

Supporting Information

Zhang et al. 10.1073/pnas.1408886111

SI Methods

Quantifying and Aligning RNA-Sequencing Data. Fastq files containing raw RNA-seq reads were aligned to the mouse genome (mm9/NCBI37) by using STAR (1) (default parameters). RNA-seq quantification was performed by using HTSeq (2), which was run in stranded mode (default parameters). Protein-coding genes were quantified by using the Ensembl annotation (3). Noncoding RNAs were quantified by using data from the NONCODE v3 database (4). Quantification values were normalized by using DESeq2 (5).

Identifying Noncoding RNAs Conserved Between Humans and Mice. We began by downloading BED files listing ncRNA coordinates for humans and mice from the NONCODE v3 database. These bed files contained 33,801 human and 36,991 mouse transcripts. To prevent overlapping ncRNAs from confounding the analysis (many of these appeared to be alternative spliceforms of the same ncRNAs), we merged all overlapping ncRNAs on the same strand by using the BEDTools suite (6). This merge step resulted to 20,042 human and 27,286 mouse transcripts. Using the coordinates for these merged transcripts and the University of California, Santa Cruz (UCSC) Genome Browser (7), we downloaded the nucleotide sequences corresponding to each of these ncRNAs in FASTA format. Next, we constructed separate human and mouse BLAST libraries from these ncRNA sequences by running the *makeblastdb* command with default parameters. Following this operation, we used BLAST (8) to align the mouse ncRNA sequences against the human ncRNA BLAST library, and vice versa. Because ncRNAs have been shown to have relaxed constraints on sequence conservation (9), we ran *blastn* by using the more permissive dc-megablast algorithm and a minimum *e*-value cutoff of $1e^{-10}$. We then mined these BLAST results for pairs of human and mouse ncRNAs that were each other's top BLAST hit (termed "reciprocal best hits"). Filtering for these reciprocal best hits left us with 1,601 human and mouse transcript pairs, we termed conserved ncRNAs. We are confident in our ability to identify conserved ncRNAs by using these relaxed BLAST parameters as we successfully found well-known, conserved ncRNAs like *Xist*, *Tsix*, *Hotair*, *H19*, and *Gas5*.

To assign names and annotation data to these conserved ncRNAs, we used BLAST to align their sequences to human and mouse RefSeq (10) transcripts. We found 585 of these conserved ncRNAs mapped to protein-coding genes (i.e., RefSeq IDs beginning with NM or XM) in the sense orientation in both humans and mice. Upon visual inspection of these ncRNAs, we found that many of these mapped along the entire length of the protein-coding transcripts. Although some ncRNAs in this list might represent noncoding isoforms of these protein-coding transcripts, we chose to take a conservative approach and removed these from further analysis. Following the removal of these transcripts, we were left with a final list of 1,016 conserved ncRNAs. We assigned biotypes [defined by GENCODE (11) and Ensembl] to these transcripts by using both Ensembl and manual annotation. Quantification and analysis of these transcripts was performed like all other RNA-seq transcript data, as described in *Quantifying and Aligning RNA-Sequencing Data*.

Identifying Novel ncRNAs. Given that RNA-seq data are not limited to a specific gene annotation, we sought to characterize novel transcripts. We began by collecting all reads that mapped across splice junctions (i.e., reads with large gaps in their alignments). Reads falling into this class are identified by STAR during alignment and stored in files with the *SI.out.tab* extension. Although this strategy will cause us to miss single-exon transcripts,

we have greater confidence that the data comes from a real, expressed transcripts if we see evidence of RNA splicing. To reduce the impact of spurious reads and noise, we required that splice junctions be mapped by a minimum of 16 reads across our entire dataset (this threshold corresponds to two reads per time point in a single organ). We chose a fairly low threshold so as not to remove junctions present in only a single organ and those circadian transcripts expressed in bursting patterns (like *Dbp*). Next, we used BEDTools to filter out any junction mapping within 1 kb of any Ensembl or RefSeq transcript, or overlapping with any NONCODE transcript. All of these steps left us with 10,452 junctions from putative transcripts. We merged all junctions within 500 bp of each other to form 5,154 putative, ncRNA transcript regions. These putative transcripts were quantified and analyzed like all other RNA-seq transcripts, as described in *Quantifying and Aligning RNA-Sequencing Data*.

Disease Genes, Drug Targets, and Other Data Sources. Disease-gene annotations were aggregated from the following sources: Online Mendelian Inheritance in Man (12), Universal Protein Resource (13), Comparative Toxicogenomics Database (14), Pharmacogenomics KnowledgeBase (15), and Literature-Derived Human Gene-Disease Network (16). Drug target genes were pulled from the DrugBank database (17). List of WHO essential medicines were downloaded from the WHO website (18). National Institutes of Health (NIH) funding estimates were for fiscal year 2013 from the NIH Research Condition, and Disease Categories system (19). MicroRNA target predictions for PTGS1 were from TargetScan (20).

Tissue Culture and Cell Maintenance. NIH 3T3 cells were purchased from ATCC. These cells were maintained in growth media containing 10% FBS (Atlanta Biologicals) by volume, 1× Penicillin/Streptomycin/Glutamine (Gibco), and 1× nonessential amino acids (NEAA; Gibco) in Dulbecco's modified Eagle's medium (DMEM; Gibco). Cells were grown in a humidified incubator at 37 °C and 5% (vol/vol) CO₂.

Transfections. All transfections were performed in the forward format. Briefly, cells were seeded in six-well dishes at a density of 2.5×10^5 cells per well, in media containing no antibiotics [DMEM, 10% FBS by volume, 1× Glutamine (Gibco), and 1× NEAA]. Cells were incubated overnight at 37 °C and 5% (vol/vol) CO₂. On the following day, cells were transfected by using Opti-MEM (Gibco) and the RNAiMAX (Invitrogen) reagent, according to the manufacturers' protocol. Cells were transfected with mirVana Negative Control #1, mmu-miR-22-3p mimic, or mmu-miR-22-5p mimic (Life Technologies) at a final concentration of 50 nM. Transfected cells were incubated for 72 h at 37 °C and 5% (vol/vol) CO₂. RNA and protein were harvested from the same well by collecting cells in ice-cold PBS and dividing these cell suspensions into two aliquots. For each well, one aliquot was processed for protein, and the other was processed for RNA.

Western blot. Whole-cell protein extracts were isolated from cells by using ice-cold RIPA buffer (Sigma), supplemented with Complete protease inhibitor mixture (Roche). Protein concentrations were quantified by using the DC protein assay (BioRad). Four micrograms of protein was resolved on precast 7.5% polyacrylamide, Tris-HCl/Glycine/SDS gels (BioRad) and transferred to PVDF membranes. Membranes were blocked for 1 h at room temperature in blocking solution [5% (wt/vol) milk, 0.05% Tween 20, 1× Tris-buffer saline], followed by overnight incubation at 4 °C with primary antibody in blocking solution.

Primary antibodies used were as follows: anti-PTGS1 (160110; Cayman Chemical) and anti-GAPDH (sc-25778; Santa Cruz). Membranes were then rinsed twice each with TBS-0.05% Tween and blocking solution. Following rinses, membranes were probed with secondary antibody at room temperature for 70 min. Those membranes treated with anti-PTGS1 were incubated with anti-mouse IgG HRP-linked secondary antibodies (NA931V; GE Healthcare), whereas membranes treated with anti-GAPDH were incubated with anti-rabbit IgG HRP-linked secondary antibodies (NA934-1ML; GE Healthcare). Membranes were then rinsed five times for 10 min in TBS-0.05% Tween, and then imaged by using standard autoradiograph techniques after the application of Western Lightning Plus ECL (PerkinElmer) Western blotting detection reagent.

1. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15–21.
2. HTSeq (2013) Analysing high-throughput sequencing data with Python. Version 0.5.4p1.
3. Flicek P, et al. (2012) Ensembl 2012. *Nucleic Acids Res* 40(Database issue):D84–D90.
4. Bu D, et al. (2012) NONCODE v3.0: Integrative annotation of long noncoding RNAs. *Nucleic Acids Res* 40(Database issue):D210–D215.
5. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
6. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
7. Meyer LR, et al. (2013) The UCSC Genome Browser database: Extensions and updates 2013. *Nucleic Acids Res* 41(Database issue):D64–D69.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
9. Washietl S, Kellis M, Garber M (2014) Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res* 24(4):616–628.
10. Pruitt KD, Tatusova T, Klimke W, Maglott DR (2009) NCBI Reference Sequences: Current status, policy and new initiatives. *Nucleic Acids Res* 37(Database issue): D32–D36.
11. Harrow J, et al. (2012) GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 22(9):1760–1774.
12. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33(Database issue):D514–D517.
13. UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res* 41:D43–7.
14. Davis AP, et al. (2013) The Comparative Toxicogenomics Database: Update 2013. *Nucleic Acids Res* 41(Database issue):D1104–D1114.
15. Whirl-Carrillo M, et al. (2012) Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 92(4):414–417.
16. Bundschuh M, Dejori M, Stetter M, Tresp V, Kriegel H-P (2008) Extraction of semantic biomedical relations from text using conditional random fields. *BMC Bioinformatics* 9:207.
17. Law V, et al. (2014) DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res* 42(Database issue):D1091–D1097.
18. WHO (2013) WHO Model Lists of Essential Medicines WHO. Available at www.who.int/medicines/publications/essentialmedicines/en/.
19. NIH (2013) *Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC)*. Available at report.nih.gov/categorical_spending.aspx. Accessed July 9, 2013.
20. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1): 15–20.

RNA extraction and quantitative PCR. RNA was extracted from cells by using TRIzol reagent (Life Technologies) with Direct-zol RNA MiniPrep kit (Zymo Research), according to manufacturer's protocol, cDNA was generated from 500 ng of RNA by using the qScript cDNA Synthesis Kit (Quanta Biosciences), and qPCR was performed on the ViiA 7 Real-Time PCR System (Life Technologies) using the PerfeCTa FastMix II, Low ROX reagent (Quanta Biosciences), according to the manufacturer's protocols. Relative expression quantification of the qPCR data were performed by using the $\Delta\Delta CT$ method with the ViiA 7 analysis software v1.2 (Life Technologies). *Ptgs1* (Mm00477214_m1; Life Technologies) was quantified by using *Gapdh* (4352661; Life Technologies) as the endogenous reference.

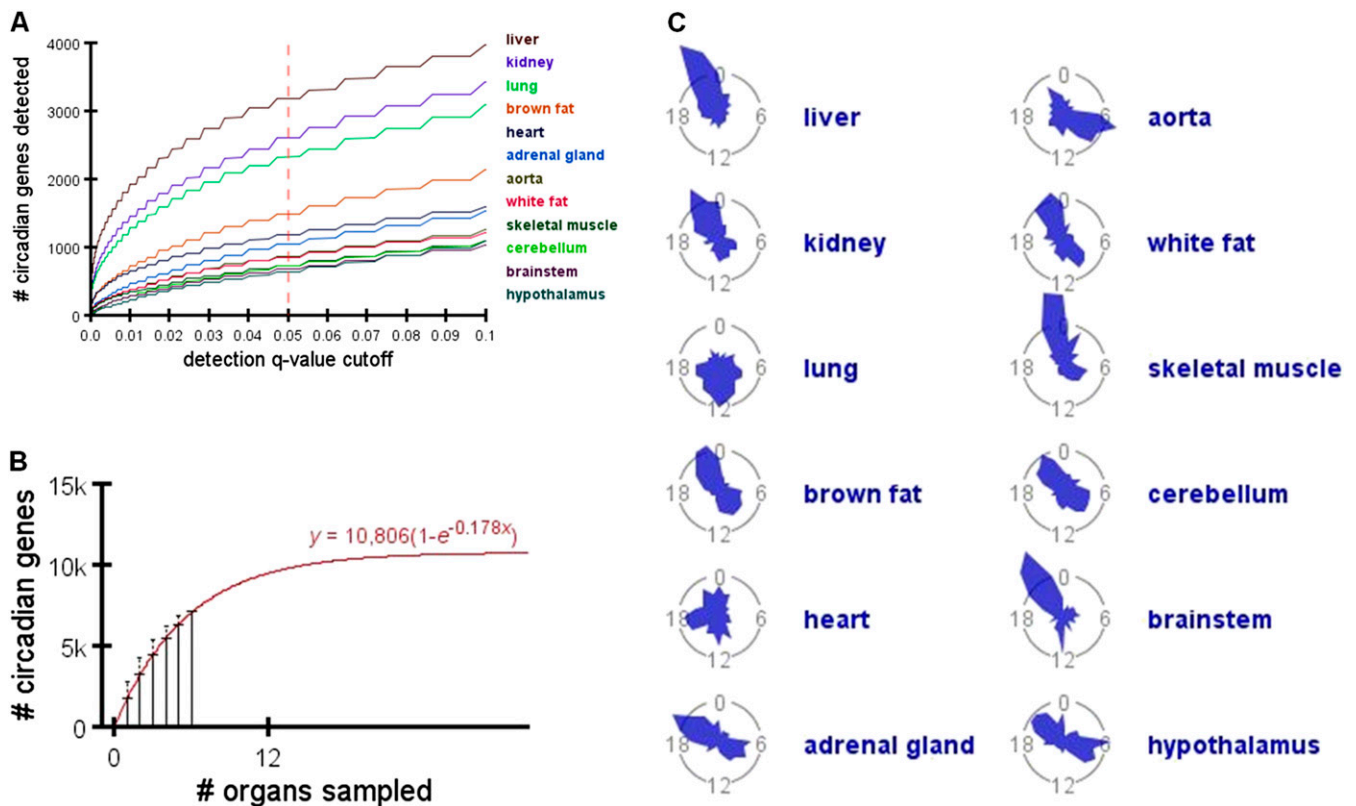


Fig. S1. Identifying oscillating transcripts across different organs. (A) The number of oscillating transcripts identified in each tissue as a function of the JTK_CYCLE significance cutoff. (B) The average total number of oscillating genes detected as a function of the number of organs sampled. This figure is a repetition of the calculation performed in Fig. 1C, using only data from organs with a more homogeneous composition (liver, kidney, lung, heart, adrenal gland, and skeletal muscle). Error bars represent SD. Best-fit model has been overlaid in red. (C) Radial diagrams showing the phase distribution of oscillating genes in each organ. These are enlarged versions of those included down the right side of Fig. 2A.

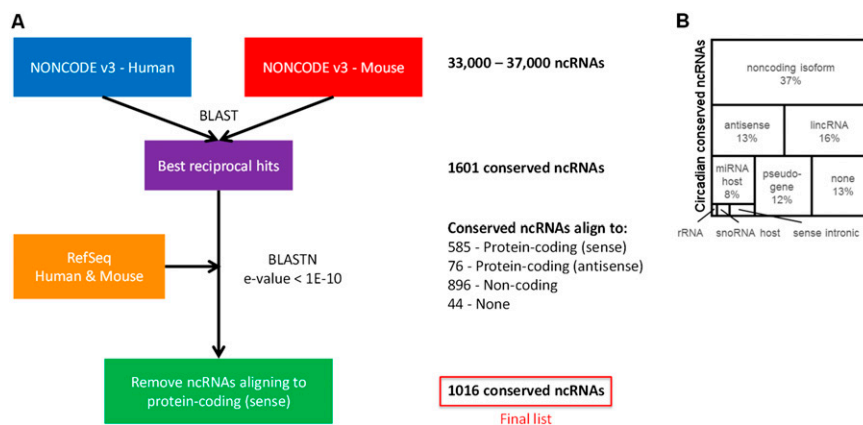


Fig. S2. Conserved circadian ncRNAs. (A) Method overview for identifying conserved ncRNAs. (B) Functional types of circadian conserved ncRNAs. Types were defined by GENCODE and Ensembl biotypes, assigned by using Ensembl and manual annotation.

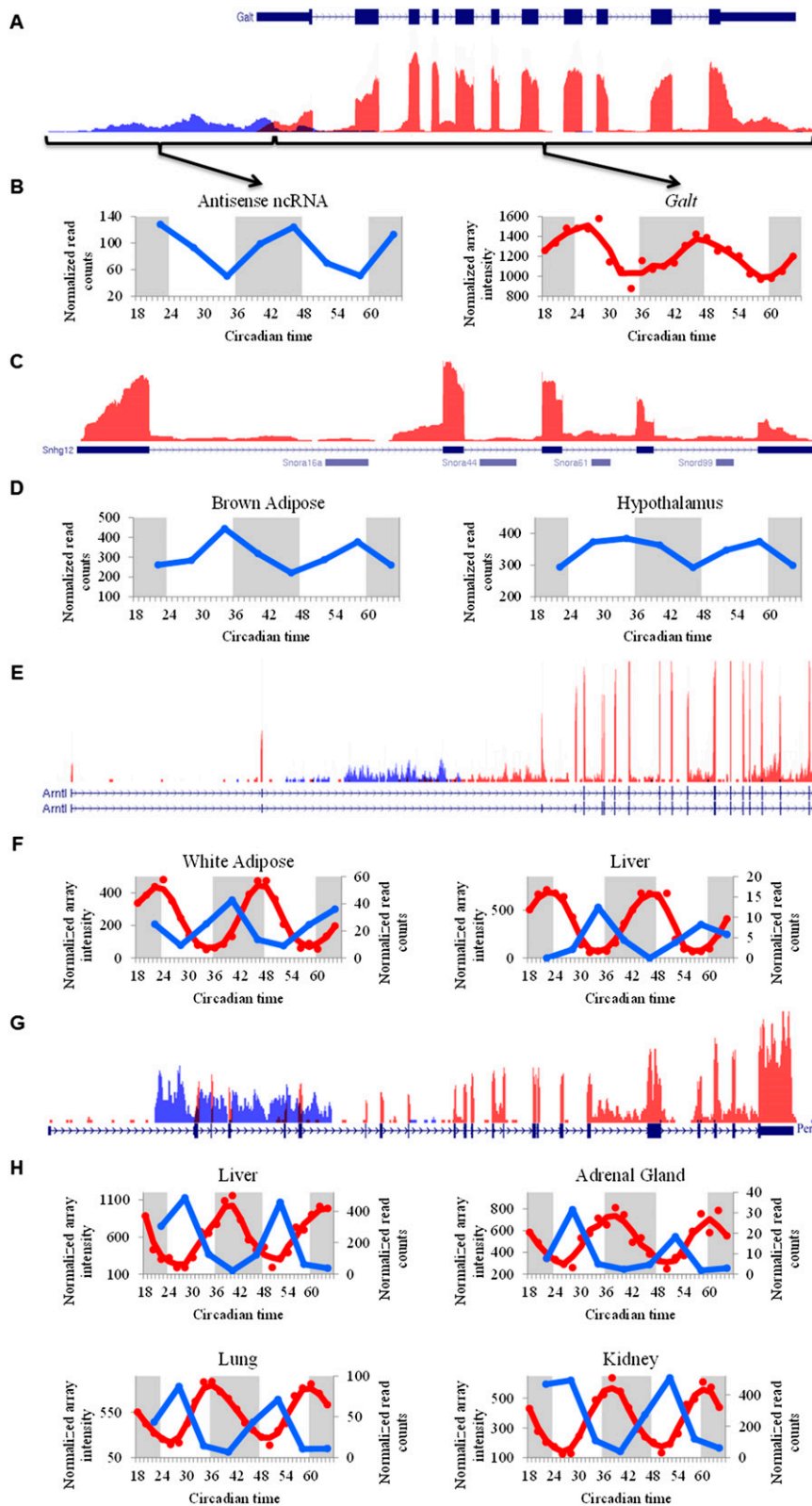


Fig. S3. Representative examples of conserved circadian ncRNAs and antisense transcripts. (A) RNA-seq coverage plot for *Galt* (red) and its antisense transcript (blue). The gene model for *Galt* is displayed above the coverage plots. (B) Expression profiles for *Galt* (red; data from microarrays) and the antisense transcripts (blue; data from RNA-seq). Gray regions indicate subjective night. (C) RNA-seq coverage plot for *Shhg12*. The gene model is displayed below the coverage plot. Note the locations of the mature snoRNA sequences located in the introns of *Shhg12*. (D) RNA-seq expression profiles for *Shhg12* in brown adipose and hypothalamus. (E) RNA-seq coverage plot for *Arntl* (red) and its antisense transcript (blue), from white adipose. The gene model for *Arntl* is displayed below the coverage plots. (F) Expression profiles for *Arntl* (red; data from microarrays) and the antisense transcripts (blue; data from RNA-seq), from white adipose and liver. (G) RNA-seq coverage plot for *Per2* (red) and its antisense transcript (blue), from white adipose. The gene model for *Per2* is displayed below the coverage plots. (H) Expression profiles for *Per2* (red) and the antisense transcript (blue) from liver, adrenal gland, lung, and kidney.

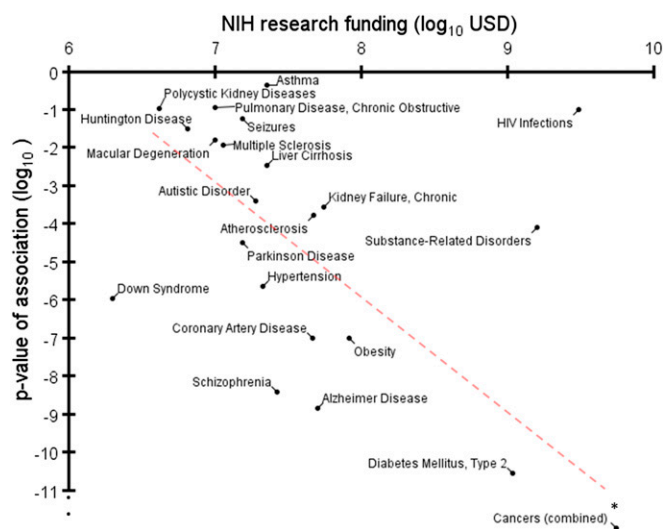


Fig. S8. Correlation between association significance of diseases with circadian genes and NIH research funding toward each disease.

Dataset S1. Drugs from the top-100 best-selling drugs list that target circadian genes

[Dataset S1](#)

Dataset S2. Drugs from the WHO Essential Medicines list that target circadian genes

[Dataset S2](#)

Dataset S3. Half-life notes on all drugs from Datasets S1 and S2

[Dataset S3](#)

Dataset S4. Oscillation phases of all conserved, non-coding RNAs

[Dataset S4](#)

Dataset S5. Antisense transcripts

[Dataset S5](#)