Supporting Information

Li et al. 10.1073/pnas.1305814110

SI Materials and Methods

Samples and Microarray Data Collection. Sample collection, including human subject recruitment and characterization, tissue dissection, and RNA extraction, was described previously (1, 2). Briefly, human subjects were recruited by the Brain Donor Program at the University of California, Irvine. Brain tissue was obtained with the consent of the next-of-kin of the deceased. Information obtained from medical examiners, coroners, medical records, and interviews of relatives was combined to record physical health, medication use, psychopathology, substance use, and details of death. To ensure the accuracy of psychiatric evaluation of the control subjects and the subjects with major depressive disorder (MDD), we relied not only on the subject's medical records but on a next-of-kin interview and a 141-item questionnaire administered to a family member. To minimize the confounding effect of agonal stress on gene expression, we assessed the agonal factor score (AFS) for each subject, defined as the degree of severity and duration of physiological stress at the time of death (TOD) (3). All subjects who were analyzed in this study had rapid death (i.e., occurred within 1 h) and an AFS of 0. The controls had no psychiatric or neurological disorders, substance abuse, or any first-degree relative with a psychiatric disorder. Patients who had MDD received a consensus diagnosis based on criteria from the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*.

Frozen coronal slabs of the brain were dissected to obtain tissue samples for specific regions. Total RNA was isolated and distributed to multiple Pritzker Consortium laboratories for replicate experiments. The microarray experiments were conducted at separate laboratories at three of the universities involved in this study: University of Michigan; University of California, Irvine; and University of California, Davis. RNA samples were analyzed on multiple microarray platforms, and in this work, we focused on Affymetrix GeneChip data from high-quality tissue (Table S3). Samples for different regions came from the same set of brains from 55 control subjects (Table S4). This group included 10 female and 45 male control subjects, with an average age of 55.8 y (SD = 13.9). The brain tissues had an average pH of 6.87 (SD = 0.23). Only two samples had a pH <6.5 (samples 6.4 and 6.3). However, some brain samples generated RNA only for a subset of regions (regions 3–6); as a result, different regions were analyzed with varying sample size (n = 29–55). Microarray experiments were performed in separate experimental cohorts, ranging from five to eight cohorts depending on the brain region. Each cohort contained a mixture of cases and controls, with most RNA samples analyzed in duplicate at two laboratories (some were analyzed in three laboratories). All laboratory procedures for running the Affymetrix GeneChips followed the manufacturer's standard labeling and hybridization protocols. The generation of probe-level intensity data (i.e., the .cel files) relied on standard Affymetrix library files, and these data were further processed using a custom annotation file (see below). Attributes files containing annotated clinical and sample quality information were maintained in an internal database.

Whereas this study focused on a subset of normal controls with TOD data [anterior cingulate cortex (AnCg; n = 55 controls), dorsolateral prefrontal cortex (DLPFC; n = 52), cerebellum (CB; n = 34), amygdala (AMY; n = 29), hippocampus (HC; n = 48), and nucleus accumbens (NAcc; n = 51)], a larger set of control samples is available that includes samples both with and without accompanying TOD data (70 AnCg, 83 DLPFC, 51 CB, 32 AMY, 63 HC, and 66 NAcc samples).

Information for the 34 patients with MDD (27 male and 7 female) is included in Table S2. The mean pH value was 6.91 (SD = 0.27). The mean age of the patients with MDD was 46.3 y (SD = 15.1). There were no significant differences between the control and MDD groups for pH values (P = 0.429). Twenty (59%) of the 34 patients with MDD died by suicide, 11 (32%) by naturally occurring sudden cardiac death, and 3 (9%) by multidrug overdose of undetermined cause (either suicide or accidental).

Toxicology screens were performed by the coroner's office, following a standard protocol in which bodily fluids (blood, urine, ocular fluid, or spinal fluid) were submitted for in-house screening across a panel of ~140 compounds. On positive screening results, a case-specific screen was ordered at the discretion of the county medical examiner to be completed at a commercial laboratory (NMS Labs, Inc.) for more quantitative measurements in a similar panel of 140 compounds. Samples submitted to the commercial laboratory were homogenized, archived brain tissue. The specific panel performed was Postmortem Toxicology-Expanded, Tissue (Forensic) Test (8052TI). Assays included headspace GC, ELISA, GC/MS, and colorimetry. Results are shown in Table S2. Results for 15 (44%) subjects with MDD were negative; of the remaining 19 (56%) positive cases, 4 (12% of the 34 cases) had lethal doses or lethal combinations of drugs, whereas the other 15 (44%) had treatment drugs within therapeutic levels.

The raw data and processed data for this complete set of controls were deposited in the National Center for Biotechnology Information Gene Expression Omnibus (accession no. GSE45642) and on our Web site (www.pritzkerneuropsych.org/?page_id=1196).

Data Processing. Data for each brain region were processed separately, using both cases (including MDD, bipolar disorder, and schizophrenia) and controls, although the analyses of circadian patterns focused only on the normal controls and patients with MDD because these are the two groups with sufficient TOD data. The reason to include both cases and controls in data processing was to conduct normalization and batch-effect correction (see below) using all samples in a batch, such that we maximized the accuracy when correcting for technical variations across batches.

Although most cohorts were analyzed on an Affymetrix U133A platform, several of the latest cohorts were analyzed on the newer Affymetrix U133Plus-v2 platform, which contains all U133A probe sets as a subset. We extracted the U133A subset of the data for these samples and combined it with data for those samples analyzed on the U133A platform. We applied robust multiarray analysis (RMA) (4, 5) to summarize probe set expression levels. RMA output in the form of logged (base 2) expression levels was generated using the custom ENTREZ12.1 chip definition files (CDFs) (6), which defined probe sets for 11,912 ENTREZ transcripts and 68 control probe sets. The reason for using our custom-defined CDFs rather than the probe annotation provided by Affymetrix was to remap all probes to the latest human genome build available and to annotate probes according to one of the most detailed gene models. The RMA results in this study thus represented 11,912 transcripts defined by ENTREZ in March 2010 and are covered by probes on the U133A microarrays. All downstream analyses were performed in R (7) using contributed packages available in early 2010.

Based on our prior experience in finding sex-specific transcripts in the human brain (8), we used 10 genes on the Y chromosome (*NLGN4Y* (*neuroligin 4, Y-linked*), *NCRNA00185* (*non-protein coding RNA 185*), *RPS4Y1* (*ribosomal protein S4, Y-linked 1*), *TTTY15* (*testis-specific transcript, Y-linked 15*), *UTY* (*ubiquitously*

transcribed tetratricopeptide repeat gene, Y-linked), *KDM5D* [*lysine (K)-specific demethylase 5D*], *USP9Y* (*ubiquitin specific peptidase 9, Y-linked*), *CYorf15B* (*chromosome Y open reading frame 15B*), *DDX3Y* [*DEAD* (*Asp-Glu-Ala-Asp*) *box polypeptide 3, Y-linked*], *EIF1AY* (*eukaryotic translation initiation factor 1A, Y-linked*)) and *XIST* [*x* (*inactive*)*-specific transcript*] on the X chromosome to infer sample sex. The analysis revealed that of >1,400 microarrays, only six NAcc samples involving two subjects had reciprocal sex switches in the database. These cases of sex misidentification were corrected. There is no evidence of sample mixing (i.e., inadvertently combining two samples) that involve a male-female sample pair, although our data could not rule out the unlikely possibility of same-sex mixing or switching.

To gain an overview of sample heterogeneity, we calculated sample-sample similarities for each region using pairwise Pearson's correlation coefficients (r) and calculated the average r of each sample compared with all other samples of the same region. We chose the threshold of average $r = 0.85$–0.94 (varying by region) to define and remove outlier microarrays. The outliers could result from either technical or biological differences. In one region (DLPFC), we removed additional microarrays corresponding to data produced at one laboratory for one cohort due to a low average r and poor match with the duplicate microarrays from the second laboratory. In all, we filtered out 52 (~3.5%) of 1,476 microarrays. The remaining 1,424 microarrays correspond to 776 unique RNA samples in six regions (Table S3).

Although the RMA method has normalized probe intensity distributions across microarrays, the resulting probe set summaries still showed between-cohort, between-microarray type (U133A vs. U133 Plus v2), and between-laboratory variations, thus requiring further normalization. For each brain region, we quantile-normalized (9) the probe set values and used pairwise correlation coefficients to define recognizable batches (10), which usually coincides with naturally occurring sample groups (>15 samples) according to cohorts or chip types. The 68 negative control probe sets on the microarray platform, representing spiked-in nonhuman transcripts, showed nearly identical batch effects as using all probes, indicating that most of the batch variation is due to technical differences in reagents and instruments rather than due to biological differences between samples in different batches. To adjust for batch effects, we median-centered the expression levels of each transcript within each batch and confirmed, using the correlation matrices, that the batch effects were removed after the adjustment.

We compared the result of this simple correction with the alternative Bayesian batch-correction approach implemented in *combat* (Combining Batches) (11), and we did not see meaningful differences in performance in terms of duplicate-sample concordance. Although this is contrary to the published comparison results showing that *combat* is a better algorithm for dealing with batch effects (11), its advantage is probably blunted in our dataset because (*i*) we have larger sample sizes per batch (typically >15) than what was tested in the published comparisons, and (*ii*) we used median centering rather than mean centering. The latter is susceptible to the influence of outlier values yet was used in earlier comparisons with *combat*. We note that *combat* has decreased the scale of variation for most transcripts (as a consequence of improving the group variance estimation) and resulted in underreporting of fold changes between sample groups. We therefore opted to maintain the use of the median-centering approach in this study.

After per-batch median centering, we quantile-normalized the resulting values and averaged the replicate microarrays for the same samples, yielding a dataset for unique subjects for each region. The RMA-normalized data and the final processed data are available (www.pritzkerneuropsych.org/?page_id=1196). Although the subjects were selected with no agonal complications (2) (Tables S2 and S4), there remains a moderate influence of expression patterns by brain pH, reflecting residual effects on gene expression due to medical conditions before death. To correct for this, we ran linear regression of expression levels against the first principal component (PC) 1 scores of the subjects, using the residuals for downstream analysis. In some brain regions (HC and NAcc), both PC1 and PC2 scores were associated with pH, and we ran linear regression against both.

Because the male and female subjects were not distributed evenly in their TOD around the 24-h day, our subsequent analysis was biased toward finding sex chromosome genes as showing circadian patterns. We therefore median-centered the male and female expression values as an additional step in data processing. This procedure primarily affected the small set of sex-specific transcripts (approximately eight chromosome Y transcripts and Xist on chromosome X). The regression with PC1 scores and sex correction each results in the reduction of only 1 df. Similarly, quantile normalization is a nonlinear rank-invariant transformation of the data. The procedures described above therefore represent relatively mild adjustments. Batch correction, on the other hand, represents a stronger adjustment, especially for batches of fewer samples.

TOD and Zeitgeber Time. To collect TOD data, deputy coroners first determined the span between time last seen alive and time found. Coroner deputies then collected a combination of data, including core temperature changes, neurological and cardiovascular changes (pupil dilation, clotted blood, pallor-pale/white, mucous membrane dryness, recent incontinence, tendon reflexes, clouded cornea, cadaveric spasm, dried blood, and tympanic abdomen resonant), rigor mortis onset (in jaw muscles, neck, fingers, wrists, elbows, shoulders, knees, and abdomen), and stages of lividity (e.g., blanches easily, blanches moderate pressure, blanches firm pressure, blanches fixed). Among recent deaths (<20 h), the combined use of these data reliably estimates TOD, accurate ±1 h in the first 6 h postmortem and ±1.5 h between 6 and 20 h postmortem (12–14). All control and MDD cases used in this study were found at less than 11 h postmortem, thus increasing the reliability in the determination of TOD. Estimates were reviewed by a board-certified forensic pathologist, compared with findings from internal forensic examination, and either confirmed or modified.

Not all subjects have documented TOD information. Only those with TOD data were included in the circadian analysis. It is known that the circadian phase in humans, as well as in other species, is synchronized to geophysical time mainly via photic cues perceived by the retina (15). Because the sunrise time varies by season and by latitude (Fig. S1), we adjusted the recorded TOD for each subject by the sunrise time of his or her date and place of death, and we used this zeitgeber time (ZT) scale for downstream analysis. In the adjusted scale, sunrise time is ZT = 0, noon is approximately ZT = 6, and midnight is approximately ZT = 18 (18 h after sunrise) or −6 (6 h before sunrise).

Discovery of Cyclic Genes. Our subjects show an uneven distribution of TOD (Fig. 1 *A* and *B*), precluding the use of standard methods intended for regular time series analysis (i.e., those involving constant intervals), including frequency domain analyses, such as the Fourier transformation. To discover cyclic genes, we fit the expression values of each gene by a sinusoidal function of time using the method of least squares, fixing the period at 24 h, and allowing the amplitude and phase to be free parameters:

$$Y_i = A\cos\frac{2\pi \times ZT_i}{24Hr} + B\sin\frac{2\pi \times ZT_i}{24Hr}. \qquad \textbf{[S1]}$$

In the expression above, Y_i is the expression level of the ith

subject, whose adjusted TOD is ZT_i, and A and B jointly determine the amplitude and phase, respectively, of the sinusoidal function. At the best-fitting parameters (A, B), we calculated the percentage of variance explained (PVE) by the fitted curve and evaluated its statistical significance by permutation (16). We randomly reassigned the ZT data across subjects 1,000 times and calculated the PVE for each round of permutation, thus obtaining a null distribution of 1,000 PVE values. The empirical P value of the actual PVE is obtained by comparing it with the null distribution. For example, when the actual PVE is larger than all but 1 of the 1,000 permutation PVEs, $P = 0.001$. When it is larger than all 1,000 permutation PVEs, the P value is between 0 and 0.001 but undetermined, and we used $P = 0.0005$ when we needed to average the logged P values across six regions.

It was necessary to fix the period at 24 h because the postmortem sampling times are limited to one 24-h cycle. Data such as ours are not amenable to discovering changes in the period of cyclic patterns. For example, if a transcript shows a lengthening of its period from 24 to 30 h in a study cohort, after sampling each subject only once and "folding" all data into the 0- to 24-h range, it is impossible to infer that the transcript has a longer than 24-h period. This limitation applies to all studies using independently sampled data.

Fisher's *P* Value, Phase, and Pathway Analysis. For each transcript, we combined the P values from six regions using Fisher's formula:

$$\chi^2 = (-2)\sum_{i=1}^{6}\log(p_i), \qquad \textbf{[S2]}$$

where p_i is the P value in region i, and the χ^2 statistic follows a χ^2 distribution with 12 df (2 * 6 brain regions), assuming independence (lack of consistency) across regions. This analysis is not intended as a formal test of overall significance or as a test for independence among regions but as a way to explore the degree of consistency among top genes. For technical reasons, some datasets (a certain region, in either controls or patients with MDD) may show systematic "inflation" or "deflation" of P values across the entire transcriptome. The P value inflation/deflation is likely to arise from technical differences affecting entire arrays that are unevenly distributed around the 24-h cycle, leading to apparently cyclic patterns affecting thousands of genes. This artifact is analogous to the phenomenon of population stratification in genetic association studies; thus, we adopted a correction method similar to the genomic control method (17). This method converts P values into a χ^2 statistic, finds the median of this statistic, and calculates the genomic control factor as the fold difference between the observed median and the expected median of a χ^2 distribution (df = 1). We then rescaled the χ^2 values by the genomic control factor and turned the corrected χ^2 values into the corrected P values. This was done for each region before calculating the Fisher's metaanalysis P value.

To identify phase, or peak time, we calculated the correlation coefficient of the actual data series for each gene with a family of 24 sinusoidal functions that are identical in shape but shifted by 1 h. The highest of the 24 correlation coefficients indicated the best-fitting curve in the family of 24 functions, thus providing the estimated peak time with a resolution of 1 h. All phase comparisons were conducted using circular statistics (in the "circular" package in R) to account for the artificial disconnect between ZT0/24 (also referred to as the "around the clock problem").

For functional analyses, we referred to "known circadian genes" as those documented by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (18) and Protein Information Resource (PIR) databases (19). These did not include those identified in previous transcriptome analyses of specific organisms and tissues.

Enrichment analysis relied on online tools at the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (20) and Pathway Analysis Using Logistic Regression (LRpath) (21), using gene symbols and ENTREZ identifications, respectively. The Web sites are http://lrpath.ncibi.org/main.jsp and http://david.abcc.ncifcrf.gov/summary.jsp, as implemented in October 2011.

Prediction. Prediction of TOD used the top 100 genes according to the Fisher's metaanalysis P value from the control dataset and relied on a PC analysis of the expression matrix for these genes in n training samples, where the samples were ordered by TOD (similar to Fig. 1*C* but with p number of genes). In this matrix, the expression values followed a wave-like form in both the x and y directions (i.e., over the different samples, ordered by TOD, and over the different genes, ordered by normal peak time). The first two eigenvectors (each a p-vector, across p genes) typically describe a sine function and a cosine function, respectively, for canonical peak-at-noon (maximum at ZT ~ 6) and peak-at-morning (maximum at ZT = 0) patterns, respectively. The relative loading of the two eigenvectors in each sample, as described by the sample's first two eigenvalues, A and B, reflects the sample's peak time and is used to calculate the predicted TOD. Specifically, when the ith sample's expression vector, $\boldsymbol{Y_i}$, can be expressed approximately as the sum of two components (similar to Eq. **S1**):

$$\boldsymbol{Y_i} \approx A_i * \boldsymbol{PC_1} + B_i * \boldsymbol{PC_2}, \qquad \textbf{[S3]}$$

where $\boldsymbol{PC_1}$ and $\boldsymbol{PC_2}$ are the first two eigenvectors. After they are verified as describing the sine and cosine functions, respectively, the quantity A_i/B_i (i.e., the ratio of the first two PC scores for each sample) is turned into the "angle" in the 24-h polar coordinate using the "arctangent" function:

$$Predicted\ \ TOD\ \ (hour) = \begin{cases} \tan^{-1}\left(A_i/B_i\right) * \dfrac{24}{(2*Pi)}, & B > 0 \\ \tan^{-1}\left(A_i/B_i\right) * \dfrac{24}{(2*Pi)} + 12, & B < 0 \end{cases}. \qquad \textbf{[S4]}$$

To avoid the potential bias in using the control patterns to predict TOD for patients with MDD, we combined the MDD cases and the controls in an undistinguished pool. We then repeatedly sampled 60 subjects from the pool, deriving PCs from this training set and using PC1 and PC2 (the first two eigenvectors) to predict TOD for the remaining MDD cases and controls, which form the test set (n = 20–29, depending on brain region). After 50 iterations, we averaged the predicted TOD for each sample as it appeared in the training set. We ran this analysis for the four regions with larger sample sizes (n = 85 for DLPFC, n = 89 for AnCg, n = 80 for HC and NAcc) and averaged TOD across the four regions to obtain the final predicted TOD for each subject.

SI Summaries and Discussions

Mammalian Circadian Molecular Machinery. To aid the understanding of our study by general readers, we provide a brief overview of the current knowledge of circadian clock machinery. The circadian clock represents an evolutionary conserved regulatory process controlling the rhythmic expression of genes involved in a wide array of physiological and behavioral activities, including the sleep/wake cycle, body temperature, hormonal secretion, and behavior. At the intracellular level, rhythmicity is generated by interlocking transcriptional and translational feedback loops involving a set of "core clock genes" that are conserved in most animals. We expand on the transcriptional regulation of these core clock genes

below, because 11 of these genes were discovered to be among the most rhythmic transcripts in the human brain (Fig. S2).

The core clock gene loop that generates 24-h periodicity centers around rhythmic transcriptional regulation at the E-box DNA binding site. This loop includes three Period genes (*PER1*, *PER2*, and *PER3*); two cryptochrome genes (*CRY1* and *CRY2*); *CLOCK* (or its homolog neuronal PAS domain-containing protein 2 or *NPAS2*), two aryl hydrocarbon receptor nuclear translocator-like genes (*ARNTL* and *ARNTL2*, also referred to as *BMAL1* and *BMAL2*), and two basic helix-loop-helix family genes, members e40 and *e41* (*BHLHE40*/*41*, also referred to as *DEC1* and *DEC2*). Within the positive limb of this feedback loop, CLOCK and ARNTL proteins form heterodimers that bind to E-box sequences to drive the transcription of *PER*, *CRY*, and *BHLHE40* and *BHLHE41* mRNAs. Negative feedback occurs when the PER and CRY proteins accumulate and dimerize in the cytoplasm and then translocate to the nucleus, where they bind to *CLOCK*/*ARNTL* to inhibit their own transcription. Rhythmicity is generated as a consequence of this feedback loop, which has an inherent tempo governed by the delayed activation and repression around the loop as determined by posttranscriptional/posttranslational modification (22, 23).

This primary loop is accompanied by two secondary loops that center around rhythmic transcriptional regulation at two other DNA binding sites, the D-box and REV-ERB/retinoid-related orphan receptor (ROR) response element (RRE). These secondary loops serve to stabilize and amplify the primary loop. One loop involves a set of transcriptional activators for the D-box: D-site of albumin promoter binding protein (*DBP*), thyrotroph embryonic factor (*TEF*), and hepatic leukemia factor (*HLF*). The other loop includes a set of transcriptional activators and repressors for the RRE, the *ROR* genes (*RORα*/*β*/*γ*), and nuclear receptor subfamily 1, group D genes (*NR1D1* and *NR1D2*, also referred to as *REV-ERBα*/*β*). In general, the transcription of these gene families is driven by ARNTL/CLOCK via E-box sequences. The D-box activators then further drive *PER* transcription, as well as the transcription of *RORα*/*β*/*γ* and *NR1D1*/*2*. Finally, activation of the RRE feeds back to drive the transcription of *ARNTL*/*CLOCK*, as well as the transcription of the nuclear factor interleukin-3–regulated gene (*NFIL3*, also known as *E4BP4*), which encodes a transcriptional repressor that binds at the D-box sequence (24) and may further regulate PER and CRY proteins (25). This regulatory system, along with epigenetic processes, controls the expression of multiple downstream, or "clock-controlled," genes (26).

Pathway Analysis. To detect biological themes represented by the cyclic genes systematically, we ran an enrichment analysis of the top 600 cyclic genes using Database for Annotation, Visualization, and Integrated Discovery online tools (DAVID) (20). These genes were selected based on their mean log-*P* value across six regions. Different annotation systems of gene function showed remarkably consistent results:

- The top-scoring Uni-Prot and PIR keyword is "biological rhythms," which includes 8 genes (*HLF*, *NPAS2*, *CRY2*, *DBP*, *PER2*, *PER1*, *PER3*, and *NFIL3*) among the top 600, with an enrichment *P* value of 1.35E-6.
- The top-scoring KEGG pathway is the hsa04710:Circadian rhythm pathway, with 9 annotated genes (*NPAS2*, *CRY2*, *NR1D1*, *PER2, PER1*, *BHLHE40*, *ARNTL*, *PER3*, and *BHLHE41*) in the top 600, with an enrichment *P* value of 6.1E-9).
- The top-scoring Gene Ontology term is GO:0048511:rhythmic process, with 15 genes (*HLF*, *FGF7*, *ARNTL*, *CCNE1*, *NPAS2*, *CRY2*, *NR1D1*, *DBP*, *PER2*, *PER1*, *NOS3*, *ADAMTS1*, *PER3*, *NFIL3*, and *FSHB*), with an enrichment *P* value of 1.56E-5.

We also adopted a logistic regression-based analysis method, LRpath (21), that does not require an arbitrary cutoff of top genes. We used *P* values for all >11,000 genes to screen pre-annotated gene sets in the Biocarta, Gene Ontology, KEGG, and Panther pathways, and we consistently found circadian rhythm to be the top "concept" (Table S1). Several other gene sets potentially related to circadian transcription, such as "PAS fold," "basic region leucine zipper," "sequence-specific DNA binding transcription factor activity," and "helix–loop–helix DNA-binding domain," also had an enrichment false discovery rate of less than 0.05.

Correlation of Statistical Significance Across Regions. The *P* values for top genes were correlated across regions (Fig. S4*A*), and the average ranks of *P* values for the most significant genes were smaller than the average ranks of the top genes in order-permutated datasets (i.e., those without region-region correlation) (Fig. S4*B*). We estimated that there were >100 genes showing consistent cyclic patterns across regions: If we removed the 100 genes with the lowest median rank (lower rank = smaller *P* value) across regions, the remaining genes would have average ranks much more similar to those in permuted datasets (Fig. S4*C*).

Similarity of Peak Time (i.e., Phase) Across Regions. The circular variance of peak times across six regions had a median of 0.089 h over 445 genes that were cyclic ($P < 0.05$) in at least two brain regions, indicating that circadian rhythms are relatively synchronized among the six extra-SCN regions analyzed. As expected, the smaller the mean *P* value (i.e., more robust and consistent cyclic patterns), the smaller were the circular variance of peak times, resulting in a median of 0.052 h for the top 50 most significant genes (Fig. S6*A*). There was no evidence of systematic phase shift between any pairs of regions, as determined by pairwise comparisons of peak times across top cyclic genes at various *P* value cutoffs.

There is a general trend that genes with higher amplitude are more likely to show smaller *P* values (Fig. S6*B*). This can be explained by the fact that genes with smaller circadian amplitude would be less likely to rise above noise and be detected.

Comparison with Results from Animal Models. We compared our results for human non-SCN regions with those from previous animal studies regarding the significance and phasing of circadian genes. Yan et al. (27) performed a metaanalysis that included gene expression data from 14 mouse tissues and reported 41 "core circadian genes," of which 27 were on the microarray platform used in our study and were rhythmic in the mouse brain outside the SCN (prefrontal cortex or whole brain).

Fig. S6 shows the comparisons of *P* values for the ~5,730 genes shared between the two studies when all the 14 tissues of the study by Yan et al. (27) are included. The 9 most significant genes in our study are highlighted in red (Fig. S7). Six of the 9 genes had a *P* value <0.01 in the mouse study, suggesting that the greatest level of concordance between the human and mouse data was found in canonical clock genes.

Because humans are a diurnal (day-active) species and most traditional laboratory rodents are nocturnal, our data provide an opportunity to compare the phase of circadian patterns in species with different chronotypes. When we compared the peak times for genes reported as rhythmic for the mouse prefrontal cortex or whole brain in the study by Yan et al. (27) and had $P < 0.01$ in our study (Fig. 3*D*), the seven top genes showed a linear relationship between the human and mouse data, but with a shift such that the phase in the mouse is delayed by ~6.5 h relative to the human. When fit with robust linear modeling (using rlm in R), they revealed a shift of 6.51 h and a slope of 1.18 ($r = 0.88$; circular correlation coefficient = 0.61). For example, NR1D1 peaks at ZT = 2 (2 h after sunrise) in our data for the human brain and

peaks at ZT = 9.8 in mouse prefrontal cortex tissues [in the study by Yan et al. (27), ZT = 0 is 7:00 AM, similar to our definition of ZT = 0 as sunrise time]. In addition, transcripts for the Period genes peaked during the day in our data (Fig. 2*B*), as has been previously reported in the human cingulate cortex (28) and in similar cortical and limbic brain regions in other diurnal species, such as *Spermophilus tridecemlineatus* (29) and *Octodon degus* (30).

Staggered Phase Pattern in Three Period Genes. In our dataset, the peak times of the three Period genes were staggered, with *Per1* peaking soon after sunrise, *Per3* peaking during midday, and *Per2* peaking in the afternoon (Fig. 2*B*). A similar staggered phase relationship is highly characteristic of Period gene expression in the SCN of laboratory rodents (Fig. S3) [e.g., mice (31), *Arvicanthis ansorgei* (32), *O. degus* (30)], but it has not been observed outside of the SCN, perhaps because the detection of such a relationship requires densely spaced sampling points around the 24-h day instead of data collected in two (e.g., ref. 27) or four (e.g., refs. 33–35) binned time points.

Effect of Sample Size in Comparison of Controls and MDD Cases. Because there were more control subjects than MDD cases in our sample collection, we asked whether the larger sample size and specific TOD distribution in controls could partially explain the much weaker evidence of circadian pattern in MDD cases. To answer this question, we selected a subset of controls so that (*i*) we had an equal number of controls as patients with MDD for each brain region and (*ii*) the TODs were matched as closely as possible between the patients with MDD and the selected controls. The sample sizes for the selected controls are as follows: 34, 33, 14, 13, 32, and 29 in the AnCg, DLPFC, CB, AMY, HC, and NAcc, respectively. The AMY and CB were not analyzed further due to small sample sizes. The *P* values for the other four regions remained much more significant in the matched subset of controls than in the patients with MDD (Fig. S8 *D* and *E*), confirming that the circadian patterns for top cyclic genes defined in controls were much weaker, if present at all, in MDD cases.

Sample-Sample Correlations Suggests Phase Shift in MDD Cases. We used the top cyclic genes ($n = 108$) to calculate sample-sample correlation in the DLPFC (Pearson's r). There was a clear positive correlation among control samples with similar TODs and a negative correlation among those with opposing TODs (Fig. S9*A*). This pattern was much weaker among MDD cases (Fig. S9*B*) or between cases and controls (Fig. S9*C*). The median absolute r value was 0.185 among controls, and it was lower among MDD cases ($r = 0.140$) and between cases and controls ($r = 0.138$). The maintenance of positively and negatively correlated samples in the MDD group despite the loss of a predictable pattern of correlation based on TOD suggests that the individual patients with MDD may continue to express a residual pattern of circadian gene expression but are desynchronized from the solar day (i.e., loss of normal circadian entrainment).

1. Evans SJ, et al. (2003) DNA microarray analysis of functionally discrete human brain regions reveals divergent transcriptional profiles. *Neurobiol Dis* 14(2):240–250.
2. Li JZ, et al. (2004) Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum Mol Genet* 13(6):609–616.
3. Tomita H, et al. (2004) Effect of agonal and postmortem factors on gene expression profile: Quality control in microarray analyses of postmortem human brain. *Biol Psychiatry* 55(4):346–352.
4. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4(2):249–264.
5. Irizarry RA, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31(4):e15.
6. Dai M, et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* 33(20):e175.
7. R Development Core Team (2005) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna).
8. Vawter MP, et al. (2004) Gender-specific gene expression in post-mortem human brain: Localization to sex chromosomes. *Neuropsychopharmacology* 29(2):373–384.
9. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2):185–193.
10. Li JZ, et al. (2007) Sample matching by inferred agonal stress in gene expression analyses of the brain. *BMC Genomics* 8:336.
11. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1):118–127.
12. Spitz WU, Spitz DJ, eds (2005) *Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation* (Thomas, Springfield, IL), 4th Ed.
13. Henssge C, et al. (2000) Experiences with a compound method for estimating the time since death. II. Integration of non-temperature-based methods. *Int J Legal Med* 113(6):320–331.
14. Kaliszan M, Hauser R, Kernbach-Wighton G (2009) Estimation of the time of death based on the assessment of post mortem processes with emphasis on body cooling. *Leg Med (Tokyo)* 11(3):111–117.
15. Roenneberg T, Kumar CJ, Merrow M (2007) The human circadian clock entrains to sun time. *Curr Biol* 17(2):R44–R45.
16. Shedden K, Cooper S (2002) Analysis of cell-cycle-specific gene expression in human cells as determined by microarrays and double-thymidine block synchronization. *Proc Natl Acad Sci USA* 99(7):4379–4384.
17. Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55(4): 997–1004.
18. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38(Database issue):D355–D360.
19. Wu CH, et al. (2003) The Protein Information Resource. *Nucleic Acids Res* 31(1): 345–347.
20. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
21. Sartor MA, Leikauf GD, Medvedovic M (2009) LRpath: A logistic regression approach for identifying enriched biological groups in gene expression data. *Bioinformatics* 25(2):211–217.
22. Bellet MM, Sassone-Corsi P (2010) Mammalian circadian clock and metabolism—The epigenetic link. *J Cell Sci* 123(Pt 22):3837–3848.
23. Masri S, Sassone-Corsi P (2010) Plasticity and specificity of the circadian epigenome. *Nat Neurosci* 13(11):1324–1329.
24. Ukai-Tadenuma M, Kasukawa T, Ueda HR (2008) Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks. *Nat Cell Biol* 10(10):1154–1163.
25. Ohno T, Onishi Y, Ishida N (2007) The negative transcription factor E4BP4 is associated with circadian clock protein PERIOD2. *Biochem Biophys Res Commun* 354(4): 1010–1015.
26. Zhang EE, Kay SA (2010) Clocks not winding down: Unravelling circadian networks. *Nat Rev Mol Cell Biol* 11(11):764–776.
27. Yan J, Wang H, Liu Y, Shao C (2008) Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLOS Comput Biol* 4(10):e1000193.
28. Cermakian N, Lamont EW, Boudreau P, Boivin DB (2011) Circadian clock gene expression in brain regions of Alzheimer's disease patients and control subjects. *J Biol Rhythms* 26(2):160–170.
29. Mrosovsky N, Edelstein K, Hastings MH, Maywood ES (2001) Cycle of period gene expression in a diurnal mammal (Spermophilus tridecemlineatus): Implications for nonphotic phase shifting. *J Biol Rhythms* 16(5):471–478.
30. Vosko AM, Hagenauer MH, Hummer DL, Lee TM (2009) Period gene expression in the diurnal degu (Octodon degus) differs from the nocturnal laboratory rat (Rattus norvegicus). *Am J Physiol Regul Integr Comp Physiol* 296(2):R353–R361.
31. Takumi T, et al. (1998) A light-independent oscillatory gene mPer3 in mouse SCN and OVLT. *EMBO J* 17(16):4753–4759.
32. Caldelas I, Poirel VJ, Sicard B, Pévet P, Challet E (2003) Circadian profile and photic regulation of clock genes in the suprachiasmatic nucleus of a diurnal mammal Arvicanthis ansorgei. *Neuroscience* 116(2):583–591.
33. Ackermann K, Dehghani F, Bux R, Kauert G, Stehle JH (2007) Day-night expression patterns of clock genes in the human pineal gland. *J Pineal Res* 43(2):185–194.
34. Wu YH, et al. (2006) Pineal clock gene oscillation is disturbed in Alzheimer's disease, due to functional disconnection from the "master clock" *FASEB J* 20(11):1874–1876.
35. Yang S, Wang K, Valladares O, Hannenhalli S, Bucan M (2007) Genome-wide expression profiling and bioinformatics analysis of diurnally regulated genes in the mouse prefrontal cortex. *Genome Biol* 8(11):R247.

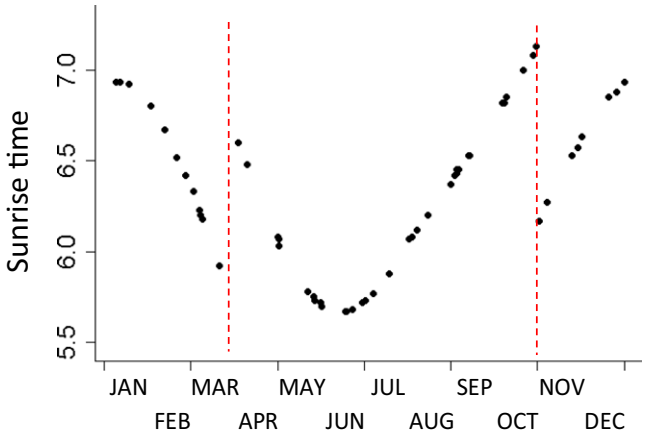


Fig. S1. Seasonal variation of sunrise time. Shown is the distribution of sunrise time vs. calendar month for the 55 subjects included in our study. Data were obtained from the online US Naval Observatory Data Services (http://aa.usno.navy.mil/data). The two red lines indicate the dates for switching to and away from Daylight Savings Time in the year of the patient's death.

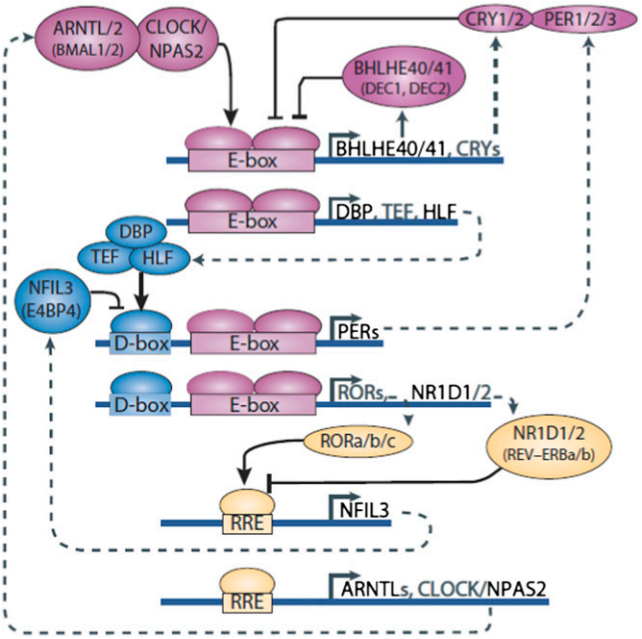


Fig. S2. Simplified diagram of the interlocking transcriptional feedback loops underlying the mammalian circadian clock, highlighting transcripts that are strongly rhythmic in the human brain according to the current study. This diagram is derived from models presented by Ukai-Tadenuma et al. (1) and Zhang and Kay (2). Squares represent the three primary types of DNA binding sites involved in the transcriptional regulation of clock genes, and they are color-coded as follows: E-box (pink), D-box (blue), and RRE (yellow). Ovals represent proteins that bind to these sites, and they were color-coded to match the three types of binding sites. Transcript names are shown as text to the right of the occupied binding sites. Most transcripts in the diagram, shown in bold typeface, rank among the top 50 rhythmic transcripts in the human brain according to our data. Black lines indicate transcriptional regulation, with arrow tips representing activation and flat lines representing repression.

1. Ukai-Tadenuma M, Kasukawa T, Ueda HR (2008) Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks. *Nat Cell Biol* 10(10):1154–1163.
2. Zhang EE, Kay SA (2010) Clocks not winding down: Unravelling circadian networks. *Nat Rev Mol Cell Biol* 11(11):764–776.

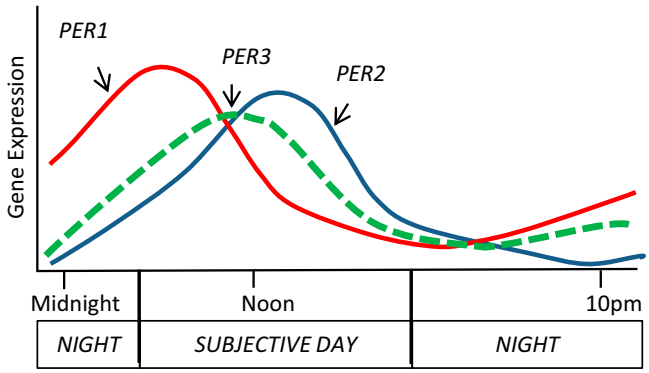


Fig. S3. Relative phasing of *PER1–3* expression in prior studies of the rodent SCN resembles that seen in the human brain. Rodent data from the SCN shows a "staggered" phase relationship between the Period genes (adapted from ref. 1), resembling the human data illustrated in Fig. 2*B*. Time is presented in ZT, with ZT0 equivalent to the time of sunrise.

1. Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* 96(2):271–290.

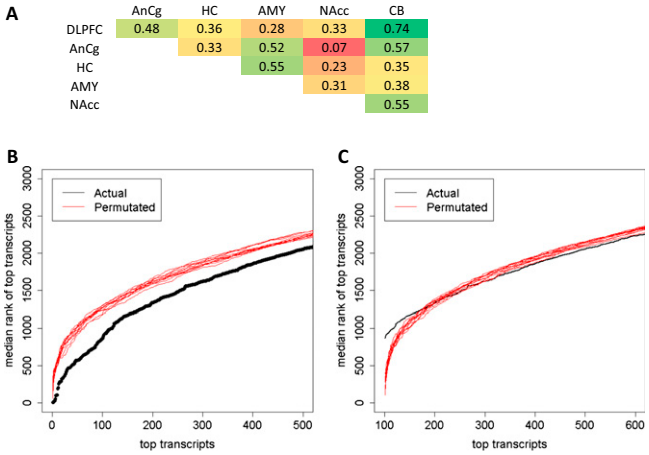


Fig. S4. Statistical significance correlates across brain regions in controls. (*A*) Spearman's rank correlation coefficients of log-*P* values between pairs of regions using 27 transcripts with the lowest median *P* values (<0.03) in the AnCg, DLPFC, HC, and NAcc, the four regions with the largest sample sizes. (*B*) Median rank of the 500 most significant transcripts in the six regions for the actual data (black) and for 10 random permutations of *P* values (red). For each permutation, observed *P* values for each region were randomly reassigned across all transcripts, and the median rank across six regions was calculated for each gene and sorted, with the 500 highest ranked transcripts plotted as a red line. (*C*) Median rank of the next 500 most significant transcripts, after removing the 100 top genes, for the actual data (black) and for 10 random permutations of *P* values (red), showing that the median ranks for the 101th to the 600th genes are similar to those in random data.

16
10
5
0
-5

Symbol	Name	DLPFC	AnCg	HC	AMY	NAcc	CB
ARNTL*	ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR-LIKE PROTEIN	-5	-4	-5	-6	-4	-5
PER2*	PERIOD, DROSOPHILA, HOMOLOG OF, 2	9	8	7	8	12	6
PER3*	PERIOD, DROSOPHILA, HOMOLOG OF, 3	6	5	6	6	5	3
NR1D1*	NUCLEAR RECEPTOR SUBFAMILY 1, GROUP D, MEMBER 1	2	1	3		2	0
DBP*	D SITE OF ALBUMIN PROMOTER-BINDING PROTEIN	5	5	6	6	5	2
SFPQ	SPLICING FACTOR, PROLINE- AND GLUTAMINE-RICH	15		14		15	12
ITIH5	INTER-ALPHA-TRYPSIN INHIBITOR, HEAVY CHAIN 5	12	13	10		14	15
LDLR	LOW DENSITY LIPOPROTEIN RECEPTOR	2	1	0	0		
PER1*	PERIOD, DROSOPHILA HOMOLOG OF, 1	3	1	1		2	3
INSIG1	INSULIN-INDUCED GENE 1	2	3	1	-2		
SLC39A14	SOLUTE CARRIER FAMILY 39 (ZINC TRANSPORTER), MEMBER 14	0	1	3	3	2	
NFIL3*	NUCLEAR FACTOR, INTERLEUKIN 3-REGULATED	-5	-4			-2	
SNTB2	SYNTROPHIN, BETA-2	5	1		8	2	
PDZRN3	PDZ DOMAIN-CONTAINING RING FINGER PROTEIN 3	3	2	5			
BHLHE40*	BASIC HELIX-LOOP-HELIX FAMILY, MEMBER E40	7	2			2	
BHLHE41	BASIC HELIX-LOOP-HELIX FAMILY, MEMBER E41	4	5			4	3
HLF	HEPATIC LEUKEMIA FACTOR		6		10		6
ETV5	ets variant gene 5	12	14	16	11	15	
TNIP2	TNFAIP3-INTERACTING PROTEIN 2		3	2	0		
ESYT1	extended synaptotagmin-like protein 1	6		3	3	3	4
ZNF394	zinc finger protein 394	6	1	3			
PION	pigeon homolog (Drosophila)	10	10			10	
GPR6	G PROTEIN-COUPLED RECEPTOR 6	2	2				6
TIMM8A	translocase of inner mitochondrial membrane 8 homolog A (yeast)		-3	11	-4		
GPR116	G protein-coupled receptor 116	12		9			12
FLRT1	fibronectin leucine rich transmembrane protein 1		13	15		11	
CSGALNACT1	chondroitin sulfate N-acetylgalactosaminyltransferase 1	16		14	13	16	
WDR41	WD repeat domain 41			10		10	
APOLD1	APOLIPOPROTEIN L DOMAIN-CONTAINING 1	3	1	0		1	
RHOB	RAS HOMOLOG GENE FAMILY, MEMBER B	-3	-1	-2		-1	
SCML1	SEX COMB ON MIDLEG, DROSOPHILA, HOMOLOG-LIKE 1	0	1	3		0	
SPRY4	SPROUTY, DROSOPHILA, HOMOLOG OF, 4	12	14	14		17	
MTR	5-METHYLTETRAHYDROFOLATE-HOMOCYSTEINE S-METHYLTRANSFERASE	10	6	9			
PLSCR1	PHOSPHOLIPID SCRAMBLASE 1		0	1		0	
EXOC1	EXOCYST COMPLEX COMPONENT 1	16	12	17		12	
KLF11	KRUPPEL-LIKE FACTOR 11	-2	0	0		-1	
SLCO4A1	SOLUTE CARRIER ORGANIC ANION TRANSPORTER FAMILY, MEMBER 4A1		-1	1	1	0	
SOCS2	SUPPRESSOR OF CYTOKINE SIGNALING 2	1	-1	1			
C10orf116	adipogenesis regulatory factor	10		12			9
ZNF286A	zinc finger protein 286A		-5	11		7	10
GAS2	GROWTH ARREST-SPECIFIC 2	-1			-2	-2	
UNC13A	UNC13, C. ELEGANS, HOMOLOG OF, A	11				11	
ATP4A	ATPase, H+,K+ EXCHANGING, ALPHA SUBUNIT		-5		10		6
RFC3	REPLICATION FACTOR C, SUBUNIT 3		2	8			12
ACOT13	acyl-CoA thioesterase 13	16	-5	15			
C7orf68	HILPDA hypoxia inducible lipid droplet-associated		1	1		3	
SYNM	DESMUSLIN	4	4		3		8
HCRTR2	HYPOCRETIN RECEPTOR 2			11			10
ZW10	ZESTE-WHITE 10	1					2
NPAS2	NEURONAL PAS DOMAIN PROTEIN 2					17	-5

Fig. S5. Phase of top cyclic genes is consistent across brain regions in controls. Shown are the times of peak expression for the top 50 genes, with the genes ordered by the average logged *P* value across six regions. Transcripts that were not significant ($P > 0.1$) in a given region are shown as blank. Phase is color-coded, such that genes that peak in expression earliest in the early morning (−5, or 5 h before sunrise) are red and those peaking latest in the evening (1) are green. Note that because our scale is linear but time itself is circular, a gene [e.g., *ACOT13* (*acyl-CoA thioesterase 13*)] may peak right before midnight in one region and right after midnight in another region, creating the artificial impression of large phase variation even though the actual peak times are only a few hours apart.

1. Shedden K, Cooper S (2002) Analysis of cell-cycle-specific gene expression in human cells as determined by microarrays and double-thymidine block synchronization. *Proc Natl Acad Sci USA* 99(7):4379–4384.

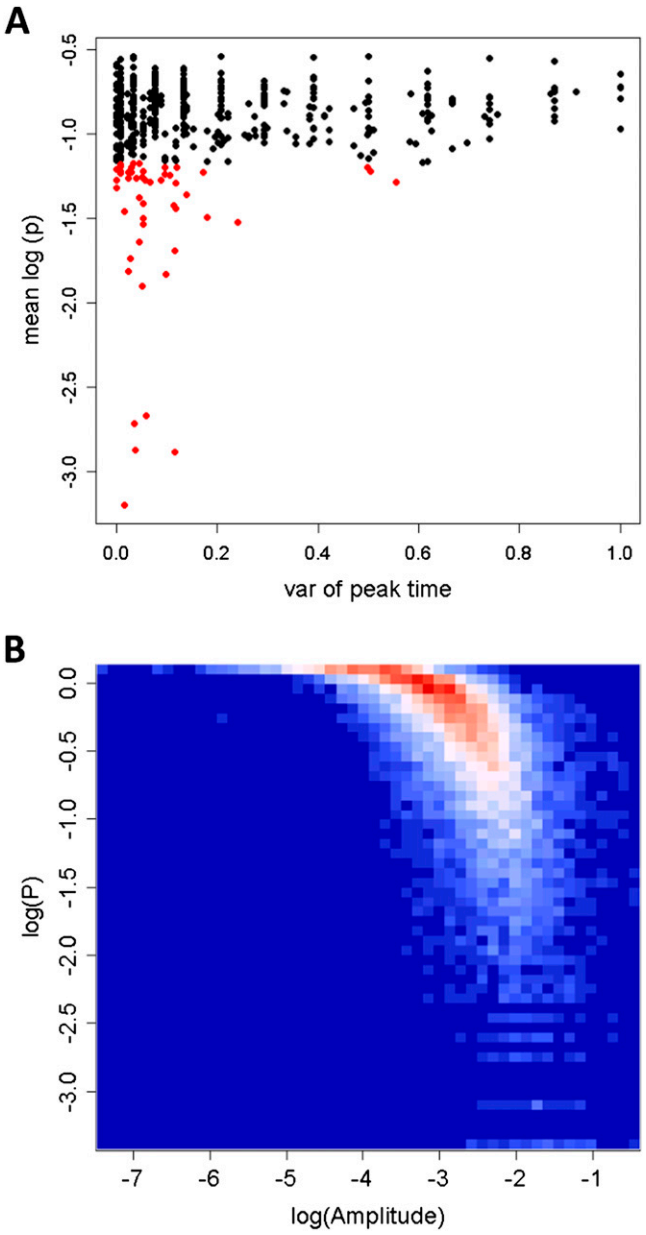


Fig. S6. Most significantly circadian genes in controls have the highest concordance of phase across brain regions, and the greatest rhythm amplitude. (*A*) Circular variance (var) of peak time (*x* axis) vs. mean log *P* value (*y* axis) across six regions for the 445 genes with $P < 0.05$ in at least two regions. The same 445 genes were shown in Fig. 3 *C* and *D*. The top 50 genes shown in Fig. 3 were colored in red and exhibited smaller circular variance of peak time than other genes. (*B*) 2D density heat map of log(Amplitude) (*x* axis) and log(*P*) (*y* axis) for all 11,979 transcripts in the DLPFC. The higher the amplitude, the more likely it was that the transcript showed a significant *P* value. Colors correspond to the number of genes concentrated in any particular part of the diagram [high density (red), low density (blue)].

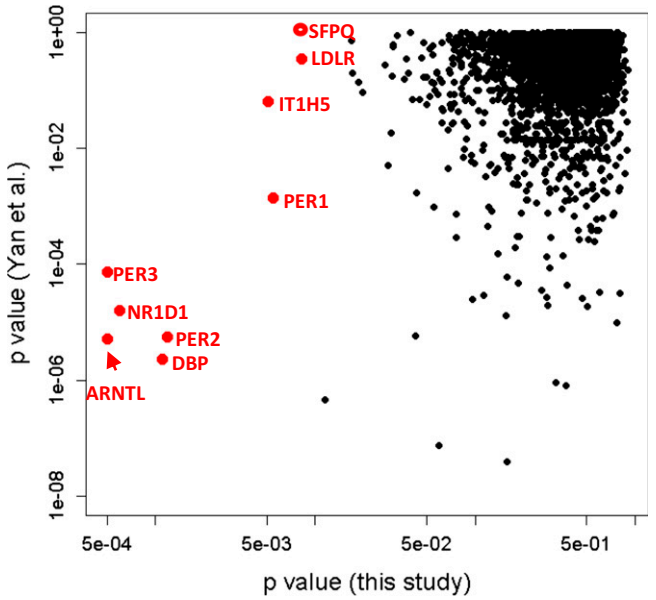


Fig. S7. Top circadian genes overlap between mouse and human brain tissues. *P* values and mean peak times for mice were obtained from supplementary table 2 of ref. 1. Comparisons of *P* values for ~5,730 genes that overlapped between our study and that of Yan et al. (1) were made, with the 9 most significant genes in our study highlighted in red, 6 of which had $P < 0.01$ in the mouse study. Six other genes had $P < 1E\text{-}6$ in the study by Yan et al. (1). Of these, 5 were shown in the plot; the sixth had $P < 1e\text{-}8$ and was out of the displayed range; however, they were not significant in our study. Note that the *P* values for mice were based on consistency across tissues using the circular range test rather than Fisher's method of metaanalysis. Alternative names for *BMAL1* and *NR1D1* are *ARNTL* and *REV-ERB*α, respectively.

1. Yan J, Wang H, Liu Y, Shao C (2008) Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLOS Comput Biol* 4(10):e1000193.

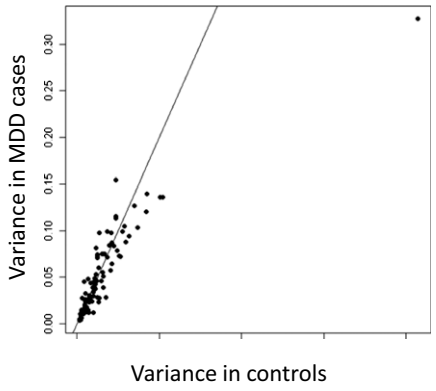
A Controls

	DLPFC	AnCG	NACC	HC	CB	AMY
ARNTL	0.41	0.36	0.37	0.23	0.47	0.28
PER2	0.43	0.30	0.25	0.25	0.53	0.27
PER3	0.49	0.45	0.43	0.53	0.51	0.27
NR1D1	0.47	0.32	0.40	0.25	0.48	0.24
DBP	0.39	0.34	0.39	0.31	0.48	0.18
SFPQ	0.28	0.09	0.28	0.25	0.37	0.15
ITIH5	0.34	0.22	0.26	0.25	0.19	0.18
LDLR	0.22	0.21	0.05	0.21	0.21	0.21
PER1	0.36	0.18	0.16	0.21	0.23	0.10
INSIG1	0.31	0.26	0.07	0.29	0.25	0.36
SLC39A14	0.26	0.38	0.26	0.44	0.18	0.42
NFIL3*	0.46	0.51	0.37	0.38	0.38	0.46
SNTB2	0.23	0.22	0.19	0.19	0.11	0.36
PDZRN3	0.40	0.39	0.10	0.38	0.11	0.21
BHLHE40	0.40	0.24	0.23	0.25	0.22	0.17
BHLHE41	0.42	0.38	0.25	0.01	0.41	0.12

B MDD patients

	DLPFC	AnCG	NACC	HC	CB	AMY
ARNTL	0.18	0.16	0.26	0.10	0.25	0.24
PER2	0.47	0.24	0.27	0.20	0.25	0.26
PER3	0.22	0.13	0.31	0.29	0.20	0.02
NR1D1	0.36	0.27	0.42	0.13	0.36	0.35
DBP	0.27	0.17	0.34	0.32	0.59	0.12
SFPQ	0.25	0.19	0.30	0.22	0.40	0.26
ITIH5	0.06	0.16	0.37	0.38	0.06	0.22
LDLR	0.24	0.09	0.16	0.11	0.52	0.21
PER1	0.30	0.17	0.14	0.18	0.35	0.11
INSIG1	0.32	0.13	0.14	0.08	0.11	0.16
SLC39A14	0.15	0.24	0.43	0.41	0.15	0.15
NFIL3*	0.18	0.20	0.35	0.17	0.23	0.23
SNTB2	0.05	0.16	0.13	0.33	0.33	0.18
PDZRN3	0.42	0.56	0.26	0.38	0.19	0.44
BHLHE40	0.24	0.07	0.28	0.27	0.41	0.11
BHLHE41	0.22	0.12	0.09	0.18	0.30	0.38

C



D Matched Controls

	DLPFC	AnCG	NACC	HC
ARNTL	0.0005	0.001	0	0.007
PER2	0.005	0.006	0.064	0.181
PER3	0.003	0.009	0.056	0.217
NR1D1	0.0005	0.003	0	0.011
DBP	0.003	0.023	0.055	0.016
SFPQ	0.018	0.351	0.023	0.037
ITIH5	0.002	0.016	0.044	0.154
LDLR	0.032	0.026	0.073	0
PER1	0.001	0.006	0.075	0.002
INSIG1	0.009	0.041	0.841	0.002
SLC39A14	0.036	0.017	0.353	0.162
NFIL3*	0.072	0.049	0.146	0.077
SNTB2	0.688	0.069	0.111	0.222
PDZRN3	0.127	0.055	0.32	0.18
BHLHE40	0.003	0.057	0.138	0.152
BHLHE41	0.056	0.209	0.537	0.902

E MDD patients

	DLPFC	AnCG	NACC	HC
ARNTL	0.121	0.072	0.005	0.732
PER2	0.015	0.083	0.075	0.213
PER3	0.42	0.652	0.29	0.432
NR1D1	0.04	0.029	0.003	0.285
DBP	0.102	0.236	0.121	0.025
SFPQ	0.135	0.124	0.047	0.165
ITIH5	0.936	0.47	0.15	0.117
LDLR	0.012	0.385	0.005	0.315
PER1	0.006	0.21	0.124	0.137
INSIG1	0.056	0.534	0.318	0.668
SLC39A14	0.641	0.21	0.157	0.301
NFIL3*	0.565	0.326	0.179	0.633
SNTB2	0.928	0.194	0.365	0.123
PDZRN3	0.13	0.003	0.075	0.503
BHLHE40	0.19	0.897	0.14	0.433
BHLHE41	0.497	0.781	0.875	0.754

Fig. S8. Decreased circadian rhythmicity in the MDD group cannot be explained by smaller sample size or lack of overall variation in expression levels. (*A* and *B*) Overall, MDD cases have lower circadian amplitude than controls. Illustrated is the amplitude for the top 16 genes (previously shown in Fig. 2*A*) in controls (*A*) and in MDD cases (*B*), with higher amplitudes coded in red and lower amplitudes in blue. Cycle amplitude was defined as the range between the maximum and minimum of the best-fitting sinusoidal curve on a $\log_2$ scale. (*C*) MDD cases and controls exhibit a similar amount of variation in gene expression levels. Shown is a comparison of the variance for the top 100 genes between controls (*x* axis) and MDD cases (*y* axis). The black straight line has a slope of 1. The gene with the largest variance (*Upper Right*) is apolipoprotein L domain containing 1 (*APOLD1*). (*D* and *E*) Small sample size cannot explain the lack of significant rhythmicity in the MDD group. Shown are the *P* values of the top 16 genes as in Fig. 2*A*, except that the controls have been selected to have an equal sample size as cases and have matched distributions of TOD. Only four regions were analyzed, because the AMY and CB contain fewer cases (n = 13 and n = 14, respectively).

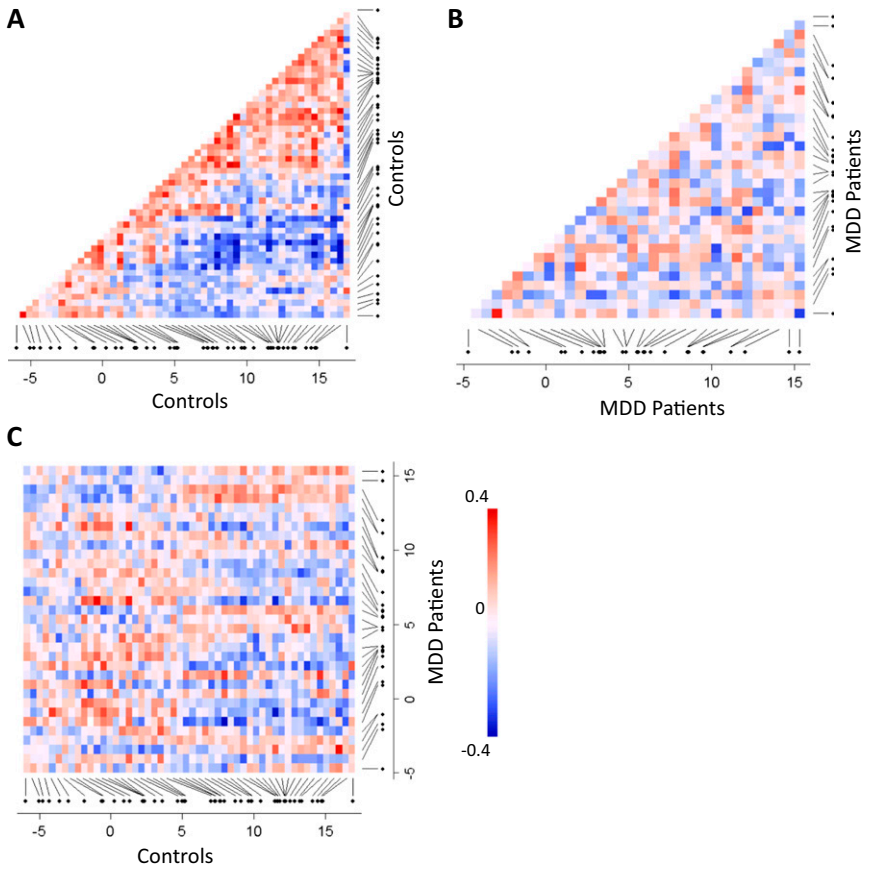


Fig. S9. Control, but not MDD, samples with a similar TOD exhibit correlated gene expression levels. Shown is a sample-sample correlation matrix (shown as heat maps) for 52 controls and 33 MDD cases in the DLPFC, using 108 genes with $P < 0.005$. Shown are the control-control matrix (*A*), the MDD-MDD matrix (*B*), and the control-MDD matrix (*C*). R values are illustrated using color, with red indicating a positive correlation ($R = 0.4$) and blue indicating a negative correlation ($R = -0.4$).

Table S1. Enrichment analysis of ranked *P* values for cyclic patterns

Name	Concept type	*P* value	FDR	Direction
Rhythmic process	GO biological process	4.12E-13	1.20E-09	Enriched
Circadian rhythm	GO biological process	3.15E-12	4.57E-09	Enriched
Circadian rhythm-mammal	KEGG pathway	2.62E-10	5.51E-08	Enriched
PAS fold	pFAM	5.63E-10	1.63E-07	Enriched
Basic region leucine zipper	pFAM	6.05E-07	8.77E-05	Enriched
Sequence-specific DNA binding transcription factor activity	GO molecular function	7.16E-06	0.0046	Enriched
Helix–loop–helix DNA-binding domain	pFAM	6.72E-05	0.0065	Enriched

Combined *P* values across six regions using Fisher's method for all transcripts were analyzed by LRpath (http://lrpath.ncibi.org). This method computes enrichment ratios for Biocarta, Gene Ontology (GO), KEGG, and Panther pathways, and reports enrichment *P* values, false-discovery rate (FDR), and whether the direction is enrichment or depletion. Shown are the pathways with FDR < 0.05. pFAM, Protein Family Database.

Table S2. Demographic and clinical information for 34 patients with MDD

Case no.	pH	Sex	Ethnicity	Cause of death	Age, y	Toxicology screen results	TOD
56	7.13	F	Caucasian	Suicide, FCO	72	Pos: propoxyphene (Darvon)	9.5
57	7.11	M	Caucasian	Suicide, asphyxiation*	19	Pos: methamphetamine, amphetamine	14.7
58	6.93	M	Caucasian	Suicide, asphyxiation*	58	Pos: fluoxetine (Prozac; D)	7.2
59	6.82	M	Caucasian	Natural, cardiac	52	Pos: fluoxetine (Prozac; D)	1.2
60	7.19	M	Caucasian	Suicide, asphyxiation*	49	Neg	5.5
61	6.91	M	Caucasian	Natural, cardiac*	46	Pos: diazepam (Valium), chlordiazepoxide (Librium), phenobarbital (D)	6.3
62	7.00	M	Caucasian	FCO	49	Pos: propoxyphene (Darvon), amitriptyline (Elavil; D), sertraline (Zoloft; D), zolpidem (Ambien; D)	−1.1
63	6.79	M	Caucasian	Suicide, asphyxiation*	39	Neg	3.5
64	7.04	M	Caucasian	Accident, asphyxiation*	35	Neg	8.6
65	6.59	F	Caucasian	Natural, cardiac	50	Pos: citalopram (Celexa; D)	−1.0
66	7.25	M	Caucasian	Suicide, asphyxiation*	47	Neg	3.3
67	6.68	F	Caucasian	Natural, cardiac	80	Pos: phenobarbital (D), butalbital (Axocet; D), doxylamine (Unisom; D)	−2.1
68	7.17	M	Caucasian	Natural, cardiac*	63	Neg	4.9
69	6.87	M	Caucasian	FCO	36	Pos: propoxyphene (Darvon) hydrocodone (Vicodin; D), clonazepam (Klonopin; D)	−4.7
70	7.05	M	Caucasian	Natural, cardiac	66	Neg	11.2
71	7.06	M	Caucasian	Suicide, asphyxiation*	19	Pos: tetrahydrocannabinol (D)	5.6
77	7.07	M	Caucasian	Suicide, asphyxiation*	28	Neg	6.0
78	7.06	F	Caucasian	Natural, cardiac	44	Pos: lidocaine (D)	6.5
79	7.16	F	Caucasian	Suicide, GSW*	53	Pos: sertraline (Zoloft; D)	15.3
80	6.90	M	Caucasian	Suicide, asphyxiation*	56	Pos: paroxetine (Paxil; D)	−1.7
81	7.13	M	Caucasian	Suicide, asphyxiation*	34	Neg	5.9
82	6.76	F	Caucasian	Suicide, asphyxiation*	46	Pos: carbamazepine (Tegretol; D), diphenhydramine (Benedryl; D), alprazolam (Xanax; D)	3.3
83	6.58	M	Caucasian	Suicide, GSW*	52	Neg	8.6
84	7.39	F	Caucasian	FCO	46	Pos: doxylamine (Unisom), propoxyphene (Darvon), mirtazapine (Remeron), citalopram (Celexa), fentanyl citrate (Fentanyl), carisoprodol (Soma; D), atropine (D), promethazine (Phenergan; D), hydroxyzine (Vistaril; D)	0.9
85	6.99	M	Caucasian	Suicide, asphyxiation*	40	Pos: O-desmethylvenlafaxine (Pristiq)	4.9
86	6.84	M	Caucasian	Natural, cardiac	50	Pos: fluoxetine (Prozac; D)	4.7
87	7.18	M	Caucasian	Suicide, asphyxiation*	39	Pos: cocaine (D), hydrocodone (Vicodin; D)	3.2
88	6.40	M	Caucasian	Natural, cardiac	77	Neg	12.0
89	7.16	M	Caucasian	Suicide, asphyxiation*	24	Neg	2.9
90	6.51	M	Caucasian	Suicide, GSW	34	Neg	9.5
91	6.33	M	Caucasian	Suicide, asphyxiation*	29	Neg	11.2
92	6.47	M	Caucasian	Natural, cardiac*	62	Neg	2.2
93	6.41	M	Caucasian	Natural, cardiac*	48	Pos: sertraline (Zoloft; D)	8.6
94	7.00	M	Asian	Suicide, asphyxiation*	31	Neg	3.6

D, detected treatment medications within therapeutic levels; FCO, fatal concentration overdose; GSW, gunshot wound; Pos, positive; Neg, negative.
*Unwitnessed death (for methodology of estimating TOD for unwitnessed death, see *SI Materials and Methods, TOD and Zeitgeber Time*).

Table S3. Counts of samples and microarrays analyzed in this study

Samples and microarrays	AnCg	DLPFC	CB	AMY	HC	NAcc	Total
All U133A microarrays	283	367	162	135	236	293	1,476
Filtered microarrays	277	337	160	135	229	286	1,424
Filtered and for controls	124	161	79	62	108	136	546
Unique samples*	153	172	104	70	137	140	776
Controls†	70	83	51	32	63	66	365
Controls with TOD	55	52	34	29	48	51	269
MDD cases with TOD	34	33	14	13	32	29	155

*Counts for "Filtered 133A microarrays" contain duplicate microarrays for most samples; hence, the counts for "Unique samples" are smaller.
†"Unique samples" consist of controls and patients with mental disorders.

Table S4. Demographic and clinical information for the 55 controls

Subject no.	pH	Sex	Ethnicity	Cause of death	Age, y	TOD
1	7.13	M	Caucasian	Cardiac*	64	−0.6
2	6.5	M	Caucasian	Cardiac	63	−0.6
3	6.4	M	African American	Cardiac	59	0.9
4	6.53	M	Caucasian	Cardiac	52	−5.0
5	6.58	M	Caucasian	Cardiac	58	12.2
6	6.63	M	Caucasian	Blunt force trauma	78	9.9
7	7.11	F	Caucasian	Exsanguination	62	11.7
8	7.04	M	Caucasian	Blunt force trauma	32	−1.9
9	7	M	Caucasian	Cardiac	79	11.5
10	n/a	M	Caucasian	Cardiac	55	7.3
11	7.15	M	Caucasian	Cardiac	30	14.1
12	6.76	M	Caucasian	Cardiac	77	8.7
13	6.96	M	Caucasian	Cardiac*	67	−4.8
14	7.14	M	Caucasian	Blunt force trauma	56	12.2
15	7.21	F	Caucasian	Cardiac	73	14.8
16	7.25	M	Caucasian	Cardiac	63	9.7
17	7.18	M	Caucasian	Cardiac	75	0.3
18	7.12	M	Caucasian	Exsanguination	69	5.2
19	6.55	F	Caucasian	Cardiac	68	3.6
20	7.18	M	Caucasian	Cardiac	55	2.4
21	7.05	F	Caucasian	Blunt force trauma	45	16.8
22	6.59	M	Caucasian	Cardiac	69	12.6
23	6.88	M	Caucasian	Cardiac	63	7.9
24	6.94	M	Caucasian	Cardiac*	66	16.9
25	6.85	M	Caucasian	Cardiac	56	7.7
26	6.59	M	Caucasian	Cardiac	60	2.3
27	6.6	F	Caucasian	Pulmonary embolism	45	11.8
28	6.98	M	Caucasian	Cardiac	56	−3.0
29	6.68	M	Caucasian	Cardiac*	49	2.2
30	7.07	M	Caucasian	Cardiac*	40	−4.3
31	7.21	F	Caucasian	Pulmonary insufficiency	74	17.1
32	6.88	M	African American	Hemorrhagic pericarditis and epicarditis	65	5.0
33	7.01	M	Caucasian	Cardiac	41	12.2
34	7.02	M	Caucasian	Electrocution*	39	11.5
35	6.69	M	Caucasian	Cardiac	67	−3.6
36	6.9	F	Caucasian	Exsanguination	70	12.9
37	6.76	M	Caucasian	Cardiac	35	2.3
38	6.3	F	Asian	Cardiac	47	14.9
39	6.64	M	Caucasian	Cardiac	53	1.3
40	6.81	M	Pacific Islander	Cardiac	39	14.2
41	6.87	M	Caucasian	Cardiac	44	−6.0
42	6.97	M	Caucasian	Electrocution	32	7.3
43	6.62	M	Caucasian	Cardiac	77	7.0
44	7.03	M	Caucasian	Cardiac*	70	5.1
45	6.61	M	Caucasian	Cardiac	54	7.7
46	6.99	F	Caucasian	Cardiac*	60	10.5
47	6.6	M	Caucasian	Cardiac	50	4.7
48	6.86	M	Caucasian	Cardiac	45	11.7
49	7.1	M	Asian	Cardiac	43	9.9
50	6.79	M	Caucasian	Cardiac	48	13.2
51	7.02	M	Caucasian	Cardiac	58	9.1
52	6.89	M	Caucasian	Cardiac	55	14.5
53	6.83	F	Caucasian	Cardiac	64	13.4
54	6.97	M	Caucasian	Drowning	18	12.2
55	6.76	M	Caucasian	Glomerulonephritis	40	3.1

Sunrise time was designated as TOD = 0, with a range of −6 to 18 h indicating 6 h before and 18 h after sunrise. Sunrise time was adjusted for season (Fig. S1). All subjects had an AFS of 0, indicating rapid death occurring within 1 h. F, female; M, male; n/a, not available; pH, brain tissue pH.
*Unwitnessed death.