrices (15), hidden Markov models (HMMs) (16). and specifically tailored algorithms (17, 18) to define candidate response elements and target genes in raw genomic sequence. Additionally, post hoc analysis employing evolutionary conservation (15, 16, 18) together with positional information of TSSs (15) and/or translational start sites (16) has helped to further define candidate elements and genes and greatly expanded our knowledge of transcriptional output regulation.

The mammalian circadian clock is an ideal system to apply these tools as it consists of integrated transcriptional regulatory loops that direct output through at least three types of transcriptional regulatory elements, the Clock/Bmal1-binding elements (E-box) (CACGTG) (19–21), DBP/E4BP4 binding elements (D-box) (TTATG[T/C]AA) (21–23), and RevErbA/ROR binding elements (RRE) ([A/T]A[A/T]NT[A/G]GGTCA) (15, 21, 24, 25). Several groups, including our own, have shown that approximately 5–10% of mammalian genes display circadian expression in central and peripheral clock tissues (26). However, for the most part, the transcriptional regulation of these thousands of clock-controlled genes has remained uncharacterized. We and others have used comparative genomics approaches to analyze E-box (21, 27, 28), D-box (21), and RRE (15, 21), highlighting the importance of both their core consensus and flanking sequences (15, 21, 27, 28) in circadian gene control. In this study, we further extend comparative genomics approaches toward a system-level understanding of the dynamic transcriptional regulations of the mammalian circadian clock.

Results and Discussion

Prediction of Direct Clock Targets Through Utilization of the Mammalian Promoter/Enhancer Database. To generate a resource that facilitates identification of clock-controlled genes, we constructed a mammalian promoter/enhancer database (http://promoter.cdb.riken.jp/) by integrating information sources such as conserved non-coding regions, TSSs and TFBSs [supporting information (SI) Fig. S1 and *SI Appendix*]. Although excellent and similar databases exist such as DBTSS (8), CisView (29) and ECRbase (30), none were tailored to specifically identify clock gene targets and having local control of the database facilitated manipulation of the underlying data (see also *SI Appendix*). We then developed a comparative genomics strategy employing this database and profile HMMs using the HMMER software package (31). Profile HMMs are powerful tools to extract the statistical properties of input sequences by representing multiple sequences as a transition probability matrix marching from one position to the neighboring position. HMMs were built and calibrated on known functional clock-controlled elements experimentally verified in our previous (15, 21) or current studies (Fig. S2 and Table S1), consisting of 12 E-boxes, 10 D-boxes and 15 RREs (Table S2). Profile HMM searches to identify new clock-controlled elements from conserved non-coding regions between human and mouse identified 1,108 E-boxes, 2,314 D-boxes, and 3,288 RREs candidate elements (see the circadian section of the mammalian promoter/enhancer database: http://promoter.cdb.riken.jp/circadian.html for element lists). To set appropriate reporting thresholds, we used the match scores of known functional clock-controlled elements (*Material and Methods*). Predicted clock-controlled elements exhibited an un-biased distribution of chromosomal position spread over whole mouse genome (Fig. 1*A* and http://promoter.cdb.riken.jp/circadian.html).

Author contributions: J.B.H. and H.R.U. designed research; Y.K., M.U., K.D.U., J.N., K.M., M.N., and Y.S. performed research; Y.K. and T.K. contributed new analytic tools; Y.K. analyzed data; and Y.K., J.B.H., and H.R.U. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[‡]Present address: Laboratory for Mammalian Epigenetic Studies, Center for Developmental Biology, RIKEN, 2-2-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan.

**To whom correspondence should be addressed. E-mail: uedah-tky@umin.ac.jp.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802636105/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0802636105

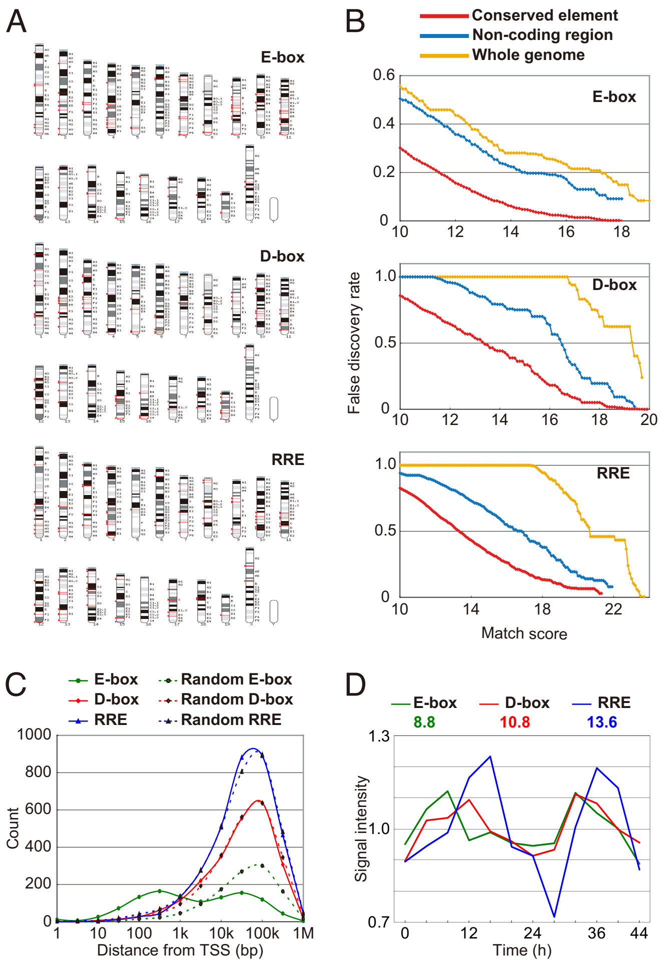


Fig. 1. Computational prediction of clock-controlled elements using HMMs. (*A*) Chromosomal distributions of predicted conserved clock-controlled elements of conserved non-coding regions mapped on the mouse genome. Chromosomal positions of the 100 most significant hits for E-boxes, D-boxes, and RREs are shown (red). (*B*) Plots of false discovery rates (FDRs) against match scores of HMM searches in three conditions: (1) searches for conserved elements within conserved non-coding regions (red, "Conserved element"), (2) searches for mouse elements within conserved or non-conserved non-coding regions (blue, "Non-coding region") and (3) searches in the entire genome relaxing both element conservation and search space (orange, "Whole genome"). FDRs in "Conserved elements" search are plotted against the average match score of human and mouse elements. (*C*) The distributions of distance from transcriptional starts sites (TSSs) for predicted conserved clock-controlled elements of conserved non-coding regions (1,108 E-boxes, 2,314 D-boxes, and 3,288 RREs). The E-box displays a biased distribution of distance from TSSs, while the D-box and RRE show unbiased, near random distributions ("Random E-box," "Random D-box," and "Random RRE," see also *SI Appendix*). (*D*) The average expression of transcripts harboring each element exhibit circadian rhythms in the liver. The average expressions of 36 genes with E-boxes, 29 genes with D-boxes, and 34 genes with RREs exhibited significant circadian oscillations ($P = 0.01$ for E-box, $P = 0.0005$ for D-box, and $P = 0.001$ for RRE). Data were normalized so that the average signal intensity over 12-point time courses is 1.0. Estimated peak times of circadian oscillation were also indicated.

However, the match score on its own does not give a good estimate of accuracy and false discovery from HMM searches. To better estimate the prediction for each HMM, we searched each against randomized genome sequence to generate a background distribution of false positive occurrences (see *SI Appendix* for detail). We found that the value of the false discovery rate (FDR) is inversely proportional to the match score of the HMM, which is a representation of the statistical significance of the candidate element (Fig. 1*B*). Importantly, we found the accuracy of the HMM-based prediction as measured by the FDR is dependent on the initial search conditions. HMM searches in conserved non-coding regions (the original condition) had the lowest FDR, while higher rates were observed in conserved or non-conserved non-coding regions, or in searches of raw genome sequence (Fig. 1*B*). These results demonstrate the value of using human/mouse conservation and a confined search space for generation of the most accurate response element predictions.

Several intriguing features resulted from this analysis. Interestingly, like cAMP-responsive elements (CREs) (16), putative E-box displayed a biased distance distribution from TSSs, while putative D-box and RRE had random localization distributions in the genome and were not more likely to be near TSSs (Fig. 1*C*). This result is consistent with and extends an earlier report that described the preferential localization of CpG containing transcription factor binding sites including the E-box and CRE to proximal promoter regions of housekeeping genes (32). In addition to housekeeping genes, we see circadian E-box sequences present in many genes with specific functions such as enzymes and signaling molecules.

Gratifyingly, we noted a significant enrichment of previously identified clock-controlled genes (15, 33) in our predicted clock-controlled elements (the 100 most significant sequences for each HMM search, E-box, D-box, and RRE, respectively). After removing the 21 clock-controlled genes used for HMM generation and training, we found an additional 19 putative clock-controlled genes (out of the 6,195 genes common in our mammalian promoter/enhancer database and U74 mouse microarray) versus the expected 10.67 that would have arisen from chance, a significant enrichment of clock-controlled genes ($P < 0.01$, see also *SI Appendix*).

The presence of statistically significant clock-controlled elements in the promoters of these genes suggests that their message levels may oscillate. To examine this possibility, we selected the 100 most significant putative clock-controlled elements for each model (E-box, D-box, and RRE) and determined their gene expression levels from previously obtained liver data (15). The average expression of 36 E-box containing genes, 29 D-box genes, and 34 RRE genes exhibited significant circadian oscillations ($P = 0.01$ for E-box, $P = 0.0005$ for D-box, and $P = 0.001$ for RRE) with a surprisingly consistent peak time of expression (E-box = 8.8, D-box = 10.8, and RRE = 13.6, Fig. 1*D*). Taken in sum, these data show that our HMM models and strategy identifies elements and genes that oscillate with a circadian period with peak phases of expression that are consistent with the previously reported literature.

***In Vitro* Validation of Putative Clock Controlled Genes.** To provide further validation of these predictions, we used an *in vitro* system of the autonomous circadian clock to empirically test candidate elements in circadian transcriptional output assays. In brief, we used a cell culture system (15, 34) that allowed the monitoring of circadian transcriptional dynamics using a destabilized luciferase reporter (dLuc) driven by known or putative clock-controlled response elements (Fig. 2*A*). We selected the ten most significant sequences for each HMM search, E-box, D-box, and RRE respectively, and located within 1kb of the TSS (Table S3). We constructed reporter vectors in which three predicted elements were inserted in front of the SV40 basic promoter driving a dLuc reporter (see *Material and Methods*). We transfected these constructs into cultured NIH 3T3 fibroblasts, stimulated them with forskolin to synchronize circadian rhythms of individual cells, and measured the sum of their transcriptional activity by monitoring bioluminescence over several days; 40% of putative E-boxes, 70% of D-boxes, and 60% of RREs generated strong circadian reporter gene activity ($P < 0.01$ and high-amplitude) in phase with the *Per1* E-box, *Per3* D-box, and *Arntl*

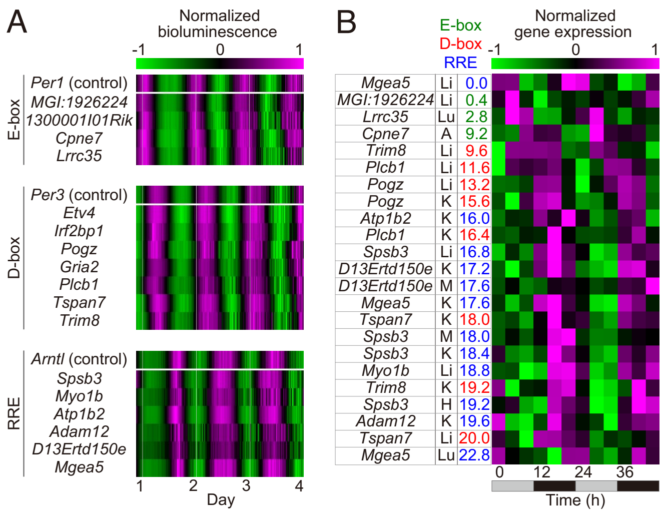


Fig. 2. Experimental validation of HMM-based predictions at cellular and organismal levels. (*A*) Circadian rhythms of bioluminescence from the predicted clock-controlled elements fused to the SV40 basic promoter driving a dLuc reporter in NIH 3T3 fibroblasts. Three known clock-controlled elements from clock genes (E-box of *Per1*, D-box of *Per3*, and RRE of *Arntl*) are used as positive controls. The bioluminescence data were detrended in baseline and amplitude and normalized so that their maximum, minimum, and average were set to 1, −1, and 0, respectively. The colors in descending order from magenta to black to green represent the detrended bioluminescence. Columns represent time points and rows represent the predicted elements on the designated genes. (*B*) Circadian rhythms of temporal mRNA expression profiles of the predicted clock-controlled genes in mouse seven tissues ('A', 'B', 'H', 'K', 'Li', 'Lu' and 'M' for aorta, bone, heart, kidney, liver, lung, and muscle, respectively). An estimated peak time with color of type of predicted clock-controlled element (green, red, and blue for E-box, D-box, and RRE, respectively) is also indicated. The colors in descending order from magenta to black to green represent the normalized data (the average and standard deviation over 12-point time courses are 0.0 and 1.0, respectively). Columns represent time points, and rows represent the predicted clock-controlled genes in the designated tissues.

RRE, respectively (Fig. 2*A* and Fig. S3 *A*–*D*). The remaining sequences generated weak, low amplitude circadian transcriptional activity or were arrhythmic (Fig. S3 *E*–*G*). To supplement our observed 40% prediction success, we constructed 14 reporters containing conserved low-scoring E-boxes and found only one exhibited high-amplitude oscillations (Fig. S3*H* and Table S3). This result indicates that our observed 40% prediction success is suggestively higher than expected ($P = 0.075$, Fisher's exact test). Taken in sum, these results demonstrate utility of this approach in finding elements within structural genes that dictate rhythmic transcription.

If these *in vitro* validated 17 elements (4 E-boxes, 7 D-boxes, and 6 RREs) play a prominent role in gene regulation *in vivo*, we would predict that the endogenous transcripts for these genes would likely oscillate in a circadian fashion. To test this, we harvested mRNA from seven tissues (aorta, bone, heart, kidney, liver, lung, and muscle) isolated from mice entrained to a 12:12 light:dark cycle and then released to free run in constant darkness. Using quantitative PCR assays, we measured expression profiles from our predicted clock-controlled elements, and evaluated their rhythmicity using a statistical method based on analysis of variance (ANOVA) followed by curve fitting to a cosine wave. These experiments revealed that circadian expression profiles ($P < 0.03$) for 13 genes (76%): 3 E-box controlled genes, 4 D-box controlled genes and 6 RRE controlled genes, respectively, with a consistent order of peak time (4.1, 15.5, and 18.8 for mean value of the peak time of putative E-box, D-box, and RRE-controlled genes, respectively) (Fig. 2*B*; See also http://promoter.cdb.riken.jp/circadian.html for detailed data). For those genes that did not confirm circadian rhythmicity, the average level of expression was lower, implying mRNA detection was limiting for these genes. Collectively, these *in vitro* and *in vivo* experiments suggest that many predicted E-box, RRE, and D-box containing genes are *bona fide* first-order clock-controlled genes.

Design and Validation of the Synthetic Regulatory Elements. One of the goals of systems biology is the synthesis of knowledge and the generation of testable (and tested) hypotheses. We reasoned that if our HMMs truly represented the functional response elements of these three transcription factor complexes, then synthetic regulatory elements derived from these models should mediate rhythmic transcription as well. To test this idea, we emitted sequences from the E-box, D-box, and RRE models, respectively, and filtered out those that naturally exist in either the human or mouse genomes. Furthermore, to not unduly focus our attention on outliers, we required that all candidates adhere to the consensus rules for each element, CACGTG for the E-box (19), TTATGTAA for the D-box (22), and [A/T]A[A/T]NT[A/G]GGTCA for the RRE (24). For the remaining sequences, we chose each one of the highest and lowest scoring synthetic representatives for three types of elements and named them "high-scoring" and "low-scoring" elements, respectively (Fig. 3*A*). We tested these elements in a synthetic reporter system as above (Fig. 3*B*). All three "high-scoring" elements showed high-amplitude circadian transcriptional activity equivalent to known elements from canonical clock genes (E-box of *Per1*, D-box of *Per3*, and RRE of *Arntl* are used as 1.0, respectively) (21) (Fig. 3*C*). On the other hand, the "low-scoring" elements emitted from the HMMs showed very low- amplitude transcriptional activity, despite the presence of "consensus" E-box, RRE, or D-box core sequences (Fig. 3*C*). These results show the utility of this comparative genomics approach in synthetic design of dynamic *cis*-acting elements, as well as highlight the contribution of flanking sequences in generating high-amplitude rhythmicity.

Investigating the Contribution of Flanking Sequences. Using these synthetic elements, we next attempted to explore the contribution of E-box flanking regions to identify critical residues that modulate amplitude and rhythmicity. We clustered their nucleotide sequences, and, interestingly, found two patterns of high-amplitude E-box flanking sequences adjacent to the core CACGTG element (Fig. S4). However, these positions do not absolutely dictate high-amplitude rhythmicity, as some elements that meet these rules exhibit lower-amplitude oscillations, possibly because they exhibit much higher GC content. In either case, these experimental results also imply that amplitude information is encoded in specific residues adjacent to the core consensus element and further strengthen the previous reports by other groups on the importance of flanking sequence of E-box (27, 28, 35, 36). Interestingly, the identified patterns in this study partly overlap with the computational models based on the evolutionarily conserved E-box structure from insects to mammals (27).

High-Amplitude Oscillations Require Appropriate Affinity Balance Between Activators and Repressors. To explore the properties of these elements that result in high amplitude oscillations, we took a simplified molecular modeling and experimental approach. First, we assumed concentrations of activators and repressors were within similar ranges (see also *SI Appendix* Discussion in more general cases). We further hypothesized that flanking region DNA sequence impacted DNA-binding affinity of clock gene regulators and therefore altered amplitude. We further hypothesized that tightly binding sequences would have higher amplitudes of circadian oscillation. To test this notion, we analyzed the DNA-binding affinity of activators and repressors

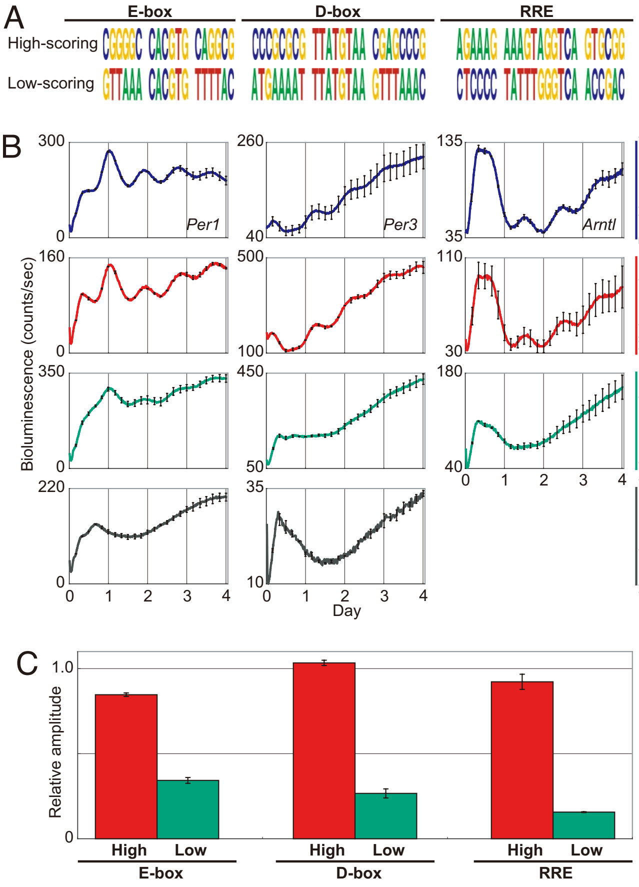


Fig. 3. Computational design and experimental characterization of high-amplitude circadian transcriptional activity. (*A*) HMM-based design of the high- and low-scoring elements. Match scores of synthetic elements are 19.84 (high-scoring E-box), −1.36 (low-scoring E-box), 22.17 (high-scoring D-box), −4.33 (low-scoring D-box), 24.45 (high-scoring RRE) and −0.46 (low-scoring RRE), respectively. All elements were filtered to identify those that do not exist in either the human or mouse genomes. (*B*) Bioluminescence from synthetic elements inserted into the SV40 basic promoter driving a dLuc reporter (SV40-dLuc) in NIH 3T3 fibroblasts. SV40-dLuc containing known clock-controlled elements (E-box of *Per1*, D-box of *Per3*, or RRE of *Arntl*) and SV40-dLuc with no insert are the positive and negative controls, respectively. The negative control of the D-box element was also used for the RRE as experiments for these elements were performed at the same time. (*C*) The amplitude of bioluminescence activity driven by synthetic elements relative to that of positive controls. The relative amplitude of positive controls is 1.0. Error bars indicate SEM determined from independent experimental duplicates.

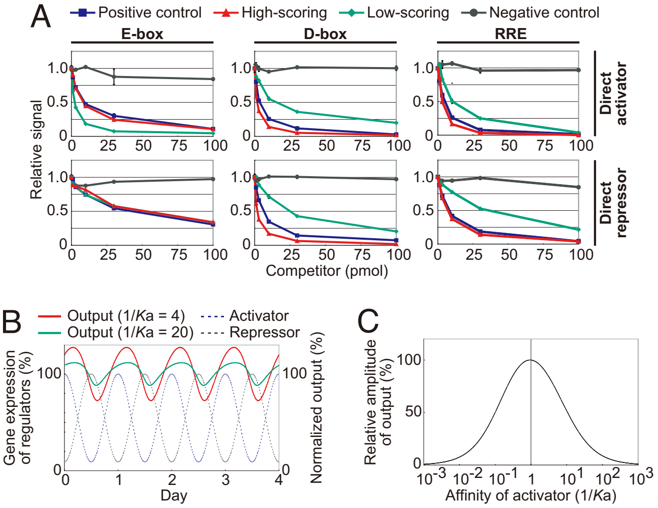


Fig. 4. Defining the relationship between affinity and amplitude. (*A*) Binding affinity between synthetic elements and DNA binding activators or repressors was detected by competitive DNA binding assays. The binding between labeled oligonucleotides of positive control elements (10 pmol) and transcription factors were competed by each of unlabeled oligonucleotides (0, 1, 3, 10, 30, and 100 pmol) for positive control elements (blue), high-scoring elements (red), low-scoring elements (green), or negative control elements (black). Known clock-controlled elements (E-box of *Per1*, D-box of *Per3*, or RRE of *Arntl*) and mutated clock-controlled elements are the positive and negative controls, respectively. *Arntl*/*Clock*, *Dbp*, *Rora* were used as DNA binding activator of E-box, D-box, and RRE, respectively. *Bhlhb2*, *Nfil3*, *Nr1d1*, were used as DNA binding repressors of E-box, D-box, and RRE, respectively. The relative signal without competitor is 1.0. Error bars indicate SEM determined from independent experimental duplicates. (*B* and *C*) *In silico* analysis of affinity to amplitude mechanism. Gene expression of idealized transcriptional activators (blue dotted line) and repressors (gray dotted line) and the normalized output of different strengths of activator binding affinity (weaker affinity 1/*Ka* = 4 is red line and stronger affinity 1/*Ka* = 20 is green line) are indicated (*B*). The relative amplitude of oscillation of the output plotted against the strength of activator binding affinity (*C*). Amplitude was normalized so that the maximum value is set to 100%.

to these response elements using competitive binding assays (Fig. 4*A* and Fig. S5*A*). For the D-box and RRE elements, "high-scoring" elements showed approximately the same DNA-binding affinity for their activators and repressors, while "low-scoring" elements of D-box and RRE showed relatively weak affinity, confirming this hypothesis. Surprisingly, in the case of E-box, "low-scoring" sequences had a *higher* affinity for the *Arntl*/*Clock* activator complex (4.8 times higher than positive control; Fig. S5*A*) than "high-scoring" sequences or the positive control, whereas the "low-scoring" E-box sequence showed approximately the same affinity to the *Bhlhb2* repressor.

To assist in interpreting these results, we used *in silico* modeling (Fig. 4 *B* and *C*) and treated affinity of activators and repressors as parameters and amplitude as the output of the model. Interestingly, this analysis showed that a high affinity activator complex coupled with a normal affinity repressor complex capitulated lower amplitude rhythms (Fig. 4*B*), suggesting that the enhanced retention of an activator alone promotes its saturation on a promoter and consequently dampens amplitude in competition-based models. Further, *in silico* analysis showed not only affinity strength of the activator and repressor (Fig. S5*B*) but also the appropriate affinity balance of activators and repressors is necessary for high-amplitude circadian oscillations (Fig. 4*C*; See also *SI Appendix* Discussions for *in silico* modeling).

Supporting this notion, a new clock gene, *clockwork orange* (*cwo*, a fly ortholog of mammalian *Bhlhb2* and *Bhlhb3*) was recently reported to directly suppress gene expression of several clock genes through E-box elements (37–39). Quantitative and qualitative impairment of *cwo* revealed an important role of this transcriptional repressor for high-amplitude oscillation of the *Drosophila* circadian clock. The findings in this report, along with the studies of the *cwo* gene, collectively show that a competitive balance between direct activator(s) and direct repressor(s) for the E-box element is important for driving high-amplitude oscillations of circadian output genes. In addition to this *in vivo* biological evidence in the fly, we listed the evolutionary conserved "low-scoring" E-boxes in the mammalian genome (predicted low-amplitude) as candidates for unbalanced affinities. Interestingly, this list includes E-boxes on core clock genes (*Cry2*, *Bhlhb3*, *Nr1d1*, and *Rorc*) and some low-amplitude clock-controlled genes such as *Id2*, and to promote follow-up, is available at http://promoter.cdb.riken.jp/circadian.html.

What could explain the discrepancy between E-box motifs and other circadian response elements? We hypothesized that these differences might be encoded at the protein sequence level of the DNA-binding domains of activators and repressors. Interestingly, the DNA binding domains of transactivators and transrepressors of the RRE and D-box are more similar to each other (65% identity and 81% homology for RRE regulators, and 44% identity and 69% homology for D-box regulators) than those of E-box transactivators and transrepressor (22∼30% identity and 55∼57% homology) (Table S4). Based on these findings, we speculate that the DNA-binding domains for transactivators and transrepressors of the RRE and D-box have evolved similar affinities. In contrast, the evolutionarily and structurally divergent regulators of E-boxes, 108 bHLH proteins including several families of activators and repressors, as well as the unrelated *period* and *cryptochrome* gene families, may have required the co-evolution of specific DNA-binding domains and E-box sequences with specific flanking regions to generate higher amplitude rhythmicity.

Conclusion

In summary, we have applied a comparative genomics strategy to the understanding of a dynamic transcriptional regulatory system, the mammalian circadian clock. Our informatics strategy employs a model-based search with excellent statistical properties, the evolutionary conservation of putative transcriptional regulatory elements across mouse and human non-coding regions, and statistical evaluation of false discovery rates in each prediction. Experimental validation of this strategy *in vitro* and *in vivo* using real-time monitoring of transcriptional activity and quantitative PCR assay has led to the identification of dozens of novel clock-controlled genes and the elements that likely dictate their rhythmicity. High-scoring conserved E-boxes (mean HMM-score = 16.15) had a 40% rate of validation, while low-scoring conserved E-boxes (mean HMM-score = 2.5143) had a 7.1% probability of generating high-amplitude rhythmicity in reporter assays. Linear interpolation from these two numbers generates an estimate of approximately 347 novel conserved E-boxes that likely confer circadian rhythmicity (see also *SI Appendix*). Moreover, to demonstrate their predictive nature, we have taken these *in silico* models and designed synthetic elements that exhibit high-amplitude transcriptional rhythmicity as well as the best canonical regulatory elements. Furthermore, experimental measurement and *in silico* analysis of affinity of regulators to synthetic elements revealed the importance of the appropriate affinity balance between activators and repressors for high-amplitude rhythmicity. Surprisingly, for E-box sequences, lower affinity DNA element generates higher amplitude rhythms. The experimental, analytical, and synthetic approaches discussed here are especially timely as genomics tools are increasingly uncovering the complexity and flexibility of transcriptional regulatory circuits. We predict the general themes and resources reported here will enhance understanding of the biology mediated by complex and dynamic transcriptional regulation including the mammalian circadian clock.

Materials and Methods

Detailed information on the construction of the mammalian promoter/enhancer database, determination of distance from TSSs for natural and randomly positioned elements, calculation of FDR for putative elements, animals, genome sequences, oligonucleotide sequences, plasmid constructions, quantitative PCR, rhythmicity analysis of real-time bioluminescence data, amplitude analysis of real-time bioluminescence data, rhythmicity analysis of quantitative PCR data, over representation analysis of clock-controlled genes, estimation of the number of high-amplitude E-boxes, microarray expression data analysis of genes with predicted clock-controlled elements, affinity analysis of competitive DNA binding data, and *in silico* analysis of affinity to amplitude mechanism are available in *SI Appendix*.

Real-Time Circadian Reporter Assays. Real-time circadian assays were performed as previously described (40) with the following modifications. NIH 3T3 cells (American Type Culture Collection) were grown in DMEM (Invitrogen) supplemented with 10% FBS (JRH Biosciences) and antibiotics (25 units ml^{-1} penicillin, 25 μg ml^{-1} streptomycin; Invitrogen). Cells were plated at 5×10^4 cells per well in 24-well plates 24 h before transfection. Cells were transfected with 0.32 μg of plasmids in total (0.13 μg reporter plasmid and 0.19 μg empty plasmid) per well using FuGENE6 (Roche Applied Science) according to the manufacturer's instructions. After 72 h, medium in each well was replaced with 500 μl of culture medium (DMEM/10% FBS) supplemented with 10 mM Hepes (pH 7.2), 0.1 mM luciferin (Promega), antibiotics and 0.01 μM forskolin (nacalai tesque). Bioluminescence was measured with photomultiplier tube (PMT) detector assemblies (Hamamatsu Photonics). The modules and cultures were maintained in a darkroom at 30 °C and interfaced with computers for continuous data acquisition until 96 h after forskolin stimulation. Photons were counted 2 min at 24-min intervals.

Construction, Search, and Design of Putative *cis*-Acting Elements. A HMM is a statistical model in which the target system is assumed to be a Markov process with unknown parameters. A HMM describes a probability distribution over input training sequences, i.e., probabilities of the state transition and emission. The extracted model can be used to find the probability of query sequence that is a product of all transition and emission probabilities at training sequences. Nucleotide sequences for known functional clock-controlled elements, 12 E-boxes (18 bp), 10 D-boxes (24 bp), and 15 RREs (23 bp), experimentally verified in previous (21) and current studies (Table S1 and Fig. S2), were used as a training dataset to construct HMMs. We also attempted to construct an HMM for the E′-box, but were unable (i.e., positive controls exhibited poor scores) because of the small number of experimentally validated E′-box (only three: *Per2*, *Bhlhb3*, and *Cry1*) and the relatively short core consensus sequence of the E-box. Thus, we did not use an E′-box HMM in this study. The lengths of these known functional elements were based on our previous experiments (21) and these were sufficient to produce circadian transcriptional activity in circadian reporter assays. These sites were aligned without gaps according to the direction of consensus sequences (TTATG[T/C]AA for the D-box; ref. 22), [A/T]A[A/T]NT[A/G]GGTCA for the RRE; ref. 24). Because the consensus sequence for E-box is palindromic (CACGTG; ref. 19), we generated all possible alignments by changing sequence directions (forward and reverse) and selected one alignment as described below. These alignments were used to build HMMs using hmmt program in the HMMER 1.8.4 software package (31) with default parameters (using sim annealing, starting kT for sim annealing run as 5.0, and multiplier for sim annealing as 0.95). We used the older version 1.8.4 package (the current version is 2.3.2) in this study because the version 2 series was optimized for analysis of protein sequences. Following construction, models were used to search genomic regions for putative clock-controlled elements using the hmmls program with default parameters (by using threshold matches score to report as 0) except use '-c' option only for bidirectionally search. The average score was used in the search for the conserved elements between human and mouse. To select only one alignment for each E-box, we constructed 2048 HMMs of all possible alignments, and calculated match scores of 12 known E-box sequences in directional HMMER search. We selected the alignment that generated the highest average match score for further work.

To design the ''high-scoring'' and ''low-scoring'' sequence of clock-controlled elements, bidirectional HMMER searches were performed against all possible sequences of the same lengths as training dataset (18 bp for E-box, 24 bp for D-box, and 23 bp for RRE) that contain ordinary consensus sequence at the center (CACGTG for E-box; ref. 19; TTATGTAA for D-box; ref. 22, [A/T]A[A/T]NT[A/G]GGTCA for RRE; ref. 24), then filtered out those that naturally exist in either the human or mouse genome. The sequence of the highest and lowest score was selected as the ''high-scoring'' and ''low-scoring'' sequences, respectively. All HMMER searches, except the directional search in the selection of E-box alignments, were performed bidirectionally. The higher score was adopted if match scores were obtained for both directions at the same position. The training data are available in Table S2. The HMMs are publicly available on the circadian section of the mammalian promoter/enhancer database: http://promoter.cdb.riken.jp/circadian.html.

Competitive DNA Binding Assays. *In vitro* transcription/translation of Flag-tagged mouse protein from pMU2-*Arntl*, pMU2-*Clock*, pMU2-*Bhlhb2*, pMU2-*Dbp*, pMU2-*Nfil3*, pMU2-*Nr1d1*, and pMU2-*Rora* were performed with TNT T7 Quick Coupled Transcription/Translation System (Promega) according to the manufacturer's specifications. *In vitro* transcribed/translated *Arntl* and *Clock* proteins were mixed in equal volume. The complementary oligonucleotides of three tandem repeats sequence of designed and control *cis*-acting elements, which were labeled with biotin on 5′-end or non-labeled (for competitor)

(Hokkaido System Science), were annealed to generate probes. Competitive DNA binding assays were performed with NoShift Transcription Factor Assay Kit (Novagen) according to the manufacturer's specifications with the following modifications. Ten pmol biotinylated annealed oligonucleotides were incubated with competitor oligonucleotides (final concentration were 0, 1, 3, 10, 30, and 100 pmol) and 5 μl of *in vitro* transcribed/translated reticulocyte lysates in the binding mixture. After the samples bound to streptavidin-coated microassay plate, the wells were washed, and Anti-Flag M2 Monoclonal Antibody-Peroxidase Conjugate (SIGMA) was applied into the each well. The wells were washed, and TMB substrate was added to each sample to develop a colorimetric signal, which was subsequently read on a spectrophotometer at 450 nm (Power Wave XS, BioTek). *Nr1d1* and *Rora* proteins were used with additional modifications. Binding reactions were performed with their own binding buffer (8 mM Tris-HCl, pH 7.5, 40 mM NaCl, 0.4 mM EDTA, 1.6 mM MgCl$_2$, 3.2% Glycerol, 0.4 mM DTT, 0.4 mg/ml BSA, and 0.5 μM poly dI;dC); 1 μM ZnSO$_4$ is further added into the binding buffer for *Rora* proteins and were incubated for 90 min at room temperature. And NoShift Wash Buffer and NoShift Antibody Dilution Buffer were diluted up to 0.5 × solution using water in dilution for a working solution.

ACKNOWLEDGMENTS. We thank Hajime Tei and Yoshiyuki Sakaki for *Arntl* and *Clock* expression vectors, Yutaka Suzuki, Sumio Sugano, and Seiichi Hashimoto for advice on the mammalian promoter/enhancer database, Takao Kondo for high-throughput monitoring systems, Hideki Ukai, Ryotaku Kito, and Kazuhiro Yagita for real-time circadian reporter assay, Rikuhiro G. Yamada, Tetsuya J. Kobayashi, and Tekeya Kasukawa for statistical analysis, and Michael Royle for critical reading. This work was supported by RIKEN Research Collaborations with Industry Program (H.R.U.), and in part, by RIKEN Center for Developmental Biology (CDB) internal grant (H.R.U.), New Energy and Industrial Technology Development Organization (NEDO) Scientific Project grant (H.R.U.) and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Genome Network Project to H.R.U.). J.B.H. is supported by grants from the National Institute of Mental Health and National Institute of Neurological Disorders and Stroke of the National Institutes of Health (P50MH074924–01, Joseph S. Takahashi, P. I, NIMH; 1R01NS054794–01A2, J.B.H, P. I, NINDS).

1. Lander ES, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
2. Waterston RH, *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.
3. Gibbs RA, *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521.
4. Ota T, *et al.* (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* 36:40–45.
5. Carninci P, *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* 309:1559–1563.
6. Gerhard DS, *et al.* (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res* 14:2121–2127.
7. Kimura K, *et al.* (2006) Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* 16:55–65.
8. Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K (2008) DBTSS: Database of transcription start sites, progress report 2008. *Nucleic Acids Res* 36:D97–D101.
9. Carninci P, *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38:626–635.
10. Matys V, *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34:D108–110.
11. Bryne JC, *et al.* (2008) JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update. *Nucleic Acids Res* 36:D102–D106.
12. Xie X, *et al.* (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* 434:338–345.
13. Schuldiner O, Shor S, Benvenisty N (2002) A computerized database-scan to identify c-MYC targets. *Gene* 292:91–99.
14. Menssen A, Hermeking H (2002) Characterization of the c-MYC-regulated transcriptome by SAGE: Identification and analysis of c-MYC target genes. *Proc Natl Acad Sci USA* 99:6274–6279.
15. Ueda HR, *et al.* (2002) A transcription factor response element for gene expression during circadian night. *Nature* 418:534–539.
16. Conkright MD, *et al.* (2003) Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness. *Mol Cell* 11:1101–1108.
17. Hoh J, *et al.* (2002) The p53MH algorithm and its application in detecting p53-responsive genes. *Proc Natl Acad Sci USA* 99:8467–8472.
18. Hallikas O, *et al.* (2006) Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* 124:47–59.
19. Hogenesch JB, Gu YZ, Jain S, Bradfield CA (1998) The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc Natl Acad Sci USA* 95:5474–5479.
20. Gekakis N, *et al.* (1998) Role of the CLOCK protein in the mammalian circadian mechanism. *Science* 280:1564–1569.
21. Ueda HR, *et al.* (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat Genet* 37:187–192.
22. Falvey E, Marcacci L, Schibler U (1996) DNA-binding specificity of PAR and C/EBP leucine zipper proteins: a single amino acid substitution in the C/EBP DNA-binding domain confers PAR-like specificity to C/EBP. *Biol Chem* 377:797–809.
23. Mitsui S, Yamaguchi S, Matsuo T, Ishida Y, Okamura H (2001) Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes Dev* 15:995–1006.
24. Harding HP, Lazar MA (1993) The orphan receptor Rev-ErbA alpha activates transcription via a novel response element. *Mol Cell Biol* 13:3113–3121.
25. Preitner N, *et al.* (2002) The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* 110:251–260.
26. Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. *Nature* 418:935–941.
27. Paquet ER, Rey G, Naef F (2008) Modeling an evolutionary conserved circadian cis-element. *PLoS Comput Biol* 4:e38.
28. Nakahata Y, *et al.* (2008) A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. *BMC Mol Biol* 9:1.
29. Sharov AA, Dudekula DB, Ko MS (2006) CisView: A browser and database of cis-regulatory modules predicted in the mouse genome. *DNA Res* 13:123–134.
30. Loots G, Ovcharenko I (2007) ECRbase: Database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics* 23:122–124.
31. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
32. Rozenberg JM, *et al.* (2008) All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC Genomics* 9:67.
33. Panda S, *et al.* (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109:307–320.
34. Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* 93:929–937.
35. Munoz E, Brewer M, Baler R (2002) Circadian Transcription. Thinking outside the E-Box. *J Biol Chem* 277:36009–36017.
36. Reinke H, *et al.* (2008) Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev* 22:331–345.
37. Matsumoto A, *et al.* (2007) A functional genomics strategy reveals clockwork orange as a transcriptional regulator in the Drosophila circadian clock. *Genes Dev* 21:1687–1700.
38. Kadener S, Stoleru D, McDonald M, Nawathean P, Rosbash M (2007) Clockwork Orange is a transcriptional repressor and a new Drosophila circadian pacemaker component. *Genes Dev* 21:1675–1686.
39. Lim C, *et al.* (2007) clockwork orange Encodes a Transcriptional Repressor Important for Circadian-Clock Amplitude in Drosophila. *Curr Biol* 17:1082–1089.
40. Sato TK, *et al.* (2006) Feedback repression is required for mammalian circadian clock function. *Nat Genet* 38:312–319.

SI Appendix

SI Materials and Methods

Construction of Promoter/Enhancer Database. Detailed database construction procedures are described here. We performed the following steps to construct the mammalian promoter/enhancer database. We mapped RefSeq (1) (15/11/2004 release 28,712 sequences for human and 8/11/2004 release 26,221 sequences for mouse), UniGene (2) (4,790,589 sequences of build #175 for human and 3,650,800 sequences of build #141 for mouse) and 5′-end sequences (3) (1,394,825 sequences) to human (NCBI Human genome build 35) or mouse (NCBI Mouse genome build 33) genomes respectively using the BLAT program (4). The results of mapping and information from Ensembl genes (19,580 known genes and 2,641 novel genes for human, and 20,718 known genes and 4,665 novel genes for mouse), Ensembl EST genes (27,678 genes for mouse) (Version 26) (5) and Vega genes (6,166 genes for human) (6) were used to identify gene structure and putative transcriptional start site (TSS). We identified 24,749 human genes with 50,373 TSSs and 26,047 mouse genes with 43,863 TSSs. Next, we determined 16,268 human-mouse orthologues (65.7% of human genes and 62.5% of mouse genes) by selecting the reciprocal best match of all genes using the BLAST program.

As a next step, we sought to define orthologous genes across species. Using positional information of adjacent orthologues, we were able to determine 434 human-mouse syntenic regions. These regions were then compared using the unit of the blocks performed by the LAGAN program (7). We determined a total of 750,043 human-mouse genome conserved regions (173 Mb and 5.6% coverage of human genome, and 172 Mb and 6.5% of mouse genome). Information regarding the coding regions was obtained from RefSeq, UniGene, Ensembl Genes and Vega Genes, allowing the non-coding genome conserved regions to be defined. Through this process, we were able to determine 893,798 human and 892,128 mouse human-mouse conserved non-coding regions (145 Mb and 4.7% coverage of human genome and 144 Mb and 5.5% of mouse genome). Next, we sought to define the transcription factor binding sites (TFBSs) by searching the known 862 consensus sequences obtained from TRANSFAC (8) at conserved regions of both genomes. In total, 7,804,559 human-mouse non-coding conserved TFBSs were predicted. Finally, data were integrated as a mammalian genome-wide promoter/enhancer database.

Determination of Genes and TSSs. Here we describe the detailed methods and conditions used in mapping RefSeq, UniGene and 5'-end sequences, determination of gene structures and TSSs, and determination of orthologues. All mRNA, EST and 5'-end sequences were mapped and selected by using blat, pslSort and pslReps programs of BLAT suit with the following parameter: "-q=rna -minIdentity=95" for blat and "-minCover=0.2 –minAli=0.98 –nearTop=0.002" for pslReps, this is determined based on the parameters used for constructing the UCSC genome browser database (9). We used repeat-masked genome sequences obtained from Ensembl (ftp://ftp.ensembl.org/) as the subject of mapping. Redundant mRNA sequences of UniGene, which contained part of the RefSeq sequences, were removed. If UniGene or the 5'-end sequence was mapped at a single continuous region, i.e. not spliced, it would not be used for further analysis in order to decrease genomic sequence contamination, transcripts of repetitive sequences, or pseudo genes, respectively. Determination of gene structures and TSSs was performed by the following five steps. First, we made transcriptional clusters by clustering overlapping chromosomal positions at the same strand of the results of mapping of RefSeq, UniGene and 5'-end sequences, and information of Ensembl genes, Ensembl EST genes and Vega genes. Second, we determined exons from the transcriptional cluster by selecting the regions with at least one or more sets of reliable information (RefSeq, Ensembl known gene or Vega gene). Third, we also determined exon-exon connections—i.e. the splicing junctions—by selecting the connections with at least one reliable set of information or two additional sets of information. Fourth, we determined the TSS or transcription termination site (TTS) containing exons by selecting exons with at least one or more sets of reliable information of 5'-end or 3'-end (5'-EST and mRNA sequence of UniGene and 5'-end sequence for TSS, and 3'-EST and mRNA sequence of UniGene for TTS). Lastly, we determined the position of TSS by selecting the most 5'-end position of reliable information or second most 5'-end position of other information. To determine human TSSs, we used 1,394,825 5'-end sequences from human full-length cDNA libraries generated by the oligo-cap method (3). In an attempt to determine mouse TSSs, we used UniGene sequences which contained 496,856 5'-end sequences from mouse full-length cDNA libraries that were generated by the FANTOM project using the cap-trapper method (10). To determine orthologues, we performed all genes to all genes BLAST homology search between the two species. Due to many of the genes having alternative splicing variants and alternative promoter variants, we used a merged gene sequence, which we were able to generate by joining determined exon sequences of the gene and using this for homology search. Following a nucleotide BLAST search with parameter "-E 1 –G 1", which reduced opening and extension of the alignment gap, and with a cutoff of E-value as 1.0e-4.0, we selected the reciprocal best match as orthologues.

Determination of Non-coding Genome Conserved Regions. Here we describe the detailed methods and conditions for determining genome conserved regions and coding regions. After the determination of syntenic regions by use of positional information of the adjacent orthologues, we used the following two steps to determine conserved regions in the genome was due to the size of the syntenic region being too long for comparing genome sequences between the two species. First, we divided the syntenic region into blocks by pairs of conserved exons between two species, which we determined by comparing the nucleotide sequence of exon regions of orthologous gene pairs using nucleotide BLAST. We then determined the conserved regions by comparing the unit of the blocks of synteny region of the two species. Genome comparisons were performed using the LAGAN program, and conserved regions were selected with 75% identity over 100 bp, as almost all of the evolutionally conserved elements previously determined (11) were detected using this cutoff value. These alignments were used for further analysis. Genome positions of the coding regions were determined by the following procedures. First, we compared the coding regions of RefSeq sequences with the mapped position of genome sequences. Second, we then compared the coding regions of mRNA and HTC sequences of UniGene with these mapped positions. Thirdly, we used the information from the coding regions of Ensembl genes and Vega genes. Finally, the genome position of all coding regions were merged and used for further analysis. In an attempt to determine the non-coding genome conserved regions, overlapping regions were removed.

Prediction of Evolutionally Conserved putative TFBSs. In this section we describe the prediction of evolutionally conserved putative TFBSs. We used 1,002 known consensus sequences of TFBS that were obtained from TRANSFAC. We searched the consensus sequences at non-coding genome conserved regions of the two species, then selected evolutionally conserved putative TFBSs that exist in the corresponding positions of the two species using alignments of sequence of genome conserved regions determined above. The search results of 140 consensus sequences of human-mouse conserved putative were not used as no conserved putative TFBS was detected at non-coding genome conserved regions or excess number (> 250,000) of conserved TFBSs were detected.

Annotation and Integration. Here we describe the annotation, integration, and interface to the database. All determined data, including genes, TSSs, non-coding genome conserved regions, and putative TFBSs, were related to gene annotations obtained from the Entrez Gene (2). Additionally, we performed a rpsblast search of the conserved domain database (CDD) (12) to estimate the molecular function of these genes. All data were integrated in a MySQL database. Finally, we constructed a web user interface for use with the mammalian promoter/enhancer database. This meant that a large amount of data, including the alignment of genome conserved regions, sequence and conservation of putative TFBSs, was made viewable by using the interface. Visualized genome conservation of genome conserved regions, which were outputted from the VISTA program (13), were also viewable. Additionally, to browse genome information annotated with determined genes, TSSs, non-coding genome conserved regions and putative TFBSs, we introduced the Ensembl Genome Browser (5) and the distributed annotation system (DAS) (14).

Determination of the distributions of distance from TSSs for natural and randomly positioned elements. To determine the distributions of distance from TSSs for clock-controlled elements, we used predicted 1,108 E-boxes, 2,314 D-boxes and 3,288 RREs (available on the circadian section of the mammalian promoter/enhancer database: http://promoter.cdb.riken.jp/circadian.html) and calculate the distance from nearest TSS. The distributions of distance from TSSs for random positioned clock-control elements were determined by randomly distribute the same number elements (1,108 for E-box, 2,314 for D-box and 3,288 for RRE) within conserved non-coding regions. The distributions of randomly positioned elements were generated 100 times and the averaged distributions were used.

Determination of FDR of putative elements. The false discovery rate (FDR) was determined for predicted elements. We used the result of predictions of a randomized genome as a background distribution of false positives. Profile HMM searches were performed in three conditions, (I) searching for conserved elements within conserved non-coding regions ("Conserved element"), (II) searching for mouse elements within the conserved non-coding regions by relaxing the requirement of element conservation ("Non-coding region") and (III) searching for mouse elements within the entire mouse genome by relaxing both element conservation and search space ("Whole genome"). The mouse genome sequence was randomized for these three conditions preserving the frequency of sequential nucleotide pair ("dinucleotide"), because a relatively small number of "CG" dinucleotides are existed in mammalian genomes. Otherwise (i.e. without preservation of dinucleotide), it will affect on the FDR for E-box, which contains "CG" dinucleotide. For the "Whole genome", a random mouse genome sequence was generated to keep the frequency of dinucleotide of the entire original mouse genome. For the "Non-coding region", a random genome sequence was generated to keep the frequency of each dinucleotide of the non-coding genome conserved regions of the original mouse genome. After randomization, genome sequence was masked except in the corresponding regions of the conserved non-coding genome regions of the original mouse genome. To reflect conservation between mouse and human among the "Conserved element", non-coding genome conserved regions of mouse and human genomes were generated to maintain the dinucleotide frequency at the parallel positions of the conserved genome region between mouse and human, including the case of a mismatch and or gap. For each condition, 200 randomized genome sequences were generated, and then HMM searches were performed. The averaged hit counts of 200 searches of these models on randomized genome sequence were used to obtain a background distribution and estimate the false discovery rate for each model taking into consideration the total number of base pairs searched. If the simulated value of FDR accidentally exceeds 1.0, the value is set to 1.0.

Animals. Male Balb/c mice (JAPS, Osaka, Japan) were purchased 5 weeks after birth. Mice were on a 12 hr light (400 lux): 12h dark (LD12:12) cycle for at least 2 weeks and were given food and water *ad lib*. Then animals were transferred to constant darkness conditions (DD) and, during the second DD cycle starting at CT0, animals were sacrificed every four hours under deep anesthesia and tissue samples were removed and frozen in liquid nitrogen for RNA extraction. This study was performed in compliance with the Rules and Regulations of the Animal Care and Use Committee, Kinki University School of Medicine, and followed the Guide for the Care and Use of Laboratory Animals, Kinki University School of Medicine.

Genome sequences. Genome sequences of NCBI Human genome build 35 and NCBI Mouse genome build 33 were downloaded from FTP site of Ensembl Project (5) (ftp://ftp.ensembl.org/) and used in this study.

Oligonucleotide sequences.

Primer sequence for quantitative PCR of mRNA.

Tbp-forward: 5'-GTTGTGCAGAAGTTGGGCTTC-3'

Tbp-reverse: 5'-TCACAGCTCCCCACCATGTT-3'

Per2-forward: 5'-TGTGCGATGATGATTCGTGA-3'

Per2-reverse: 5'-GGTGAAGGTACGTTTGGTTTGC-3'

Cry1-forward: 5'-TGAGGCAAGCAGACTGAATATTG-3'

Cry1-reverse: 5'-CCTCTGTACCGGGAAAGCTG-3'

Arntl-forward: 5'-CCACCTCAGAGCCATTGATACA-3'

Arntl-reverse: 5'-GAGCAGGTTTAGTTCCACTTTGTCT-3'

MGI:1926224-forward: 5'-GACCTGGCGGTGGATGG-3'

MGI:1926224-reverse: 5'-AACACATTTGCGTCCTGCC-3'

Lrrc35-forward: 5'-TGCTGCAGGCCTAATCCTTT-3'

Lrrc35-reverse: 5'-CGGTTGGGTTGGATGAGACT-3'

1300001I01Rik-forward: 5'-AAGATGGTTTTGCACTGGTTCA-3'

1300001I01Rik-reverse: 5'-TTTCGTGTCTTCAATTAGGCCTC-3'

Cpne7-forward: 5'-ATGGAAAGGGTGGTGAAGGG-3'

Cpne7-reverse: 5'-TCTCCACACGATCAAATGGC-3'

Gria2-forward: 5'-CTGAGTGCCTTACACAATGGTTTC-3'

Gria2-reverse: 5'-CGGATGCCTCTCACCACTTT-3'

Etv4-forward: 5'-GCCTCTGCCTAGGTCTTGCTC-3'

Etv4-reverse: 5'-ACACTGGATCTCTGTGGTGGG-3'

Pogz-forward: 5'-TTTATGCCACCACTCCCAGC-3'

Pogz-reverse: 5'-CGGCGTTCCTAATAACCCAC-3'

Plcb1-forward: 5'-AGTGCACGCCTTGCAACTC-3'

Plcb1-reverse: 5'-CTTCTTGAGGCTGTCGGACAC-3'

Irf2bp1-forward: 5'-TCGTGGCTTGCCTTTTCC-3'

Irf2bp1-reverse: 5'-CTTCCCCGCCCCCTG-3'

Trim8-forward: 5'-ACCCTCTTTCTAGCCGGAAGTT-3'

Trim8-reverse: 5'-GGTTTGAAGATGCCAAAGGC-3'

Tspan7-forward: 5'-GTATGGCATCGAGGAGAATGG-3'

Tspan7-reverse: 5'-ATGAGGAGGGTTTTGAGACAGG-3'

Atp1b2-forward: 5'-CTCGAATTTTGGAGCCGTCT-3'

Atp1b2-reverse: 5'-CACACACCGCCTAGAAGCAA-3'

Spsb3-forward: 5'-CAGGGACATCTCTGGTTCATTCA-3'

Spsb3-reverse: 5'-GGCTGAGCGCCGTATAAGAA-3'

Mgea5-forward: 5'-CCTTAATAGCAGATCCGCATGTG-3'

Mgea5-reverse: 5'-CAGTCCCCTTACCCTTACTTAACAAT-3'

Adam12-forward: 5'-CACTGTCCAGCCAATGTGTACC-3'

Adam12-reverse: 5'-AGTAACCATCCACGCCCTGA-3'

Myo1b-forward: 5'-ACGAGTGTTTGTCTCTCTCTCCCT-3'

Myo1b-reverse: 5'-CAGACTTCAGCAGCCCTTTAGC-3'

D13Ertd150e-forward: 5'-CGCTTTTGCAACCAGGTGTT-3'

D13Ertd150e-reverse: 5'-GGTGGGAGCGAACGTGG-3'

The oligonucleotide sequence for competitive binding assay.

Canonical consensus sequences for E-box (CACGTG), D-box (TTATGTAA) and RRE ([A/T]A[A/T]NT[A/G]GGTCA) are indicated in bold while mutated core sequences for each element are indicated in bold and italics.

Per1 E-box-forward: 5'-CAAGTC**CACGTG**CAGGGACAAGTC**CACGTG**CAGGGACAAGTC**CACGTG**CAGGGA-3'

Per1 E-box-reverse: 5'-TCCCTG**CACGTG**GACTTGTCCCTG**CACGTG**GACTTGTCCCTG**CACGTG**GACTTG-3'

Mutated *Per1* E-box-forward: 5'-CAAGTC***ACCGGT***CAGGGACAAGTC***ACCGGT***CAGGGACAAGTC***ACCGGT***CAGGGA-3'

Mutated *Per1* E-box-reverse: 5'-TCCCTG***ACCGGT***GACTTGTCCCTG***ACCGGT***GACTTGTCCCTG***ACCGGT***GACTTG-3'

High-scoring E-box-forward: 5'-CGGGGC**CACGTG**CAGGCGCGGGGC**CACGTG**CAGGCGCGGGGC**CACGTG**CAGGCG-3'

High-scoring E-box-reverse: 5'-CGCCTG**CACGTG**GCCCCGCGCCTG**CACGTG**GCCCCGCGCCTG**CACGTG**GCCCCG-3'

Low-scoring E-box-forward: 5'-GTTAAA**CACGTG**TTTTACGTTAAA**CACGTG**TTTTACGTTAAA**CACGTG**TTTTAC-3'

Low-scoring E-box-reverse: 5'-GTAAAA**CACGTG**TTTAACGTAAAA**CACGTG**TTTAACGTAAAA**CACGTG**TTTAAC-3'

Per3 D-box-forward: 5'-

CCCGCGCG**TTATGTAA**GGTACTCGCCCGCGCG**TTATGTAA**GGTACTCGCCCGCGCG**TTATGTAA**GGTACTCG-3'

Per3 D-box-reverse: 5'-

CGAGTACC**TTACATAA**CGCGCGGGCGAGTACC**TTACATAA**CGCGCGGGCGAGTACC**TTACATAA**CGCGCGGG-3'

Mutated *Per3* D-box-forward: 5'-

CCCGCGCG***CACCCGGC***GGTACTCGCCCGCGCG***CACCCGGC***GGTACTCGCCCGCGCG***CACCCGGC***GGTACTCG-3'

Mutated *Per3* D-box-reverse: 5'-

CGAGTACC***GCCGGGTG***CGCGCGGGCGAGTACC***GCCGGGTG***CGCGCGGGCGAGTACC***GCCGGGTG***CGCGCGGG-3'

High-scoring D-box-forward: 5'-

CCCGCGCG**TTATGTAA**CGAGCCCGCCCGCGCG**TTATGTAA**CGAGCCCGCCCGCGCG**TTATGTAA**CGAGCCCG-3'

High-scoring D-box-reverse: 5'-

CGGGCTCG**TTACATAA**CGCGCGGGCGGGCTCG**TTACATAA**CGCGCGGGCGGGCTCG**TTACATAA**CGCGCGGG-3'

Low-scoring D-box-forward: 5'-

ATGAAAAT**TTATGTAA**GTTTAAACATGAAAAT**TTATGTAA**GTTTAAACATGAAAAT**TTATGTAA**GTTTAAAC-3'

Low-scoring D-box-reverse: 5'-

GTTTAAAC**TTACATAA**ATTTTCATGTTTAAAC**TTACATAA**ATTTTCATGTTTAAAC**TTACATAA**ATTTTCAT-3'

Arntl RRE-forward: 5'-

AGGCAG**AAAGTAGGTCA**GGGACGAGGCAG**AAAGTAGGTCA**GGGACGAGGCAG**AAAGTAGGTCA**GGGACG-3'

Arntl RRE-reverse: 5'-

CGTCCC**TGACCTACTTT**CTGCCTCGTCCC**TGACCTACTTT**CTGCCTCGTCCC**TGACCTACTTT**CTGCCT-3'

Mutated *Arntl* RRE-forward: 5'-

AGGCAG***AAAGTCCTAGC***GGGACGAGGCAG***AAAGTCCTAGC***GGGACGAGGCAG***AAAGTCCTAGC***GGGACG-3'

Mutated *Arntl* RRE-reverse: 5'-

CGTCCC***GCTAGGACTTT***CTGCCTCGTCCC***GCTAGGACTTT***CTGCCTCGTCCC***GCTAGGACTTT***CTGCCT-3'

High-scoring RRE-forward: 5'-

AGAAAG**AAAGTAGGTCA**GTGCGGAGAAAG**AAAGTAGGTCA**GTGCGGAGAAAG**AAAGTAGGTCA**GTGCGG-3'

High-scoring RRE-reverse: 5'-

CCGCAC**TGACCTACTTT**CTTTCTCCGCAC**TGACCTACTTT**CTTTCTCCGCAC**TGACCTACTTT**CTTTCT-3'

Low-scoring RRE-forward: 5'-

CTCCCC**TATTTGGGTCA**ACCGACCTCCCC**TATTTGGGTCA**ACCGACCTCCCC**TATTTGGGTCA**ACCGAC-3'

Low-scoring RRE-reverse: 5'-

GTCGGT**TGACCCAAATA**GGGGAGGTCGGT**TGACCCAAATA**GGGGAGGTCGGT**TGACCCAAATA**GGGGAG-3'

The oligonucleotide sequence for construction of expression plasmids.

Underlines indicate linker sequence, which incorporated the recognition sequence for a restriction enzyme.

Arntl-forward: 5'-ATTACCCTGTTATCCCTAATGCGGACCAGAGAATGGAC -3'

Arntl-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCACTACAGCGGCCATGGCAAGTC-3'

Clock-forward: 5'-ATTACCCTGTTATCCCTAATGTGTTTACCGTAAGCTGTAG -3'

Clock-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCACTGTGGCTGGACCTTGG -3'

Bhlhb2-forward: 5'-ATTACCCTGTTATCCCTAATGAACGGATCCCCAGCGC -3'

Bhlhb2-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCAGTCTTTGGTTTCTAAG -3'

Dbp-forward: 5'-ATTACCCTGTTATCCCTAATGCGCGGCCTCTGAGCGAC -3'

Dbp-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCACAGTGTCCCATGCTGG -3'

Nfil3-forward: 5'-ATTACCCTGTTATCCCTAATCAGCTGAGAAAAATGCAG -3'

Nfil3-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCACCTGGAGTCCGAAGCCG -3'

Nr1d1-forward: 5'-ATTACCCTGTTATCCCTAATACGACCCTGGACTCCAATAAC -3'

Nr1d1-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCACTGGGCGTCCACCCGG -3'

Rora-forward: 5'-ATTACCCTGTTATCCCTAATTATTTTGTGATCGCAGCG -3'

Rora-reverse: 5'-ACCCATAATACCCATAATAGCTGTTTGCCACCCATCGATTTGCATGGCTG-3'

Plasmid constructions. The SV40-dLuc (15) and *Per2*-dLuc (16) reporters are described elsewhere. The sequences containing three tandem repeats of putative *cis*-acting elements were inserted into *Mlu*I/*Bgl*II site of the SV40-dLuc vector. Detailed information on putative *cis*-acting elements is available in **Tables S1 and S3**. The DNA sequences of all constructs generated in this study were verified by standard methods. For construction of expression plasmids, we amplified the full length coding sequence of mouse *Arntl*, *Clock*, *Bhlhb2*, *Dbp*, *Nfil3*, *Nr1d1* and *Rora* from pCI-*Arntl* (17), pCI-*Clock* (17), pSPORT6-*Bhlhb2* (MGC clone # 3707474, Invitrogen, Carlsbad, CA), pSPORT6-*Rora* (MGC clone # 3592667, Invitrogen) or NIH3T3 cDNA library by PCR with forward primers containing *I-Sce*I recognition sequence and reverse primers containing *PI-Psp*I recognition sequence (Hokkaido System Science, Hokkaido, Japan). PCR product was digested with *I-Sce*I (NEW ENGLAND BioLabs, Ipswich, MA) and *PI-Psp*I (NEW ENGLAND BioLabs), and cloned into pMU2 vector (18) and termed as pMU2-*Arntl*, pMU2-*Clock*, pMU2-*Bhlhb2*, pMU2-*Dbp*, pMU2-*Nfil3*, pMU2-*Nr1d1* and pMU2-*Rora*. Those genes were fused in-frame with 1 × Flag Tag at N-terminal by *I-Sce*I recognition site, and regulated by T7 promoter in pMU2.

Quantitative PCR. Quantitative PCR was performed with the ABI PRISM 7900HT and SYBR Green Reagents (Applied Biosystems, Foster City, CA). cDNAs were synthesized from 0.25 μg of total RNA using SuperScript II Reverse Transcriptase (Invitrogen) and Random Primers (Promega). Samples contained 1 × SYBR Green PCR Master mix, 0.8 μM primers and 1/50 synthesized cDNA in a 10 μl volume. The PCR conditions were as follows: 10min at 95°C, then 45 cycles of 15 sec at 95°C and 1min at 59°C. Absolute cDNA abundance was calculated using the standard curve obtained from mouse genomic DNAs. *Tbp* expression levels were quantified and used as the internal control. Detailed data of quantitative PCR is available at the circadian section of the mammalian promoter/enhancer database.

Rhythmicity analysis of real-time bioluminescence data. Bioluminescence time-series data beginning 21 h after forskolin stimulation were used for analysis to distinguish endogenous circadian oscillations from acute effects of stimulation. Bioluminescence data were detrended by using the trend curve calculated by the smoothing spline method, and statistical significance and the period of circadian oscillation in the detrended data was evaluated as previously described ($p < 0.01$) (16). To visualize the normalized bioluminescence data (i.e. the oscillatory component of the bioluminescence data) shown in **Fig. 2*A***, the moving average of the absolute value of the detrended bioluminescence data was calculated first. The window size of the moving average was set to half of the period calculated above. Then, the oscillatory component of the detrended data was calculated by dividing the data by the moving average of the data at each time point.

Amplitude analysis of real-time bioluminescence data. Bioluminescence time-series data of 21–96 h after forskolin stimulation were normalized so that the average bioluminescence is 1.0, and then detrended as described above. Detrended and normalized time-series data were used in further analysis. To determine the amplitude of each sample, 151 "reference" time-series data were generated for E-box, D-box and RRE, respectively, by multiplying time-series data of positive controls (*Per1* E-box, *Per3* D-box and *Arntl* RRE) (11) with the value of "relative amplitude" from 0 to 1.5 with 0.01 step. Time-series data for each sample were compared with reference data using the least-squares method to determine the best fit for reference data, the "relative amplitude" of which was used as the control amplitude and used for comparisons.

Rhythmicity analysis of quantitative PCR data. Two statistical tests, cosine fitting and analysis of variance (ANOVA), were combined to identify circadian expression profiles with high-amplitude. To evaluate the wave form of expression profiles, statistical analysis (cosine fitting test) was performed in parallel on two independent expression profiles (two days each). 2-cycle cosine waves of 24-h period with a different phase were generated by shifting phase with 0.4-h interval (total 60 waves). A two day expression profile was compared with 60 cosine test waves by calculating the correlation coefficient to determine the most correlated cosine wave, the correlation coefficient of which was used as "maximum correlation coefficient" for the expression profile. Statistical significance of the maximum correlation coefficient for the expression profile was evaluated by calculating the maximum correlation coefficient for 100,000 random expression profiles. Two *P*-values calculated on two independent expression profiles (two days each) were combined by Fisher's probability combination method (19) to calculate the synthetic *P*-value for cosine fitting test. To evaluate the amplitude of expression profile, statistical analysis (one-way ANOVA test) was also performed on experimental duplicates. *P*-values for cosine fitting and ANOVA tests were combined by Fisher's probability combination method to calculate the synthetic *P*-value, which was used to identify circadian expression profiles with high-amplitude ($p < 0.03$). The peak time of an expression profile was estimated by the peak time of the best correlated cosine wave.

Over representation analysis of clock-controlled genes. To determine the significance of the enrichment of clock-controlled genes within the genes having predicted clock-controlled elements, we first selected the 100 most significant predicted sequences for each clock-controlled elements (E-box, D-box and RRE) that were mapped to 98 genes on U74 microarray after removing the 21 clock-controlled genes used for HMM generation and training. As the dataset of clock-controlled genes, we used previously identified clock-controlled genes in the SCN and liver by using mouse U74 microarray (15, 20). After removing the 21 clock-controlled genes in the above, we found additional 19 putative clock-controlled out of the 6,195 genes common in our mammalian promoter/enhancer database and U74 mouse microarray, which is significantly higher than the expected 10.67 genes that would have arisen from chance ($p = 0.009$).

Estimation of the number of high-amplitude E-boxes. To estimate the number of clock-controlled conserved E-boxes that likely confer circadian rhythmicity, we used 1,108 E-boxes (minimum HMM score = 11.56; available on http://promoter.cdb.riken.jp/circadian.html) except E-boxes used for HMM generation and training. Linear interpolation from the true positive rate (40%) of high-scoring conserved E-boxes (mean HMM score = 16.15) and the false negative rate (7.1%) of low- scoring conserved E-boxes (mean HMM score = 2.5143) was used to estimate the probability of rhythmicity for each E-box. The sum of the probability of 1,108 E-boxes was calculated to estimate the number of high-amplitude E-boxes.

Microarray expression data analysis of genes with predicted clock-controlled elements. To investigate the averaged expression of genes with predicted clock-controlled elements, we used time-series microarray expression data of mouse liver in our previous study (15). Averaged expression value under LD and DD conditions were used for this analysis. We selected 100 most significant sequences for each clock-controlled elements (E-box, D-box and RRE). 21 genes, which were used for training data set of HMMs, were excluded in the selection. We then related them to microarray probe set data (36 genes for E-box, 29 genes for D-box and 34 genes for RRE respectively). In case that more than one probe sets exists for a single gene, averaged expression data was used. Then, averaged expression value for each time points were determined for each clock-controlled element. Data were then normalized so that the average expression value over 2-day 12-point time courses is 1.0. To statistical tests, cosine fitting was performed. 2-cycle cosine waves of 24-h period with a different phase were generated by shifting phase with 0.4-h interval (total 60 waves). A 2-day expression profile was compared with 60 cosine test waves by calculating the correlation coefficient to determine the most correlated cosine wave, the correlation coefficient of which was used as "maximum correlation coefficient" for the expression profile. Statistical significance of the maximum correlation coefficient for the expression profile was evaluated by calculating the maximum correlation coefficient for 100,000 random expression profiles. *P*-values for cosine fitting were used to evaluate circadian oscillation of averaged expression profiles. The peak times of the averaged expression profile were estimated by the peak time of the best correlated cosine wave.

Sequence logos. Sequence logos shown in **Fig. S4** and **Table S2** were created by WebLogo (http://weblogo.berkeley.edu/) (21).

Calculation of relative affinity from competitive DNA binding assay data. A relative binding affinity between a regulator (R) and an unlabeled competitive DNA element (D_u) in comparison with that to the labeled control DNA element (D_l) was determined by competitive DNA binding assays. In each competitive DNA binding assay, the concentration of bound regulator to the labeled element ($[RD_l]$) can be described as follows;

$$[RD_l] = ([R]_{all} - [R])\frac{[D_l]}{[D_l] + \frac{A_u}{A_l}[D_u]} \qquad \text{(Eq. 1)}$$

, where $[R]_{all} \equiv [RD_l] + [RD_u] + [R]$ is a total concentration of a regulator,

$[RD_l]$ is a concentration of the regulator-labeled element complex,

$[RD_u]$ is a concentration of the regulator-unlabeled element complex,

$[R]$ is a concentration of a free regulator,

$[D_l]$ is a concentration of a free labeled element,

$[D_u]$ is a concentration of a free unlabeled element,

$A_l \equiv \frac{[RD_l]}{[R][D_l]}$ is an affinity constant between a regulator and a labeled element, and

$A_u \equiv \frac{[RD_u]}{[R][D_u]}$ is an affinity constant between a regulator and an unlabeled element.

Since an excessive amount of a labeled or unlabeled element is usually applied in the competitive DNA binding assay, the concentration of a labeled ($[D_l]$) or unlabeled ($[D_u]$) element is much greater than the concentration of bound regulator to the labeled ($[RD_l]$) or unlabeled ($[RD_u]$) element. Thus, the total concentration of the labeled ($[D_l]_{all} \equiv [D_l] + [RD_l]$) or unlabeled ($[D_u]_{all} \equiv [D_u] + [RD_u]$) element can be approximated as follows;

$$[D_l]_{all} \cong [D_l]$$

$$[D_u]_{all} \cong [D_u]$$

Since an amount of a labeled or unlabeled element vastly exceeds the amount of regulator, the amount of a free regulator ($[R]$) is much less than the total amount of a regulator ($[R]_{all}$) in the competitive DNA binding assay. Thus, $[R]_{all} - [R]$ can be also approximated as follows;

$$[R]_{all} - [R] \cong [R]_{all}$$

In addition, since $[RD_l]$ is proportional to the measured value of competitive DNA binding assay (M_{450}), Equation 1 can be rewritten as follows;

$$M_{450} \propto [RD_l] \cong [R]_{all} \frac{[D_l]_{all}}{[D_l]_{all} + \frac{A_u}{A_l}[D_u]_{all}}$$

Furthermore, as $[R]_{all}$ is constant in the assay, we can derive the following equation;

$$M_{450} = C\frac{[D_l]_{all}}{[D_l]_{all} + A*[D_u]_{all}} \qquad \text{(Eq. 2)}$$

, where C is a proportional constant, and $A \equiv \frac{A_u}{A_l}$ is an relative affinity constant of the unlabeled element in comparison with that of the labeled element.

To determine the relative affinity constant from the measured value of competitive DNA binding assay, we first determined the value of $[D_l]_{all}$ and C from series of data for unlabeled known clock-controlled element (positive control) where A was defined as 1.0. In details, by changing the value of $[D_l]_{all}$ and C, the most fitting $[D_l]_{all}$ and C values were determined using the least-square method applied to series of measured values and model data calculated from Equation 2. We then determined the values of A for unlabeled elements including the "high-scoring", "low-scoring" and "negative control" by changing the value of A.

***In silico* analysis of affinity to amplitude mechanism.** By modifying the previous described formula (11), we formulated transcriptional activity $T(t)$ at time t regulated by competition between a clock-controlled activator and a repressor as follows:

$$T(t) \equiv \frac{(\frac{A(t)}{K_a})^n}{1 + (\frac{A(t)}{K_a})^n + (\frac{R(t)}{K_b})^n} + \alpha$$

where $1/K_a$ and $1/K_b$ represent the affinity of an activator and a repressor, respectively. α represents transcriptional activity that does not depend on the circadian clock. n represents the Hill coefficient at competitive regulation. $A(t)$ and $R(t)$ represent expression of a clock-controlled activator and repressor, which are defined as follows:

$$A(t) \equiv \beta_a (1 + Cos(2\pi \frac{t-a}{24})) + \gamma_a$$

and

$$R(t) \equiv \beta_b (1 + Cos(2\pi \frac{t-b}{24})) + \gamma_b$$

where β_a and β_b represent half amplitude of expression of an activator and a repressor. γ_a and γ_b represent expression of an activator and a repressor that does not depend on the circadian clock. $a(-12 \le a \le 12)$ and $b(-12 \le b \le 12)$ represent phases of expression of an activator and a repressor. Then, we formulated output $P(t)$ at time t which depended on transcriptional activity $T(t)$ as follows:

$$\frac{d}{dt} P(t) = T(t) - \lambda P(t)$$

where λ represent decay constant and defined as follows:

$\lambda \equiv \log 2 / T_{1/2}$

where $T_{1/2}$ represent half life. For simplicity, we used $a = 0$, $b = 12$, $\alpha = 0.2$, $\beta_a = \beta_b = 1$, $\gamma_a = \gamma_b = 0.2$, $T_{1/2} = 3$, $P(0) = 0$ and $48 \le t \le 144$ in the analysis. We used $n = 1$ and $K_b = 1$ for the analysis of activator affinity to amplitude mechanism (**Fig. 4 *B* and *C***). Output time-series data were normalized so that the center value of maximum and minimum value of time-series data is 100%. The difference between maximum and minimum values of normalized output time-series data was used to define amplitude.

SI Discussion

Utility of the Promoter/Enhancer Database. As a tool for understanding systems-level transcriptional regulation in mammals, we constructed the mammalian promoter/enhancer database (http://promoter.cdb.riken.jp) by integrating information of conserved non-coding regions, TSSs, and TFBSs. Users can input a gene name or symbol, or UniGene or RefSeq identifiers, and get back a page that summarizes promoter information with additional links to outside databases such as SymAtlas and NCBI. The promoter sequences are available in a default view 1000 bp 5' of the TSS; this default view can be changed arbitrarily by the user 5' or 3' of the TSS, and the DNA sequence information can be downloaded in FASTA format. Users can also highlight conserved sequence regions between humans, mice, and rats, as well as TSSs, exons and TFBSs. Where known, links to alternative promoters are available. Although we applied this database to the understanding of circadian transcriptional regulation, it is generically useful and can be applied to any aspect of mammalian transcriptional regulation.

In silico* modeling of affinity amplitude mechanism.** From competitive DNA binding assays, the "high-scoring" D-box and RRE show approximately the same affinity as positive control whereas the "low-scoring" D-box and RRE show relatively weak affinity (**Fig. 4*A and **Fig. S5*A***). These results can be reasonably explained by using *in silico* model that has been extended from our previous model for gene expression of transcriptional activators and repressors to generate high-amplitude circadian output (11) by introducing expression dynamics of an activator and a repressor, affinity of a DNA element to an activator and a repressor, and half life of output (see also ***SI Materials and Methods***). This *in silico* analysis shows that, if a DNA element has a weak affinity to both an activator and a repressor, then the transcription system exhibits low-amplitude of output oscillations, which can be intuitively elucidated (**Fig. S5*B***). In this *in silico* analysis also shows that the amplitude of output oscillations depend not only on the strength of affinity but also on a hill coefficient (i.e. a parameter for nonlinearity in transcriptional response) (**Fig. S5*B***).

On the other hand, the "high-scoring" E-box shows approximately the same affinity as positive control, whereas the "low-scoring" E-box shows, surprisingly, 4.8-times higher affinity only to *Arntl/Clock* activator (**Fig. 4*A* and Fig. S5*A***). Since "low-scoring" E-box shows approximately the same affinity to *Bhlhb2* repressor as positive control, this result suggests that the "low-scoring" E-box has an unbalanced affinity stronger for *Arntl/Clock* activator than *Bhlhb2* repressor. In order to interpret this seemingly complicated result, we also performed *in silico* analysis of affinity-amplitude relationship especially in the case that a DNA element has an unbalanced affinity between an activator and a repressor. This *in silico* analysis shows that, if a DNA element has a 5-times higher affinity to an activator than a repressor, then the transcription system exhibits less than half amplitude of output oscillations (**Fig. 4*B***). Further *in silico* analysis also shows that, if a DNA element has a lower affinity to an activator than a repressor, then the transcription system exhibits the reduced amplitude of output oscillations (**Fig. 4*C***). Collectively, these results suggest that not only the strength of affinity to regulators but also the balance of affinity between an activator and a repressor are important in generating high-amplitude outputs. In this study, we supposed the same level of expression for a clock-controlled activator and repressor for simplicity in the modeling, and drew the conclusion that appropriate affinity balance between activators and repressors is important. We can easily generalize it into different levels of expression between activator and repressor. In such a case, the conclusion also holds for the product of affinity and concentration instead (i.e. appropriate balance in the product of concentration and affinity is important).

SI References

1. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:D61-65.

2. Wheeler DL, *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36:D13-D21.

3. Kimura K, *et al.* (2006) Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* 16:55-65.

4. Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12:656-664.

5. Flicek P, *et al.* (2008) Ensembl 2008. *Nucleic Acids Res* 36:D707-D714.

6. Wilming LG, *et al.* (2008) The vertebrate genome annotation (Vega) database. *Nucleic Acids Res* 36:D753-D760.

7. Brudno M, *et al.* (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13:721-731.

8. Matys V, *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 34:D108-110.

9. Karolchik D, *et al.* (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* 36.

10. Okazaki Y, *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563-573.

11. Ueda HR, *et al.* (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat Genet* 37:187-192.

12. Marchler-Bauer A, *et al.* (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 35:D237-240.

13. Mayor C, *et al.* (2000) VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16:1046-1047.

14. Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L (2001) The distributed annotation system. *BMC Bioinformatics* 2:7.

15. Ueda HR, *et al.* (2002) A transcription factor response element for gene expression during circadian night. *Nature* 418:534-539.

16. Sato TK, *et al.* (2006) Feedback repression is required for mammalian circadian clock function. *Nat Genet* 38:312-319.

17. Hida A, *et al.* (2000) The human and mouse Period1 genes: five well-conserved E-boxes additively contribute to the enhancement of mPer1 transcription. *Genomics* 65:224-233.

18. Ukai H, *et al.* (2007) Melanopsin-dependent photo-perturbation reveals desynchronization underlying the singularity of mammalian circadian clocks. *Nat Cell Biol* 9:1327-1334.

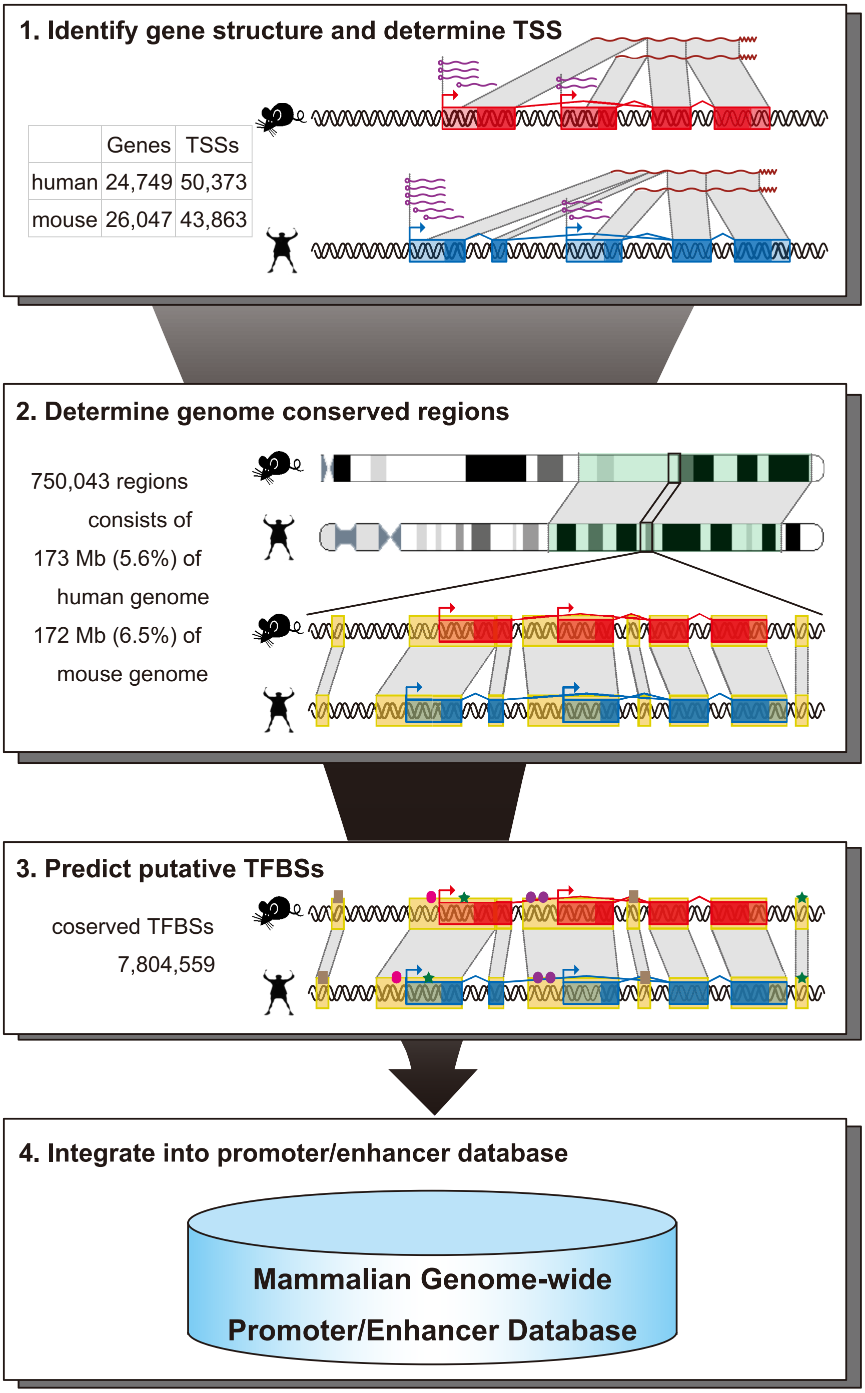
19. Fisher RA (1970) Statistical methods for research workers.

20. Panda S, *et al.* (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109:307-320.

21. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188-1190.

Supporting Information

Kumaki *et al.* 10.1073/pnas.0802636105



1. Identify gene structure and determine TSS

	Genes	TSSs
human	24,749	50,373
mouse	26,047	43,863

2. Determine genome conserved regions

750,043 regions consists of 173 Mb (5.6%) of human genome 172 Mb (6.5%) of mouse genome

3. Predict putative TFBSs

coserved TFBSs 7,804,559

4. Integrate into promoter/enhancer database
Mammalian Genome-wide
Promoter/Enhancer Database

Fig. S1. Construction of the mammalian promoter/enhancer database. Mammalian full-length cDNA and EST sequences were mapped onto mammalian genome sequences to identify 24,749 human genes with 50,373 TSSs and 26,047 mouse genes with 43,863 TSSs. These mammalian genes were then compared to identify 16,268 human-mouse orthologs (65.7% of human genes and 62.5% of mouse genes). The positional information of adjacent orthologs was used to determine 434 human-mouse syntenic regions, which contained 750,043 human-mouse conserved genomic regions. The 862 consensus TFBSs from TRANSFAC were then mapped on these conserved genomic regions to identify the 7,804,559 sites conserved between human and mouse in non-coding regions. Human-rat comparisons were performed using the same procedure. Finally, visualization of putative promoter/enhancer and TFBSs data, and curation of current genes were integrated into a free and publicly accessible website (Mammalian Promoter/Enhancer Database; http://promoter.cdb.riken.jp/).

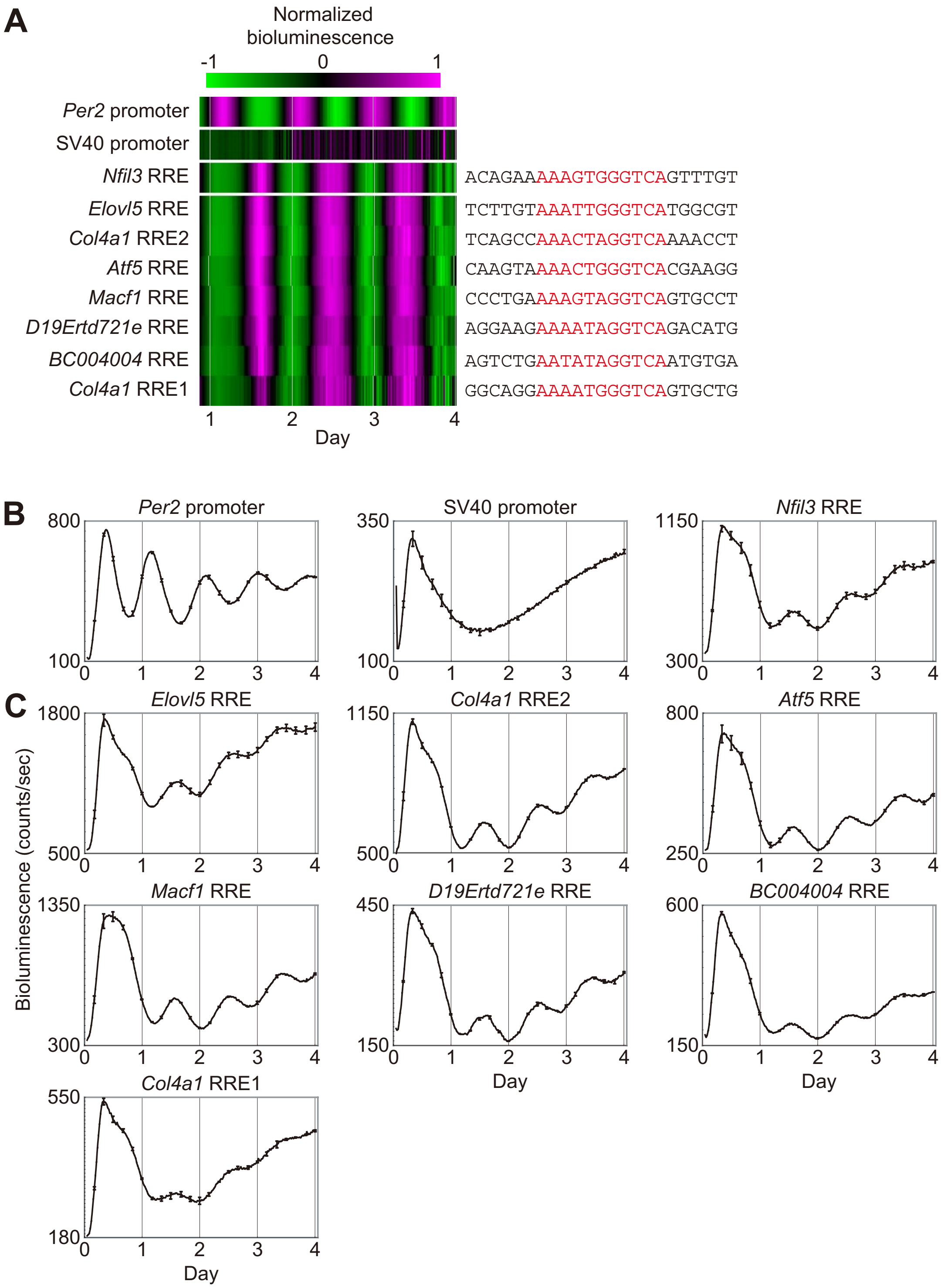


Fig. S2. Functional clock-controlled elements used to train HMMs. (*A*) Bioluminescence from functional clock-controlled elements fused to the SV40 basic promoter driving a dLuc reporter (SV40-dLuc) in NIH 3T3 fibroblasts. Known clock-controlled promoter (*Per2* promoter), known clock-controlled element fused to the SV40 basic promoter (*Nfil3* RRE), or the SV40 basic promoter alone (SV40 promoter) driving luciferase were used as controls. The colors in descending order from magenta to black to green represent the detrended bioluminescence. Columns represent time points, and rows represent the predicted elements on the designated genes. (*B* and *C*) Raw bioluminescence data of positive and negative controls (*B*), or functional clock-controlled elements inserted into SV40-dLuc reporters (*C*). Error bars indicate the standard error of mean (SEM) determined from independent experimental duplicates for each condition.

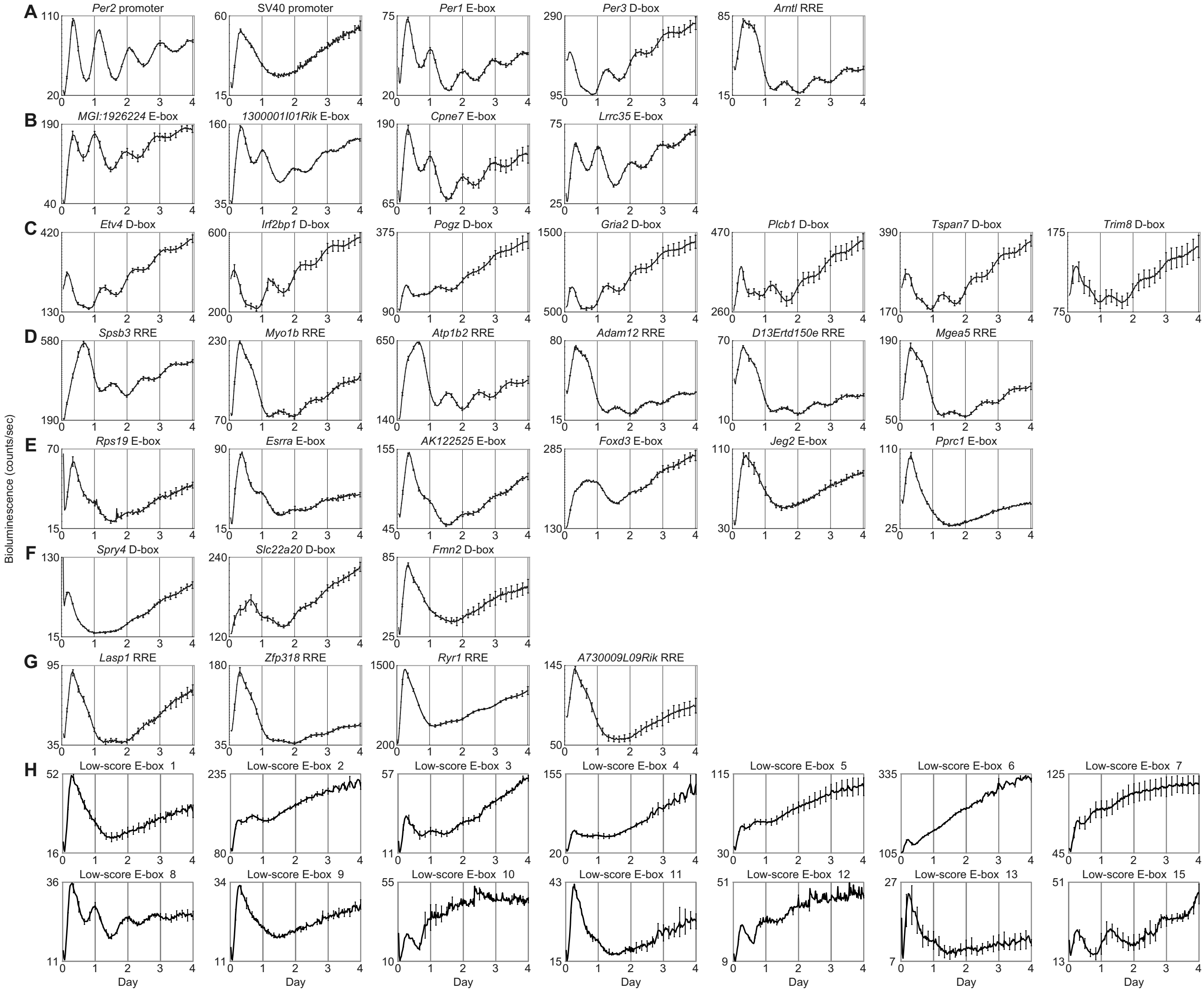
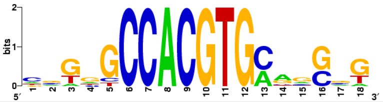


Fig. S3. Detailed information on experimental validation of HMM-based predictions at the cellular level. Raw bioluminescence data from positive and negative controls (*A*), and predicted clock-controlled elements inserted into the SV40-dLuc reporter that generate strong circadian transcriptional activity with high amplitude of the (*B*) E-box, (*C*) D-box, and (*D*) RRE, and weak circadian transcriptional activity with low amplitude or arrhythmic transcriptional activity of the (*E*) E-box, (*F*) D-box, and (*G*) RRE. Raw bioluminescence data from 14 low-score E-boxes inserted into the SV40-dLuc reporter (*H*). Only the low-score E-box No. 8 generated strong circadian transcriptional activity with high amplitude. The low-score E-box No. 14 was not measured due to the difficulty of reporter construction. Error bars indicate the standard error of mean (SEM) determined from independent experimental replicates ($n = 2$ or 3) for each element tested.

Pattern1	
Per1 E-box2	TAGAGCCACGTGAGGGCG
Per1 E-box3	TTTAGCCACGTGACAGTG
Per1 E-box5	CAAGTCCACGTGCAGGGA
Nr1d1 E-box1	CGGGCCCACGTGCTGCAT
Nr1d1 E-box2	GGGTGCCACGTGCGAGGG
Nr1d2 E-box1	ACTGGCCACGTGCACGGT
Dbp E-box	CCTCGCCACGTGAGTCCG
Bhlhb2 E-box	CCGGGCCACGTGAAGCGT
Rorc E-box	AGGTGCCACGTGCACCAG
High-scoring E-box	CGGGGCCACGTGCAGGCG

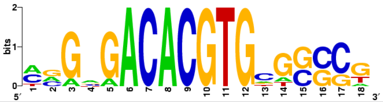
Pattern2	
Per1 E-box1	AGGGAACACGTGCAGGCT
Per1 E-box4	TAACGACACGTGGGCCCT
Nr1d2 E-box2	CGGAGACACGTGAGGCCG
MGI:1926224 E-box	AAGAGACACGTGCTGGGG
Lrrc35 E-box	ACGTGACACGTGCGGCGG
1300001I01Rik E-box	CGGGGACACGTGCGCGCA
Cpne7 E-box	CCGAGACACGTGTGCCCG

Fig. S4. Two patterns revealed in the adjacent sequence of high-amplitude E-boxes. High-amplitude E-box sequences classified into two patterns were shown. Information contents of the relative frequency of each nucleic acid at the position of the pattern were shown by sequence logos.

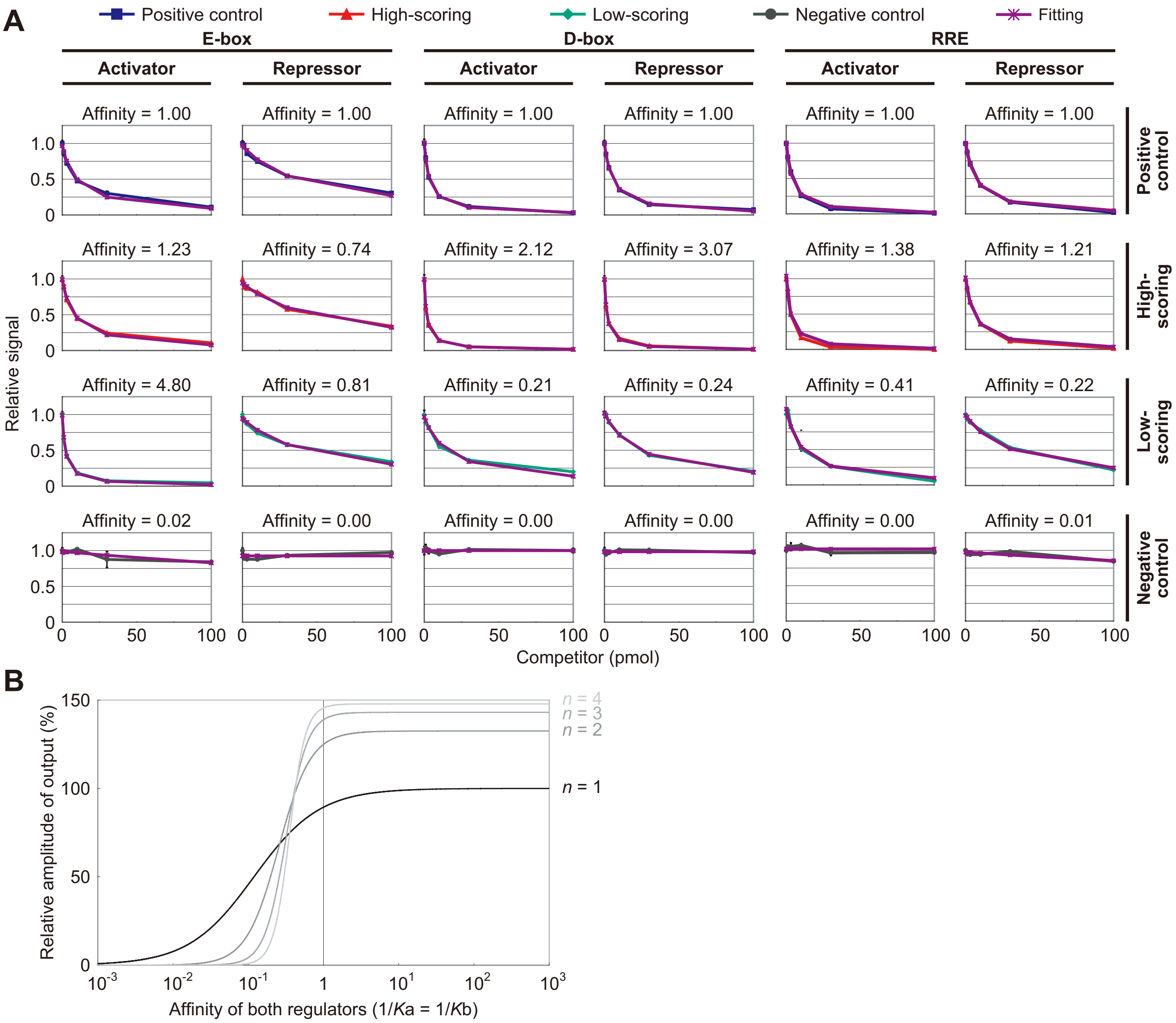


Fig. S5. Analysis of affinity to amplitude mechanism. (*A*) Affinity analysis of competitive binding data. The relative affinities of regulators for known clock controlled elements vs. synthetic elements were determined from competitive binding assay data shown in Fig. 4. A series of signal of binding between labeled oligonucleotide of positive control element and regulators, which were challenged with unlabeled oligonucleotides of positive control element (blue), high-scoring element (red), low-scoring element (green), or negative control element (black) in competition assays. This data were fitted by the model data (purple) using the least-squares method and the affinities relative to that of positive control were determined from fitting data. The value of relative affinities (positive control is 1.0) are indicated. See also *SI Appendix* for more detail. (*B*) *In silico* analysis of affinity to amplitude mechanism when affinity of activator and repressor are balanced. The relative amplitudes of oscillation of output plotted against strength of regulators binding affinity when the affinity of an activator and a repressor are the same; *n* indicates the Hill coefficient at competitive regulation. Amplitude was normalized so that the maximum value at $n = 1$ is 100%.

Other Supporting Information Files

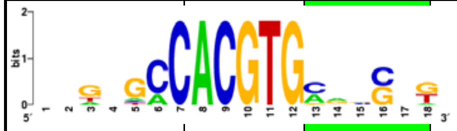
SI Appendix

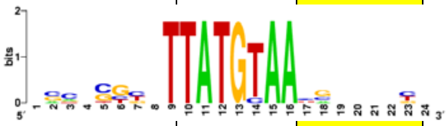
Table S1. Functional RREs from known clock-controlled genes used in this study

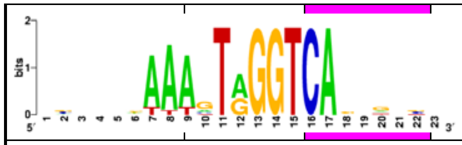
<table>
<tr><th>Gene</th><th>GeneID</th><th>Element</th><th>Organism</th><th>chr</th><th>Start</th><th>End</th><th>Sequence</th><th>Evidence</th><th>Affymetrix Probe ID</th></tr>
<tr><td rowspan="2"><i>Elov15</i></td><td rowspan="2">68801</td><td rowspan="2">RRE</td><td>Mouse</td><td>9</td><td>78317303</td><td>78317325</td><td rowspan="2">Mouse TCTTGTAAATTGGGTCATGGCGT
      | || ||| | ||||||| |||
Human TTGTGCAAAGTAGGTCATGCCGT</td><td rowspan="2">ref. 1, ref. 2</td><td rowspan="2">93496_at</td></tr>
<tr><td>Human</td><td>6</td><td>53273932</td><td>53273910</td></tr>
<tr><td rowspan="2"><i>BC004004</i></td><td rowspan="2">80748</td><td rowspan="2">RRE</td><td>Mouse</td><td>17</td><td>27851244</td><td>27851266</td><td rowspan="2">Mouse AGTCTGAATATAGGTCAATGTGA
        | |||| ||||||| | |||
Human TTTGGGAATCTAGGTCATTCTGA</td><td rowspan="2">ref. 1, ref. 2</td><td rowspan="2">95517_i_at</td></tr>
<tr><td>Human</td><td>6</td><td>36927126</td><td>36927148</td></tr>
<tr><td rowspan="2"><i>Macf1</i></td><td rowspan="2">11426</td><td rowspan="2">RRE</td><td>Mouse</td><td>4</td><td>121994185</td><td>121994207</td><td rowspan="2">Mouse CCCTGAAAAGTAGGTCAGTGCCT
      |||||||||||||||||| ||||
Human CCCTGAAAAGTAGGTCAGCGCCT</td><td rowspan="2">ref. 2</td><td rowspan="2">98402_at</td></tr>
<tr><td>Human</td><td>1</td><td>39277738</td><td>39277716</td></tr>
<tr><td rowspan="2"><i>Atf5</i></td><td rowspan="2">107503</td><td rowspan="2">RRE</td><td>Mouse</td><td>7</td><td>32213618</td><td>32213640</td><td rowspan="2">Mouse CAAGTAAAACTGGGTCACGAAGG
        || |||||||||||| |||
Human GGAGGGTAACTGGGTCACGCAGG</td><td rowspan="2">ref. 1, ref. 2</td><td rowspan="2">103006_at</td></tr>
<tr><td>Human</td><td>19</td><td>55118372</td><td>55118350</td></tr>
<tr><td rowspan="2"><i>Col4a1</i></td><td rowspan="2">12826</td><td rowspan="2">RRE1</td><td>Mouse</td><td>8</td><td>11309356</td><td>11309378</td><td rowspan="2">Mouse GGCAGGAAAATGGGTCAGTGCTG
       ||||||||||||||||||| ||
Human AGCAGGAAAATGGGTCAGTGATG</td><td rowspan="2">ref. 1, ref. 2</td><td rowspan="2">101093_at</td></tr>
<tr><td>Human</td><td>13</td><td>109748860</td><td>109748882</td></tr>
<tr><td rowspan="2"><i>Col4a1</i></td><td rowspan="2">12826</td><td rowspan="2">RRE2</td><td>Mouse</td><td>8</td><td>11337102</td><td>11337124</td><td rowspan="2">Mouse TCAGCCAAACTAGGTCAAAACCT
      ||||||||| |||||||||||
Human TCAGCCAAAATAGGTCAAAACAG</td><td rowspan="2">ref. 1, ref. 2</td><td rowspan="2">101093_at</td></tr>
<tr><td>Human</td><td>13</td><td>109781193</td><td>109781215</td></tr>
<tr><td rowspan="2"><i>D19Ertd721e</i></td><td rowspan="2">225896</td><td rowspan="2">RRE</td><td>Mouse</td><td>19</td><td>7972840</td><td>7972862</td><td rowspan="2">Mouse AGGAAGAAAATAGGTCAGACATG
      | ||||||| ||||||||||||
Human AAGAAGAAATTAGGTCAGACATC</td><td rowspan="2">ref. 1, ref. 2</td><td rowspan="2">97240_g_at, 97241_at</td></tr>
<tr><td>Human</td><td>11</td><td>62201916</td><td>62201894</td></tr>
<tr><td colspan="10">References
1. Ueda HR, <i>et al</i>. (2002) A transcription factor response element for gene expression during circadian night. <i>Nature</i> 418:534-539.
2. Panda S, <i>et al</i>. (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. <i>Cell</i> 109:307-320.</td></tr>
</table>

Element information on the gene symbol ('Gene'), NCBI GeneID ('GeneID'), element type ('Element'), the position ('Start' and 'End') on a chromosome ('chr') of mouse and human ('Organism') and the sequence alignment of human-mouse RREs ('Sequence') are indicated. Experimental evidence on clock-controlled genes ('Evidence') and Affymetrix probe ID detecting circadian expression of the genes ('Affymetrix Probe ID') are also indicated. The canonical consensus sequences for the RRE ([A/T]A[A/T]NT[A/G]GGTCA) is indicated in red. The function of these mouse RREs sequences in the context of a luciferase reporter was experimentally determined in **Fig. S2**.

Table S2. Functional clock-controlled elements used as train HMMs

Gene	GeneID	Element	Organism	chr	Start	End	Score of HMMER	Sequence	Evidence
Per1	18626	E-box 1	Mouse	11	68707450	68707433	15.39	AGGGAACACGTGCAGGCT	ref. 1
Per1	18626	E-box 2	Mouse	11	68707634	68707617	17.87	TAGAGCCACGTGAGGGCG	ref. 1
Per1	18626	E-box 3	Mouse	11	68710224	68710241	11.48	TTTAGCCACGTGACAGTG	ref. 1
Per1	18626	E-box 4	Mouse	11	68710970	68710987	12.10	TAACGACACGTGGGCCCT	ref. 1
Per1	18626	E-box 5	Mouse	11	68711333	68711350	13.45	CAAGTCCACGTGCAGGGA	ref. 2
Nr1d1	217166	E-box 1	Mouse	11	98445288	98445271	14.71	CGGGCCCACGTGCTGCAT	ref. 2
Nr1d1	217166	E-box 2	Mouse	11	98445008	98444991	15.77	GGGTGCCACGTGCGAGGG	ref. 2
Nr1d2	353187	E-box 1	Mouse	14	14904545	14904528	16.38	ACTGGCCACGTGCACGGT	ref. 2
Nr1d2	353187	E-box 2	Mouse	14	14904442	14904425	17.71	CGGAGACACGTGAGGCCG	ref. 2
Dbp	13170	E-box	Mouse	7	33110149	33110166	15.23	CCTCGCCACGTGAGTCCG	ref. 2
Bhlhb2	20893	E-box	Mouse	6	109049680	109049697	18.55	CCGGGCCACGTGAAGCGT	ref. 2
Rorc	19885	E-box	Mouse	3	94357502	94357519	16.45	AGGTGCCACGTGCACCAG	ref. 2
									
Per1	18626	D-box	Mouse	11	68711473	68711496	17.30	GCCTGGCATTATGCAACCCGCCTC	ref. 2
Per2	18627	D-box	Mouse	1	91377818	91377795	16.48	TGTGCGTCTTATGTAAAGAGAGCG	ref. 2
Per3	18628	D-box 1	Mouse	4	148930820	148930797	18.62	CCCGCGCGTTATGTAAGGTACTCG	ref. 2
Per3	18628	D-box 2	Mouse	4	148930773	148930750	16.10	GCCCGCGGTTATGTAACCCCCGCC	ref. 2
Nr1d1	217166	D-box	Mouse	11	98445195	98445218	19.41	GGAGCTCATTATGTAACGAGGCCG	ref. 2

Nr1d2	353187	D-box	Mouse	14	14904456	14904479	17.41	AGCTCGCATTATGTAATGCTGCGT	ref. 2
Rora	19883	D-box 1	Mouse	9	69195245	69195222	11.78	AAGCTGTTTTATGTAATAGCTTTG	ref. 2
Rora	19883	D-box 2	Mouse	9	68926158	68926135	16.84	CACTGCTGTTATGTAACCAAACGT	ref. 2
Rora	19883	D-box 3	Mouse	9	68894279	68894302	17.42	CGAGCGGGTTATGTAACAGGGTTA	ref. 2
Rorb	225998	D-box	Mouse	19	18304970	18304993	14.84	TCCAGTTCTTATGTAATGAATATA	ref. 2
									
Arntl	11865	RRE	Mouse	7	100478066	100478088	19.72	AGGCAGAAAGTAGGTCAGGGACG	ref. 2
Npas2	18143	RRE	Mouse	1	39516271	39516293	14.66	GAAAAATATGTAGGTCAGTGGAA	ref. 2
Nfil3	114519	RRE 1	Rat	17	18094293	18094271	15.64	AGTGTGTTAGTAGGTCAGTTCCG	ref. 2
Nfil3	18030	RRE 2	Mouse	13	52011542	52011520	21.79	ACAGAAAAAGTGGGTCAGTTTGT	ref. 2
Clock	12753	RRE	Mouse	5	75037420	75037442	15.91	AGGAATAAAGTGGGTCACAAGGC	ref. 2
Cry1	12952	RRE 1	Mouse	10	84829678	84829656	20.81	GACTAGAAAGTAGGTCATTGTGA	ref. 2
Cry1	12952	RRE 2	Mouse	10	84829601	84829623	16.47	GTTTCTAAAGTAGGTCATCGCTA	ref. 2
Rorc	19885	RRE	Mouse	3	94352279	94352301	19.40	GGAATAAAAGTGGGTCATCTTGT	ref. 2
Elov15	68801	RRE	Mouse	9	78317303	78317325	17.19	TCTTGTAAATTGGGTCATGGCGT	This study
BC004004	80748	RRE	Mouse	17	27851266	27851244	17.53	AGTCTGAATATAGGTCAATGTGA	This study
Macf1	11426	RRE 2	Mouse	4	121994185	121994207	20.05	CCCTGAAAAGTAGGTCAGTGCCT	This study
Atf5	107503	RRE	Mouse	7	32213618	32213640	14.65	CAAGTAAAACTGGGTCACGAAGG	This study
Col4a1	12826	RRE 1	Mouse	8	11309378	11309356	19.92	GGCAGGAAAATGGGTCAGTGCTG	This study
Col4a1	12826	RRE 2	Mouse	8	11337124	11337102	14.10	TCAGCCAAACTAGGTCAAAACCT	This study
D19Ertd721e	225896	RRE	Mouse	19	7972862	7972840	18.30	AGGAAGAAAATAGGTCAGACATG	This study

References

1. Hida A, *et al*. (2000) The human and mouse Period1 genes: five well-conserved E-boxes additively contribute to the enhancement of mPer1 transcription. *Genomics* 65:224-233.
   and unpabsihed data of Ueda, H.R. *et al*.
2. Ueda HR, *et al*. (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat Genet* 37:187-192.

Element information on the gene symbol ('Gene'), NCBI GeneID ('GeneID'), element type ('Element'), the position ('Start' and 'End') on a chromosome ('chr') of mouse and rat ('Organism') and the sequence ('Sequence') are indicated. Experimental evidence on functional clock-controlled elements ('Evidence') and match score of HMMER search performed using the constructed HMMs ('Score of HMMER') is also indicated. Canonical consensus sequences for the E-box (CACGTG), D-box (TTATG[T/C]AA) and RRE ([A/T]A[A/T]NT[A/G]GGTCA) are indicated in red. Information contents of the relative frequency of each nucleic acid at the position of the pattern for each element were shown by sequence logos.

Table S3. Predicted clock-controlled elements used for experimental validation

<table>
<tr><th>No.</th><th>Gene</th><th>GeneID</th><th>Element</th><th>FDR</th><th>Organism</th><th>chr</th><th>Start</th><th>End</th><th>Score of HMMER</th><th>Sequence</th></tr>
<tr><td rowspan="2">-</td><td rowspan="2"><i>Per1</i></td><td rowspan="2">18626</td><td rowspan="2">E-box</td><td rowspan="2">0.043</td><td>Mouse</td><td>11</td><td>68711333</td><td>68711350</td><td>13.45</td><td rowspan="2">Mouse CAAGTCCACGTGCAGGGA
|| ||||||||||
Human CAGGTCCACGTGCGCCCG</td></tr>
<tr><td>Human</td><td>17</td><td>7996626</td><td>7996609</td><td>16.13</td></tr>
<tr><td rowspan="2">1</td><td rowspan="2"><i>Esrra</i></td><td rowspan="2">26379</td><td rowspan="2">E-box</td><td rowspan="2">0.002</td><td>Mouse</td><td>19</td><td>6635826</td><td>6635809</td><td>18.03</td><td rowspan="2">Mouse CGGACGCACGTGGCCCCG
||||||||||||||||||
Human CGGACGCACGTGGCCCCG</td></tr>
<tr><td>Human</td><td>11</td><td>63828852</td><td>63828869</td><td>18.03</td></tr>
<tr><td rowspan="2">2</td><td rowspan="2"><i>Foxd3</i></td><td rowspan="2">15221</td><td rowspan="2">E-box</td><td rowspan="2">0.014</td><td>Mouse</td><td>4</td><td>97986368</td><td>97986351</td><td>17.23</td><td rowspan="2">Mouse CGCGGTCACGTGGCCCCG
| ||||||||||||||||
Human CACGGTCACGTGGCCCCG</td></tr>
<tr><td>Human</td><td>1</td><td>63500197</td><td>64500214</td><td>15.64</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2"><i>MGI:1926224</i></td><td rowspan="2">268859</td><td rowspan="2">E-box</td><td rowspan="2">0.017</td><td>Mouse</td><td>16</td><td>6032279</td><td>6032296</td><td>14.92</td><td rowspan="2">Mouse AAGAGACACGTGCTGGGG
||| ||||||||| || |
Human AAGGGACACGTGCGGGCG</td></tr>
<tr><td>Human</td><td>16</td><td>6473013</td><td>6473030</td><td>17.35</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2"><i>AK122525</i></td><td rowspan="2">331623</td><td rowspan="2">E-box</td><td rowspan="2">0.021</td><td>Mouse</td><td>10</td><td>43388228</td><td>43388211</td><td>17.00</td><td rowspan="2">Mouse CCCCCGCACGTGGCGCCC
||||| ||||||||||||
Human CCCCCCCACGTGGCGCCC</td></tr>
<tr><td>Human</td><td>6</td><td>107543034</td><td>107543051</td><td>15.19</td></tr>
<tr><td rowspan="2">5</td><td rowspan="2"><i>Lrrc35</i></td><td rowspan="2">272589</td><td rowspan="2">E-box</td><td rowspan="2">0.021</td><td>Mouse</td><td>9</td><td>42544300</td><td>42544317</td><td>16.09</td><td rowspan="2">Mouse ACGTGACACGTGCGGCGG
||||||||||||||||||
Human ACGTGACACGTGCGGCGG</td></tr>
<tr><td>Human</td><td>11</td><td>120399989</td><td>120399972</td><td>16.09</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2"><i>1300001I01Rik</i></td><td rowspan="2">74148</td><td rowspan="2">E-box</td><td rowspan="2">0.021</td><td>Mouse</td><td>11</td><td>74262482</td><td>74262499</td><td>15.91</td><td rowspan="2">Mouse CGGGGACACGTGCGCGCA
||||||||||||||||||
Human CGGGGACACGTGCGCGCA</td></tr>
<tr><td>Human</td><td>17</td><td>2561349</td><td>2561332</td><td>15.91</td></tr>
<tr><td rowspan="2">7</td><td rowspan="2"><i>Cpne7</i></td><td rowspan="2">102278</td><td rowspan="2">E-box</td><td rowspan="2">0.023</td><td>Mouse</td><td>8</td><td>122590893</td><td>122590910</td><td>13.74</td><td rowspan="2">Mouse CCGAGACACGTGTGCCCG
||||| |||||| |||||
Human CCGAGCCACGTGCGCCCG</td></tr>
<tr><td>Human</td><td>16</td><td>88169660</td><td>88169677</td><td>17.87</td></tr>
<tr><td rowspan="2">8</td><td rowspan="2"><i>Pprc1</i></td><td rowspan="2">226169</td><td rowspan="2">E-box</td><td rowspan="2">0.024</td><td>Mouse</td><td>19</td><td>45397616</td><td>45397633</td><td>15.71</td><td rowspan="2">Mouse CGGGTCCACGTGGGGGCG
||||||||||||||||||
Human CGGGTCCACGTGGGGGCG</td></tr>
<tr><td>Human</td><td>10</td><td>103883430</td><td>103883447</td><td>15.71</td></tr>
</table>

9	*Jag2*	16450	E-box	0.024	Mouse	12	107682267	107682284	15.64	Mouse CCTGGCCACGTGGGCGCG |||||||||||||||||| Human CCTGGCCACGTGGGCGCG
					Human	14	104704954	104704971	15.64	
10	*Rps19*	20085	E-box	0.024	Mouse	7	13623927	13623944	15.62	Mouse CGCGGCCACGTGCGAGCG |||||||||||||||||| Human CGCGGCCACGTGCGAGCG
					Human	19	47056319	47056336	15.62	
-	*Per3*	18628	D-box	0	Mouse	4	148930797	148930820	18.62	Mouse CCCGCGCGTTATGTAAGGTACTCG |||||||||||||||| | | | Human CCCGCGCGTTATGTAACGCGCCCC
					Human	1	7778455	7778432	21.11	
1	*Gria2*	14800	D-box	0.051	Mouse	3	81194005	81193982	17.31	Mouse CGGGGCTGTTACATAATGCCCACC |||||||||||||||| ||||||| Human CGGGGCTGTTACATAACGCCCACC
					Human	4	158501155	158501178	17.63	
2	*Spry4*	24066	D-box	0.122	Mouse	18	39043332	39043309	16.38	Mouse AGGTGCGTTTACATAACGCCGGGC ||||||||||||||||| || ||| Human AGGTGCGTTTACATAACACCAGGC
					Human	5	141683774	141683751	16.80	
3	*Etv4*	18612	D-box	0.181	Mouse	11	101455967	101455990	17.09	Mouse CACACGTCTTATGTAACCCAGTTC |||||||||||| || | Human ACACGTTCTTATGTAACCGAGCCC
					Human	17	38978809	38978832	15.20	
4	*Pogz*	229584	D-box	0.181	Mouse	3	94832639	94832662	16.89	Mouse CCCCTGTGTTATGTAATCCCGCTC ||| |||||||||||||| | Human CCCTTGTGTTATGTAATCTCTGCT
					Human	1	148244037	148244014	15.15	
5	*Fmn2*	54418	D-box	0.209	Mouse	1	174628879	174628902	14.89	Mouse ACCGCGCATTATGCAAAGCGGCAG ||||||||||||||||||||| | Human GCCGCGCATTATGCAAAGCGGCGG
					Human	1	236581260	236581283	16.89	
6	*Plcb1*	18795	D-box	0.325	Mouse	2	134535724	134535747	15.25	Mouse GGGGCGCGTTATGCAATGGGGCGC |||||||||||||||||||| Human GGGGCGCGTTATGCAATGGGCGCA
					Human	20	8061473	8061496	15.25	
7	*Irf2bp1*	272359	D-box	0.363	Mouse	7	10526613	10526636	16.21	Mouse CCCGCGCGTTATGTAACTTTCCCT | ||| ||||||||||||||||| Human CAGGCGTGTTATGTAACTTTCCCT
					Human	19	51081280	51081257	13.67	
8	*Slc22a20*	381203	D-box	0.363	Mouse	19	5775101	5775078	14.95	Mouse CTGCCTTTTTACATAAGGCCTGGG | |||| | ||||||| ||| ||| Human CGGCCTCTCTACATAAAGCCGGGG
					Human	11	64737873	64737896	14.81	
9	*Trim8*	93679	D-box	0.387	Mouse	19	45844359	45844336	14.62	Mouse GACACTCATTACATAAACAGCAGC |||||||||||||||||||||||| Human GACACTCATTACATAAACAGCAGC

					Human	10	104395331	104395308	14.62	
10	*Tspan7*	21912	D-box	0.388	Mouse	X	8817583	8817560	13.68	Mouse AGAGGGTCTTACATAAGCCAGGGG ||||||||||||||||||| |||| Human AGAGGGTCTTACATAAGCCGGGGG
					Human	X	38177725	38177702	15.27	
-	*Arntl*	11865	RRE	0.066	Mouse	7	100478066	100478088	20.38	Mouse AGGCAGAAAGTAGGTCAGGGACG ||||||||||||||||||||||| Human AGGCAGAAAGTAGGTCAGGGACG
					Human	11	13255933	13255955	20.38	
1	*Atp1b2*	11932	RRE	0.066	Mouse	11	69217609	69217631	20.61	Mouse GCGGAGAAAGTAGGTCACTGCCG |||||||||||||||||| |||| Human GCGGAGAAAGTAGGTCACAGCCG
					Human	17	7495878	7495856	19.61	
2	*Spsb3*	79043	RRE	0.083	Mouse	17	23484948	23484970	18.81	Mouse GAACGGAAAGTGGGTCAGCGCCG | ||||||||||||||| |||| Human GGGCGGAAAGTGGGTCAGGGCCG
					Human	16	1772556	1772534	19.64	
3	*Zfp318*	57908	RRE	0.104	Mouse	17	44564473	44564495	19.35	Mouse GAGAGAAAAGTGGGTCATTGAGA ||||||||||||||||||||| Human AAGAGAAAAGTGGGTCATTGAGC
					Human	6	43417992	43417970	18.25	
4	*Mgea5*	76055	RRE	0.112	Mouse	19	45123691	45123713	17.41	Mouse ATCTCCAAAGTAGGTCAGTGTCT | ||||||||||||| | Human CCATAGAAAGTAGGTCAGTTCTT
					Human	10	103568543	103568565	20.12	
5	*Lasp1*	16796	RRE	0.139	Mouse	11	97476169	97476191	16.54	Mouse TTCGTGAAAGTGGGTCATGGTCT | | |||||||||||||||||| Human TGTGAGAAAGTGGGTCATGGTCT
					Human	17	34284865	34284887	19.34	
6	*Ryr1*	20190	RRE	0.148	Mouse	7	18096374	18096352	17.94	Mouse AGGCTCTGACCTATTTAAATTCT || |||||||||||||||||||| Human AGACTCTGACCTATTTAAATTCT
					Human	19	43616749	43616771	17.45	
7	*Adam12*	11489	RRE	0.148	Mouse	7	121856769	121856791	15.44	Mouse TACTTAAAAGTAGGTCAGAAAAA |||| |||||||||||| || | Human AACTTGAAAGTAGGTCAGTAAGA
					Human	10	128077170	128077148	19.88	
8	*A730009L09Rik*	402727	RRE	0.168	Mouse	16	84900535	84900513	16.17	Mouse TCTTAATGACCCAATTTCTAAAT ||||||||||||| |||||||||| Human TCTTAATGACCCACTTTCTAAAT
					Human	21	25632353	25632331	18.49	
9	*Myo1b*	17912	RRE	0.179	Mouse	1	52270010	52269988	18.02	Mouse TGGTGCTGACCCACTTTCCTCTT | | ||||| ||||||| ||| Human GGATATTGACCTACTTTCCCCTT

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>Human</td><td>2</td><td>191953461</td><td>191953483</td><td>16.47</td><td></td></tr>
<tr><td rowspan="2">10</td><td rowspan="2"><i>D13Ertd150e</i></td><td rowspan="2">52548</td><td rowspan="2">RRE</td><td rowspan="2">0.203</td><td>Mouse</td><td>13</td><td>37218548</td><td>37218570</td><td>16.99</td><td rowspan="2">Mouse GATGCGAAAGTGGGTCAGGAATG
      || ||||||||||||||||||||
Human GACGCGAAAGTGGGTCAGGAATG</td></tr>
<tr><td>Human</td><td>6</td><td>6997855</td><td>6997877</td><td>16.67</td></tr>
<tr><td rowspan="2">1</td><td rowspan="2"><i>Shox2</i></td><td rowspan="2">20429</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>3</td><td>67055808</td><td>67055791</td><td>0.64</td><td rowspan="2">Mouse GATAAACACGTGTGTATC
      ||||||||||||| ||||
Human GATAAACACGTGTATATC</td></tr>
<tr><td>Human</td><td>3</td><td>159121615</td><td>159121598</td><td>0.64</td></tr>
<tr><td rowspan="2">2</td><td rowspan="2"><i>1700057H15Rik</i></td><td rowspan="2">78460</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>4</td><td>122900755</td><td>122900738</td><td>1.39</td><td rowspan="2">Mouse TATTCTCACGTGATAAAC
      ||||||||||||||||||
Human TATTCTCACGTGATAAAC</td></tr>
<tr><td>Human</td><td>1</td><td>38322547</td><td>38322564</td><td>1.39</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">-</td><td rowspan="2">435253</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>11</td><td>35821433</td><td>35821450</td><td>1.83</td><td rowspan="2">Mouse GACAAGCACGTGCCAGAC
      ||||||||||||||||||
Human GACAAGCACGTGCCAGAC</td></tr>
<tr><td>Human</td><td>5</td><td>167485618</td><td>167485601</td><td>1.83</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2"><i>Tbx15</i></td><td rowspan="2">21384</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>3</td><td>99126851</td><td>99126834</td><td>1.83</td><td rowspan="2">Mouse GTTTTTCACGTGCTTGAC
      |||||||||||| ||||
Human GTTTTTCACGTGTTTGAT</td></tr>
<tr><td>Human</td><td>1</td><td>119301655</td><td>119301672</td><td>2.39</td></tr>
<tr><td rowspan="2">5</td><td rowspan="2"><i>Stk4</i></td><td rowspan="2">58231</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>2</td><td>164147516</td><td>164147533</td><td>2.12</td><td rowspan="2">Mouse GAAGAGCACGTGATCTGC
      ||| | |||||| | |||
Human GAAAACCACGTGGTTTGC</td></tr>
<tr><td>Human</td><td>20</td><td>43041498</td><td>43041515</td><td>2.28</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2"><i>Ube1x</i></td><td rowspan="2">22201</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>X</td><td>19009993</td><td>19009976</td><td>2.61</td><td rowspan="2">Mouse GTTACTCACGTGAGGTAC
      ||||||||||||||||||
Human GTTACTCACGTGAGGTAC</td></tr>
<tr><td>Human</td><td>X</td><td>46825349</td><td>46825332</td><td>2.61</td></tr>
<tr><td rowspan="2">7</td><td rowspan="2">-</td><td rowspan="2">211823</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>3</td><td>129597003</td><td>129596986</td><td>2.80</td><td rowspan="2">Mouse TGCAAACACGTGATTTCC
      ||||||||||||||||||
Human TGCAAACACGTGATTTCC</td></tr>
<tr><td>Human</td><td>4</td><td>112078186</td><td>112078203</td><td>2.80</td></tr>
<tr><td rowspan="2">8</td><td rowspan="2"><i>Rnf44</i></td><td rowspan="2">105239</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>13</td><td>53770222</td><td>53770239</td><td>2.80</td><td rowspan="2">Mouse GTAAAACACGTGGATTTT
      ||||||||||||| ||||
Human GTAAAACACGTGGGTTTT</td></tr>
<tr><td>Human</td><td>5</td><td>175887078</td><td>175887095</td><td>2.80</td></tr>
<tr><td rowspan="2">9</td><td rowspan="2"><i>Pabpc1</i></td><td rowspan="2">18458</td><td rowspan="2">low-score E-box</td><td rowspan="2">-
-</td><td>Mouse</td><td>15</td><td>36663158</td><td>36663141</td><td>2.93</td><td rowspan="2">Mouse AAATACCACGTGTTGAAC
      ||||||||||||||||||
Human AAATACCACGTGTTGAAC</td></tr>
<tr><td>Human</td><td>8</td><td>101802352</td><td>101802335</td><td>2.93</td></tr>
<tr><td rowspan="2">10</td><td rowspan="2">-</td><td rowspan="2">-</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>16</td><td>74421611</td><td>74421628</td><td>3.03</td><td rowspan="2">Mouse GCTAAGCACGTGGAAGTC
      ||||||||||||||||||
Human GCTAAGCACGTGGAAGTC</td></tr>
<tr><td>Human</td><td>3</td><td>77581572</td><td>77581555</td><td>3.03</td></tr>
</table>

<table>
<tr><td rowspan="2">11</td><td rowspan="2"><i>C330011F03</i></td><td rowspan="2">328837</td><td rowspan="2">low-score E-box</td><td rowspan="2">--</td><td>Mouse</td><td>17</td><td>49541248</td><td>49541231</td><td>3.08</td><td rowspan="2">Mouse GTTTCCCACGTGTTTGGC
      ||||||||||||||||||
Human GTTTCCCACGTGTTTGGC</td></tr>
<tr><td>Human</td><td>3</td><td>17954468</td><td>17954451</td><td>3.08</td></tr>
<tr><td rowspan="2">12</td><td rowspan="2"><i>Prkce</i></td><td rowspan="2">18754</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>17</td><td>85008129</td><td>85008146</td><td>3.20</td><td rowspan="2">Mouse ATATAACACGTGCTAAAA
     ||||||||||||||||||
Human ATATAACACGTGCTAAAA</td></tr>
<tr><td>Human</td><td>2</td><td>46068028</td><td>46068045</td><td>3.20</td></tr>
<tr><td rowspan="2">13</td><td rowspan="2"><i>Map3k9</i></td><td rowspan="2">338372</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>12</td><td>77032273</td><td>77032290</td><td>2.44</td><td rowspan="2">Mouse GACAAACACGTGTGCGTC
      ||||||||||||| |
Human GACAAACACGTGTATGCA</td></tr>
<tr><td>Human</td><td>14</td><td>70276648</td><td>70276665</td><td>4.03</td></tr>
<tr><td rowspan="2">14</td><td rowspan="2"><i>Zfhx1b</i></td><td rowspan="2">24136</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>2</td><td>45057246</td><td>45057229</td><td>3.29</td><td rowspan="2">Mouse GTATTACACGTGAAAAGC
      ||||||||||||||||||
Human GTATTACACGTGAAAAGC</td></tr>
<tr><td>Human</td><td>2</td><td>145095987</td><td>145095970</td><td>3.29</td></tr>
<tr><td rowspan="2">15</td><td rowspan="2"><i>Ppfibp2</i></td><td rowspan="2">19024</td><td rowspan="2">low-score E-box</td><td rowspan="2">-</td><td>Mouse</td><td>7</td><td>94860226</td><td>94860243</td><td>4.93</td><td rowspan="2">Mouse GTTTCCCACGTGTGTCCC
      ||||||||||||| | |
Human GTTTCCCACGTGTTTGTC</td></tr>
<tr><td>Human</td><td>11</td><td>7489362</td><td>7489379</td><td>1.76</td></tr>
</table>

The element information including the gene symbol ('Gene'), NCBI GeneID ('GeneID'), element type ('Element'), the position ('Start' and 'End') on a chromosome ('chr') of mouse and human ('Organism'), the sequence alignments of human and mouse ('Sequence') and the match score of HMMER search ('Score of HMMER') are indicated. The estimated accuracy of the HMM-based prediction is indicated as false discovery rate ('FDR'). The mouse elements were used for experimental validation. The element type 'low-score E-box' indicate evolutionary conserved 15 most low-score E-boxes with the core consensus sequence 'CACGTG' in non-coding regions. The function of these element sequences in the context of a luciferase reporter was experimentally determined in **SI Fig. S3**.

Table S4. Similarity between DNA binding domains of DNA binding activators and repressors

Element	Gene	Alignment	Accession	Domain position	Domain name	Type	Homology
E-box	*Arntl*	-KNAREAHSQIEKRRRDKMNSFIDELASLVPTCNAM--SRKLDKLTVLRMAVQHMKTLR----	NP_031515.1	71-126	HLH	Activator	Identity 30%, Similarity 57% to *Bhlhb2*
	Clock	DKAKRVSRNKSEKKRRDQFNVLIKELGSMLPGN-----ARKMDKSTVLQKSIDFLRKHKE---	NP_031741.1	32-86		Activator	Identity 22%, Similarity 55% to *Bhlhb2*
	Bhlhb2	KETYKLPHRLIEKKRRDRINECIAQLKDLLPEHLKLTTLGHLEKAVVLELTLKHVKALTNLID	NP_035628.1	50-112		Repressor	-
		: : .: **:***::* * :* .::* :::* .**. ::..::					
D-box	*Dbp*	KDEKYWSRRYKNNEAAKRSRDARRLKENQISVRAAFLEKENALLRQEVVAVRQE	NP_058670.1	254-307	bZIP_2	Activator	Identity 44%, Similarity 69% to *Nfil3*
	Nfil3	KDAMYWEKRRKNNEAAKRSREKRRLNDLVLENKLIALGEENATLKAELLSLK--	NP_059069.1	72-123		Repressor	-
		** **.:* **********: ***:: :. : * :*** *: *:::::					
RRE	*Rora*	IPCKICGDKSSGIHYGVITCEGCKGFFRRSQQSNATYS-CPRQKNCLIDRTSRNRCQHCRLQKCLAVGMSRDAVKFG	NP_038674.1	71-146	zf-C4	Activator	Identity 65%, Similarity 81% to *Nr1d1*
	Nr1d1	-LCKVCGDVASGFHYGVHACEGCKGFFRRSIQQNIQYKRCLKNENCSIVRINRNRCQQCRFKKCLSVGMSRDAVRFG	NP_663409.2	132-207		Repressor	
		:* :**:**** :*********** *.* *. * :::** * * .*****:** ::***:********:**					

The gene symbol ('Gene') of DNA binding activators and repressors, target element ('Element'), alignment ('Alignment') of DNA binding domain, accession No. ('Accession') of protein sequence, position of DNA binding domain ('Domain position') at the protein sequence and name of the DNA binding domain ('Domain name') are indicated. Homology ('Homology') between DNA binding domains of DNA binding activators and repressors is also indicated.