Evidence for widespread dysregulation of circadian clock progression in human cancer

Jarrod Shilts [1] , **Guanhua Chen** [2] , **Jacob J Hughey** [Corresp. 1, 3]

1 Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States

2 Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, United States

3 Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, United States

Corresponding Author: Jacob J Hughey
Email address: jakejhughey@gmail.com

The ubiquitous daily rhythms in mammalian physiology are guided by progression of the circadian clock. In mice, systemic disruption of the clock can promote tumor growth. In vitro, multiple oncogenes can disrupt the clock. However, due to the difficulties of studying circadian rhythms in solid tissues in humans, whether the clock is disrupted within human tumors has remained unknown. We sought to determine the state of the circadian clock in human cancer using publicly available transcriptome data. We developed a method, called the clock correlation distance (CCD), to infer circadian clock progression in a group of samples based on the co-expression of 12 clock genes. Our method can be applied to modestly sized datasets in which samples are not labeled with time of day and coverage of the circadian cycle is incomplete. We used the method to define a signature of clock gene co-expression in healthy mouse organs, then validated the signature in healthy human tissues. By then comparing human tumor and non-tumor samples from twenty datasets of a range of cancer types, we discovered that clock gene co-expression in tumors is consistently perturbed. Subsequent analysis of data from clock gene knockouts in mice suggested that perturbed clock gene co-expression in human cancer is not caused solely by the inactivation of clock genes. Furthermore, focusing on lung cancer, we found that human lung tumors showed systematic changes in expression in a large set of genes previously inferred to be rhythmic in healthy lung. Our findings suggest that clock progression is dysregulated in many solid human cancers and that this dysregulation could have broad effects on circadian physiology within tumors. In addition, our approach opens the door to using publicly available data to infer circadian clock progression in a multitude of human phenotypes.

1 Evidence for widespread dysregulation of circadian clock
2 progression in human cancer

4 Jarrod Shilts[1], Guanhua Chen[2], Jacob J. Hughey[1,3]

5 [1] Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN,
6 United States
7 [2] Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison,
8 WI, United States
9 [3] Department of Biological Sciences, Vanderbilt University, Nashville, TN, United States

10 Corresponding author:
11 Jacob J. Hughey[1,3]

12 Email address: jakejhughey@gmail.com

Abstract

14 The ubiquitous daily rhythms in mammalian physiology are guided by progression of the
15 circadian clock. In mice, systemic disruption of the clock can promote tumor growth. In vitro,
16 multiple oncogenes can disrupt the clock. However, due to the difficulties of studying circadian
17 rhythms in solid tissues in humans, whether the clock is disrupted within human tumors has
18 remained unknown. We sought to determine the state of the circadian clock in human cancer
19 using publicly available transcriptome data. We developed a method, called the clock correlation
20 distance (CCD), to infer circadian clock progression in a group of samples based on the co-
21 expression of 12 clock genes. Our method can be applied to modestly sized datasets in which
22 samples are not labeled with time of day and coverage of the circadian cycle is incomplete. We
23 used the method to define a signature of clock gene co-expression in healthy mouse organs,
24 then validated the signature in healthy human tissues. By then comparing human tumor and
25 non-tumor samples from twenty datasets of a range of cancer types, we discovered that clock
26 gene co-expression in tumors is consistently perturbed. Subsequent analysis of data from clock
27 gene knockouts in mice suggested that perturbed clock gene co-expression in human cancer is
28 not caused solely by the inactivation of clock genes. Furthermore, focusing on lung cancer, we
29 found that human lung tumors showed systematic changes in expression in a large set of genes
30 previously inferred to be rhythmic in healthy lung. Our findings suggest that clock progression is
31 dysregulated in many solid human cancers and that this dysregulation could have broad effects
32 on circadian physiology within tumors. In addition, our approach opens the door to using publicly
33 available data to infer circadian clock progression in a multitude of human phenotypes.

Introduction

35 Daily rhythms in mammalian physiology are guided by a system of oscillators called the
36 circadian clock (Dibner, Schibler & Albrecht, 2010). The core clock consists of feedback loops
37 between several genes and proteins, and based on work in mice, is active in nearly every tissue

38   in the body (Yoo et al., 2004; Zhang et al., 2014). The clock aligns itself to environmental cues,
39   particularly cycles of light-dark and food intake (Damiola et al., 2000; Asher et al., 2010; Eckel-
40   Mahan et al., 2013). In turn, the clock regulates various aspects of metabolism (Nakahata et al.,
41   2009; Cho et al., 2012; Neufeld-Cohen et al., 2016; Guerrero-Vargas et al., 2017) and is tightly
42   linked to the cell cycle (Matsuo et al., 2003; Gréchez-Cassiau et al., 2008; Geyfman et al., 2012;
43   Feillet et al., 2014; Bieler et al., 2014; Matsu-Ura et al., 2016; El-Athman et al., 2017).

44   Consistent with the tight connections between the circadian clock, metabolism, and the cell
45   cycle, multiple studies have found that systemic disruption of the circadian system can promote
46   cancer. In humans, long-term rotating shift work and night shift work, which perturb sleep-wake
47   and circadian rhythms, have been associated with breast, colon, and lung cancer
48   (Schernhammer et al., 2001, 2003, 2013; Wegrzyn et al., 2017). In mice, environmental
49   disruption of the circadian system (e.g., through severe and chronic jet lag) increases the risk of
50   breast cancer and hepatocellular carcinoma (Van Dycke et al., 2015; Kettner et al., 2016).
51   Furthermore, both environmental and genetic disruption of the circadian system promote tumor
52   growth and decrease survival in a mouse model of human lung adenocarcinoma
53   (Papagiannakopoulos et al., 2016). While these studies support the link from the clock to cancer,
54   other studies have established a link in the other direction, whereby multiple components of a
55   tumor, including the RAS and MYC oncogenes, can induce dysregulation of the circadian clock
56   (Relógio et al., 2014; Michael et al., 2015; Altman et al., 2015). Despite this progress, however,
57   whether the clock is actually disrupted within human tumors has remained unclear. Given recent
58   findings that stimulating the circadian clock slows tumor growth in a mouse model of melanoma
59   (Kiessling et al., 2017), it is important to determine the extent of clock disruption across human
60   cancers, in order to delineate the general potential of anti-tumor strategies based on restoring or
61   improving clock function.

62   When the mammalian circadian clock is progressing or "ticking" normally, clock genes and clock-
63   controlled genes show characteristic rhythms in expression throughout the body and in vitro
64   (Balsalobre, Damiola & Schibler, 1998; Yoo et al., 2004; Zhang et al., 2014). These rhythms can
65   be used to monitor clock progression in humans (Laing et al., 2017; Hughey, 2017).
66   Measurements of circadian rhythms through time-course experiments have revealed that the
67   clock is altered or perturbed in some human breast cancer cell lines (Rossetti et al., 2012; Xiang
68   et al., 2012). Computational methods for analyzing rhythmicity require that samples be labeled
69   with time of day (or time since start of experiment) and acquired throughout the 24-h cycle
70   (Hughey, Hastie & Butte, 2016; Thaben & Westermark, 2016; Wu et al., 2016). Unfortunately,
71   existing data from resected human tumors meet neither of these criteria. In this scenario, one
72   approach might be to look for associations between the expression levels of clock genes and
73   other biological and clinical variables. For example, in human breast cancer, the expression
74   levels of several clock genes have been associated with metastasis-free survival (with the
75   direction of association depending on the gene) (Cadenas et al., 2014). However, because a
76   functional circadian clock is marked less by the levels of gene expression than by rhythms in
77   gene expression, this type of analysis cannot necessarily be used to determine whether the
78   clock is progressing normally.

79   A more sophisticated approach is to assume the presence of rhythms and to infer a cyclical
80   ordering of samples, using methods such as Oscope or CYCLOPS (Leng et al., 2015; Anafi et

81   al., 2017). By applying CYCLOPS to transcriptome data from hepatocellular carcinoma, Anafi et
82   al. found evidence for weaker or disrupted rhythmicity of several clock genes, as well as genes
83   involved in apoptosis and JAK-STAT signaling, in tumor samples compared to non-tumor
84   samples (Anafi et al., 2017). Although CYCLOPS does not require that samples be labeled with
85   time of day, it does require that the samples cover the entire cycle. Consequently, the authors
86   recommend that CYCLOPS be applied to datasets from humans with at least 250 samples
87   (Anafi et al., 2017). In addition, although CYCLOPS can be used to infer rhythmicity in the
88   expression of individual genes, it is not designed to quantify differences in the overall pattern of
89   those rhythms (e.g., the relative phasing between genes) between conditions (e.g., healthy and
90   diseased).

91   Rather than attempting to infer an oscillation, an alternative approach might be to take
92   advantage of the pattern of co-expression (e.g., pairwise correlation) that results from different
93   clock genes having rhythms with different phases. Indeed, a previous study found different levels
94   of co-expression between a few clock genes in different subtypes and grades of human breast
95   cancer (Cadenas et al., 2014). Although this finding was an important first step, its
96   generalizability has been limited because the correlations in expression were not examined for
97   all clock genes, in other human cancer types, or in healthy tissues where the circadian clock is
98   known to be functional. Thus, a definitive answer to whether the circadian clock is progressing
99   normally in human tumors is still lacking.

100 In this study, we developed a computational method to characterize the extent of dysregulation
101 of circadian clock progression in human cancer. Using transcriptome data from mice, we defined
102 a robust signature of clock progression based on the co-expression of clock genes. We validated
103 the signature using transcriptome data from various organs in humans, then examined the
104 extent to which the signature was perturbed in tumor compared to non-tumor samples from The
105 Cancer Genome Atlas (TCGA) and from multiple independent datasets. Our findings suggest
106 that dysregulation of circadian clock progression is present in a wide range of human cancers, is
107 not caused solely by the inactivation of core clock genes, and is accompanied by systematic
108 changes in broader circadian gene expression.

Materials and Methods

Study design

111 The main goals of this study were to develop a co-expression-based signature of the circadian
112 clock, to validate the signature in healthy human organs, and to use the signature to infer the
113 extent of normal circadian clock progression in human cancer. We focused on transcriptome
114 data because of its wide availability. We selected the datasets of circadian gene expression in
115 mice (both for defining the reference signature and for comparing clock gene knockouts to wild-
116 type) to represent multiple organs, light-dark regimens, and microarray platforms.

117 We selected the datasets from healthy human organs to include as many circadian studies as
118 possible and to include a range of organs. The dataset from human skin consisted of samples
119 taken at only three time-points for each subject (9:30am, 2:30pm, and 7:30pm). Datasets from
120 human blood consisted of multiple samples taken throughout the 24-h cycle for each subject.

121 Datasets from human brain were based on postmortem tissue from multiple anatomical areas,
122 and zeitgeber time for each sample was calculated using the respective donor's date and time of
123 death and geographic location. Datasets from cells cultured in vitro were based on time-courses
124 following synchronization by dexamethasone, serum, or alternating temperature cycles. For
125 GSE45642 (human brain), we only included samples from control subjects (i.e., we excluded
126 subjects with major depressive disorder).

127 For datasets of human cancer, we analyzed data from the Cancer Genome Atlas (TCGA), as it
128 contains samples from various cancer types. To complement the breadth of cancer types in
129 TCGA data, we further included human liver cancer and lung cancer datasets from NCBI GEO.
130 We analyzed human cancer datasets that had at least 30 non-tumor samples. When analyzing
131 clock gene expression in human cancer, unless otherwise noted, we included all tumor and non-
132 tumor samples, not just those from patients from whom both non-tumor and tumor samples were
133 collected. For details of the datasets, all of which are publicly available, see Table S1.

Processing the gene expression data

135 For TCGA samples, we obtained the processed RNA-seq data (in units of transcripts per million,
136 TPM, on a gene-level basis) and the corresponding metadata (cancer type, patient ID, etc.) from
137 GSE62944 (Rahman et al., 2015). For E-MTAB-3428, we downloaded the RNA-seq read files
138 from the European Nucleotide Archive, used Salmon to quantify transcript-level abundances in
139 units of TPM (Patro et al., 2017), then used the mapping between Ensembl Transcript IDs and
140 Entrez Gene IDs to calculate gene-level abundances.

141 For the remaining datasets, raw (in the case of Affymetrix) or processed microarray data were
142 obtained from NCBI GEO and processed using MetaPredict, which maps probes to Entrez Gene
143 IDs and performs intra-study normalization and log-transformation (Hughey & Butte, 2015).
144 MetaPredict processes raw Affymetrix data using RMA and customCDFs (Irizarry et al., 2003;
145 Dai et al., 2005). As in our previous study, we used ComBat to reduce batch effects between
146 anatomical areas in human brain and between subjects in human blood (Johnson, Li &
147 Rabinovic, 2007; Hughey & Butte, 2016). We also used ComBat to normalize expression
148 between between subjects in the circadian human skin dataset (Fig. S8). ComBat is an empirical
149 Bayes method to standardize the mean and variance of gene expression values between groups
150 of samples.

Analyzing clock gene expression and co-expression

152 We first focused on the expression of 12 genes that are considered part of the core clock or are
153 directly controlled by the clock and that show strong, consistently phased rhythms in multiple
154 mouse organs (Zhang et al., 2014; Hughey, Hastie & Butte, 2016). We calculated times of peak
155 expression (Fig. S2) and strengths of circadian rhythmicity of expression in wild-type and
156 knockout mice (Fig. S15) using ZeitZeiger (Hughey, Hastie & Butte, 2016), with three knots for
157 the periodic smoothing splines (Helwig & Ma, 2014). ZeitZeiger uses the time of day information
158 for each sample.

159 We quantified the relationship between expression values of pairs of genes using the Spearman
160 correlation (rho), which is rank-based and therefore invariant to monotonic transformations such

161   as the logarithm and less sensitive to outliers than the Pearson correlation. Using the biweight
162   midcorrelation (Song, Langfelder & Horvath, 2012), which is also robust to outliers, gave very
163   similar results. All heatmaps of clock gene co-expression in Fig. 1, Fig. 2, and the supplemental
164   figures have the same mapping of correlation value to color, so they are directly visually
165   comparable.

166   For a group of samples from a given dataset, we estimated the strength of clock gene co-
167   expression (Fig. S7) as the difference between the 95th and 5th percentiles of the distribution of
168   Spearman correlations between pairs of the 12 clock genes.

169   We calculated the reference Spearman correlation for each pair of clock genes (Table S2) using
170   a fixed-effects meta-analysis (Hedges & Olkin, 1985) of the eight mouse datasets shown in Fig.
171   S1. First, we applied the Fisher z-transformation ($arctanh$) to the correlations from each
172   dataset. Then we calculated a weighted average of the transformed correlations, where the
173   weight for dataset i was $n_i - 3$ (corresponding to the inverse variance of the transformed
174   correlation), where n_i is the number of samples in dataset i. Finally, we applied the
175   inverse transformation ($tanh$) to the weighted average.

176   To quantify the similarity in clock gene co-expression between the reference and a group of
177   samples from another transcriptome dataset, we calculated the Euclidean distance between the
178   respective Spearman correlation vectors, which contain all values in the strictly lower (or strictly
179   upper) triangular part of the correlation matrix. We call this distance the clock correlation
180   distance (CCD). A smaller CCD indicates greater similarity in clock gene co-expression.
181   Although here we used Euclidean distance, other distance metrics could be used as well.

182   To estimate the probability that the observed CCD is less than would be expected by chance, we
183   randomly selected 12 genes (from all genes measured in that dataset) 1000 times, in order to
184   estimate a null distribution of CCDs. We then calculated the exact one-sided p-value using the
185   number of random samplings that gave a CCD less than or equal to the observed CCD and the
186   method of Phipson and Smyth in the statmod R package (Phipson & Smyth, 2010). For the eight
187   mouse datasets that form the reference, we calculated the CCD between each unique pair of
188   datasets, essentially taking one dataset in each pair to be the reference.

189   To quantify the extent to which clock gene co-expression is perturbed in samples from one
190   condition (e.g., tumor) relative to samples from another condition (e.g., non-tumor) in a given
191   dataset, we calculated the CCD for each condition relative to the mouse reference, then
192   calculated the difference between those CCDs. We call this difference the delta clock correlation
193   distance (ΔCCD). In this study, a positive ΔCCD indicates that clock gene co-expression in
194   tumor (or knockout) samples is less similar to the reference than is clock gene co-expression in
195   non-tumor (or wild-type) samples.

196   To evaluate the probability that the ΔCCD is greater than would be expected by chance, we
197   permuted the condition labels of the samples and recalculated the ΔCCD 1000 times, always
198   keeping the reference fixed. We then calculated the exact one-sided p-value using the number
199   of permutations that gave a ΔCCD greater than or equal to the observed ΔCCD and the method
200   of Phipson and Smyth in the statmod R package (Phipson & Smyth, 2010). Since we used the

201   one-sided p-value, the alternative hypothesis was that non-tumor (or wild-type) is more similar to
202   the reference than is tumor (or knockout).

203   To calculate the ΔCCD for individual tumor grades in TCGA data, we used the clinical metadata
204   provided in GSE62944. We analyzed all combinations of TGCA cancer type and tumor grade
205   that included at least 50 tumor samples. In each case, we calculated the ΔCCD using all non-
206   tumor samples of the respective cancer type.

207   To compare ΔCCD and tumor purity, we used published consensus purity estimates for TCGA
208   tumor samples (Aran, Sirota & Butte, 2015). The estimates are based on DNA methylation,
209   somatic copy number variation, and the expression of immune genes and stromal genes (none
210   of which are clock genes).

211   To quantify differential expression between tumor and non-tumor samples and between
212   knockout and wild-type samples, we used limma and voom (Smyth, 2004; Law et al., 2014). To
213   ensure a fair comparison between human and mouse data, we ignored time of day information
214   in the mouse samples. We quantified the variation in expression of clock genes in each dataset
215   and condition using the median absolute deviation (MAD), which is less sensitive to outliers than
216   the standard deviation.

217   Analyzing circadian gene expression in human lung cancer

218   We obtained the set of rhythmic transcripts (identified by microarray probe ID) inferred by
219   CYCLOPS to be rhythmic in healthy human lung (Anafi et al., 2017). This set contains
220   transcripts whose abundance was well described by a sinusoidal function of CYCLOPS phase in
221   samples from both the Laval and GRNG sites of GSE23546 (Bonferroni-corrected P < 0.05 and
222   peak/trough ratio > 2) and whose orthologs were rhythmic in mouse lung. We mapped the
223   microarray probes to Entrez Gene IDs and calculated the acrophase for each gene as the
224   circular mean of the provided acrophase for the corresponding probes. Phase values inferred by
225   CYCLOPS are relative, so Anafi et al. adopted the convention of setting phase π (in radians)
226   to the average acrophase of the PAR bZip transcription factors (DBP, HLF, and TEF) (Anafi et
227   al., 2017).

228   Because some of the module preservation statistics calculated by NetRep use the expression
229   matrix directly, we used ComBat (Johnson, Li & Rabinovic, 2007) to merge the expression data
230   from the Laval and Groningen sites of GSE23546. We then followed WGCNA's recommended
231   procedure for identifying gene modules, i.e., groups of genes with correlated expression across
232   samples (Langfelder & Horvath, 2008). Briefly, starting with the merged expression data, we
233   calculated the Spearman correlation matrix for the 1,292 genes, then used the signed method
234   with a soft thresholding power of 12 to calculate the adjacency matrix, then calculated the
235   topological overlap dissimilarity matrix. We used hierarchical clustering of the dissimilarity matrix
236   and adaptive branch pruning to define modules of genes, followed by a procedure to merge
237   closely related modules. We then used the DAVID web application (version 6.8) to discover
238   functional categories enriched in each module (Huang, Sherman & Lempicki, 2009).

239   We used NetRep (Ritchie et al., 2016) to calculate seven module preservation statistics between
240   healthy human lung and the non-tumor and tumor samples from five datasets of human lung

241   cancer (LUAD and LUSC from TCGA and GSE19188, GSE19804, and GSE32863 from NCBI
242   GEO). We excluded GSE10072, because it included expression levels for only 876 of the 1,292
243   genes (the other five datasets included at least 1,199 of the 1,292 genes). When calculating
244   module preservation, NetRep automatically removed any genes that were not measured in both
245   healthy human lung and the group of samples to which it was being compared. For non-tumor
246   and tumor samples from each dataset, we calculated the Spearman correlation matrix and
247   adjacency matrix using the same procedure as for healthy human lung.

248   To evaluate the statistical significance of the difference in each module preservation statistic
249   between non-tumor and tumor samples, we performed permutation testing and calculated one-
250   sided p-values similarly to the procedure for ΔCCD, permuting the sample labels (non-tumor or
251   tumor) in the test dataset and recalculating module preservation 1000 times. Note this is
252   different from the standard way WGCNA and NetRep perform permutations, which is to shuffle
253   the gene labels.

254   Periodic smoothing splines of log_2 fold-change as a function of acrophase in healthy lung were fit
255   using ZeitZeiger (Hughey, Hastie & Butte, 2016). P-values of non-zero amplitude of the spline
256   fits were calculated using 1000 permutations of the relationship between acrophase and log_2
257   fold-change and the method of Phipson and Smyth (Phipson & Smyth, 2010). For Phase Set
258   Enrichment Analysis (Zhang et al., 2016), we used the canonical pathways gene sets from
259   MSigDB (Liberzon et al., 2011) and looked for gene sets with a q-value ≤ 0.1 against a uniform
260   distribution and vector-average value within 0.4 radians of either 0.07 π or 1.06 π (the mean
261   phases of the peak and trough of the spline fits, respectively). No gene sets met the criteria for
262   the former.

Results

Consistent co-expression of clock genes in mice

265   The progression of the mammalian circadian clock is marked by daily rhythms in gene
266   expression throughout the body (Zhang et al., 2014). We hypothesized that the relative phasing
267   of different genes' rhythms would result in those genes having a characteristic pattern of co-
268   expression. Such a pattern could be used to infer the progression of the clock within a group of
269   samples, even if the samples are not labeled with time of day, as long as they were taken from
270   multiple parts of the circadian cycle.

271   Although most circadian gene expression is tissue-specific, the rhythms of a select group of
272   genes show consistent relative phasing in multiple organs in mice and humans (Hughey & Butte,
273   2016). We focused on 12 genes from this group, each of which is part of the core clock or
274   directly controlled by the clock, in order to define a general signature of circadian clock
275   progression. For the rest of the paper, we refer to these 12 genes as "clock genes."

276   As a first test of our hypothesis, we assembled eight publicly available datasets of genome-wide
277   circadian gene expression from various healthy mouse organs (Oster et al., 2006; Hoogerwerf et
278   al., 2008; Hughes et al., 2009; Negoro et al., 2012; Geyfman et al., 2012; Haspel et al., 2014;
279   Zhang et al., 2014) (Table S1). The organs were collected from mice under both constant

280 darkness and alternating light-dark cycles. For each dataset, we calculated the Spearman
281 correlation between expression values (over all samples) of each pair of clock genes.

282 The pattern of co-expression was highly consistent across datasets (for 59 of 66 gene pairs, the
283 sign of the Spearman correlation was the same in at least 6 of 8 datasets; Fig. S1). Co-
284 expression revealed two groups of genes, where the genes within a group tended to be
285 positively correlated with each other and negatively correlated with genes in the other group.
286 Genes in the first group (Arntl, Npas2, and Clock), which form the positive arm of the clock
287 (Partch, Green & Takahashi, 2014), peaked in expression shortly before zeitgeber time 0 (ZT0,
288 which corresponds to time of lights on or sunrise; Fig. S2). Genes in the second group (Cry2,
289 Nr1d1, Nr1d2, Per1, Per2, Per3, Dbp, and Tef), which form the negative arms of the clock,
290 peaked in expression near ZT10. Cry1, which was part of the first group in some datasets and
291 the second group in others, tended to peak in expression near ZT18. These results indicate that
292 the progression of the circadian clock in various mouse organs produces a consistent pattern of
293 co-expression between clock genes.

294 To construct a single reference pattern of clock gene co-expression, we combined the eight
295 datasets in a fixed-effects meta-analysis (Fig. 1A-B and Materials and Methods). To compare
296 clock gene co-expression between the reference and a group of samples from another dataset,
297 we developed a metric we call the clock correlation distance (CCD), which corresponds to the
298 Euclidean distance between the respective correlation vectors (Fig. 1A). A smaller CCD
299 indicates a greater similarity in clock gene co-expression. To estimate the probability that the
300 observed CCD is less than would be expected by chance, we estimated the distribution of CCD
301 that would result from using 12 randomly selected genes instead of the 12 clock genes. As a
302 positive control, we calculated the Euclidean distance between the correlation vectors
303 corresponding to all pairs of the eight mouse datasets ($P \leq 0.001$ by permutation test for each
304 pair; Fig. 1C and Fig. S3).

305 We used the reference pattern and the CCD metric to verify that clock gene co-expression
306 arises from the phasing of the genes' rhythms relative to each other, not their phasing relative to
307 time of day. Thus, the pattern is not affected by phase differences between groups of samples,
308 such as those caused by daytime feeding in mice (Vollmers et al., 2009) ($P \leq 0.001$ for CCD
309 relative to reference; Fig. S4).

310 Most computational methods for quantifying circadian rhythmicity and inferring the status of the
311 clock require that samples be acquired over the entire 24-h cycle. Because our approach does
312 not attempt to infer oscillations, we wondered if it would be robust to partial coverage of the 24-h
313 cycle. We therefore analyzed clock gene co-expression in three of the original eight datasets, in
314 samples acquired during the first 8 h of the day (or subjective day) or the first 8 h of the night (or
315 subjective night). In each dataset, clock gene co-expression was preserved in both daytime and
316 nighttime samples ($P \leq 0.001$; Fig. S5). Thus, our approach can detect the signature of clock
317 progression in groups of samples without using time of day information, even if the samples'
318 coverage of the 24-h cycle is incomplete.

Validation of the co-expression pattern in human tissues

319

320 We next examined clock gene co-expression in nine publicly available datasets of circadian
321 gene expression from human tissues: one from skin, two from brain, three from blood, and three
322 from cells cultured in vitro (Hughes et al., 2009; Spörl et al., 2012; Möller-Levet et al., 2013; Li et
323 al., 2013; Janich et al., 2013; Archer et al., 2014; Hoffmann et al., 2014; Arnardottir et al., 2014;
324 Chen et al., 2016) (Table S1). Samples from these datasets are labeled with time of day (or in
325 the case of cells in vitro, time since synchronization) and were taken from multiple time-points in
326 the circadian cycle. Other than the two datasets of U2OS cells, which are derived from an
327 osteosarcoma but show normal rhythms of the core clock genes (Vollmers, Panda & DiTacchio,
328 2008; Maier et al., 2009), all samples were from healthy tissue.

329 Although gene expression rhythms in mice and humans show different phasing relative to the
330 light-dark cycle (Hughey & Butte, 2016), the pattern of clock gene co-expression in the human
331 datasets significantly resembled the reference pattern from mice (P ≤ 0.001 for 3 of 3 in vitro
332 datasets and 4 of 6 in vivo datasets; Fig. 1C-D and Fig. S6). The CCDs of the datasets varied:
333 skin and in vitro cells had the smallest (i.e., the strongest similarity to mice), brain was
334 intermediate, and blood had the largest (Fig. 1C). Correlations between clock genes in the latter
335 two organs were generally weaker than those in mice (Fig. S7), consistent with noisier circadian
336 rhythms in those datasets (Hughey & Butte, 2016). The co-expression pattern in the human skin
337 dataset, for which samples were acquired from each individual at three time-points (9:30am,
338 2:30pm, and 7:30pm), was due to co-expression both across individuals at each time-point and
339 across time-points within each individual (Fig. S8).

340 To confirm our findings in additional human organs, we analyzed eight transcriptome datasets
341 from healthy human lung, liver, skin, and adipose tissue (Schadt et al., 2008; Innocenti et al.,
342 2011; Bossé et al., 2012; Grundberg et al., 2012; Bonder et al., 2014) (Table S1). Samples from
343 these datasets were not collected for the purpose of studying circadian rhythms. Thus, the
344 samples are not labeled with time of day and may not cover the entire circadian cycle.
345 Nonetheless, clock gene co-expression in these datasets was similar to that of mice (P ≤ 0.001
346 for all 8 datasets; Fig. 1C, Fig. 1E, and Fig. S9), with CCDs comparable to those of datasets
347 designed to quantify circadian rhythms in humans in vivo. The generally weaker co-expression
348 patterns and higher CCDs in human in vivo datasets compared to mouse datasets is likely due
349 to data collected from humans having more sources of variation (including technical,
350 environmental, and genetic variation). We conclude that our approach can detect the signature
351 of circadian clock progression in publicly available transcriptome data from various human
352 tissues, even in datasets not designed to quantify circadian rhythms.

Aberrant patterns of clock gene co-expression in human cancer

353

354 We then turned our attention to clock gene co-expression in human cancer. We started with
355 RNA-seq data collected by The Cancer Genome Atlas (TCGA) and reprocessed using the
356 Rsubread package (Rahman et al., 2015). TCGA samples are from surgical resections
357 performed prior to neoadjuvant treatment. We analyzed TCGA data from the 12 cancer types
358 that included at least 30 samples from adjacent non-tumor tissue (Table S1). To study human
359 hepatocellular carcinoma and non-small cell lung cancer in more depth, we also analyzed eight
360 datasets from the NCBI Gene Expression Omnibus (GEO), each of which had tumor and

361 adjacent non-tumor samples (Landi et al., 2008; Hou et al., 2010; Lu et al., 2010; Roessler et al.,
362 2010; Lamb et al., 2011; Selamat et al., 2012; Lim et al., 2013; Villa et al., 2016) (Table S1). For
363 all data from TCGA and GEO, the times of day of surgery are not available. Presumably, most to
364 all of the samples were acquired during surgical working hours (6:00am to 6:00pm).

365 To quantify the extent to which clock gene co-expression is perturbed in tumor samples
366 compared to non-tumor samples in a given dataset, we calculated the difference between the
367 CCDs (relative to the mouse reference) for each condition. We call this difference the delta clock
368 correlation distance (ΔCCD), where a ΔCCD > 0 indicates that clock gene co-expression in
369 tumor samples is less similar to the reference than that of non-tumor samples (Fig. 2A). We
370 continued to use the reference from mouse data, instead of one based on human data, in order
371 to maintain a strict separation of training and validation data and because in the mouse data, the
372 effects of other sources of variation and confounding are minimized. Thus, the mouse data
373 offered the clearest window into the clock gene co-expression that results from circadian clock
374 progression in vivo. To evaluate the probability that the ΔCCD is greater than would be expected
375 by chance, we permuted the condition labels of the samples and recalculated the ΔCCD 1000
376 times.

377 Clock gene co-expression in non-tumor samples was similar to that observed in solid healthy
378 human tissues (P ≤ 0.005 for CCD relative to the mouse reference in 14 of 20 datasets; Fig. 2B-
379 C and Fig. S10). In contrast, co-expression in tumor samples often lacked a significant similarity
380 to the expected pattern (P > 0.05 for CCD in 10 of 20 datasets; Fig. 2B-C and Fig. S10).
381 Consequently, tumor CCDs were significantly higher than non-tumor CCDs (P = $1.4 \cdot 10^{-8}$ by
382 paired t-test) and each human cancer dataset had a ΔCCD > 0 (P ≤ 0.005 in 15 of 20 datasets;
383 Fig. 2D and Fig. S11A). Among the three TCGA cancer types with the lowest ΔCCD, prostate
384 adenocarcinoma had a relatively high non-tumor CCD, whereas renal clear cell carcinoma and
385 thyroid carcinoma each had a relatively low tumor CCD (Fig. S11B). Interestingly, we observed
386 no clear trend between ΔCCD and histological tumor grade (Fig. S12). Overall, these results
387 suggest that circadian clock progression is perturbed in a range of human cancers.

388 Tumors are a complex mixture of cancer cells and various non-cancerous cell types. The
389 proportion of cancer cells in a tumor sample is called the tumor purity and is an important factor
390 to consider in genomic analyses of bulk tumors (Aran, Sirota & Butte, 2015). We therefore
391 examined the relationship between ΔCCD and tumor purity in the TCGA data. With the
392 exception of thyroid carcinoma and prostate adenocarcinoma, ΔCCD and median tumor purity
393 across TCGA cancer types were positively correlated (Fig. S13; Spearman correlation = 0.67, P
394 = 0.059 by exact test). These findings suggest that at least in some tumors, clock progression is
395 perturbed more strongly in cancer cells than in non-cancerous cells.

Distinct patterns of clock gene expression in human cancer and clock gene

397 knockouts in mice

398 To better understand the nature of the dysregulation of clock progression, in particular to
399 determine if the pattern of clock gene co-expression in human tumors could be explained by
400 inactivation of core clock genes, we compared clock gene expression in human cancer and in
401 clock gene knockouts in mice. We assembled seven datasets of circadian gene expression,

402   measured in various tissues, that included samples from wild-type mice and from mice in which
403   at least one core clock gene was knocked out, either in the entire animal or in a specific cell type
404   (Vollmers et al., 2009; Cho et al., 2012; Nikolaeva et al., 2012; Paschos et al., 2012; Dyar et al.,
405   2014; Young et al., 2014; Dudek et al., 2016) (Table S1). For each dataset, we calculated clock
406   gene co-expression in wild-type and knockout samples (Fig. S14).

407   The two datasets with the highest ΔCCD (>50% higher than any ΔCCD we observed in human
408   cancer) were those in which the knockout mice lacked not one, but two components of the clock
409   (Cry1 and Cry2 in GSE13093; Nr1d1 and Nr1d2 in GSE34018; Fig. 3A). The ΔCCDs for the
410   other five datasets were similar to or somewhat lower than the ΔCCDs we observed in human
411   cancer. Due to the smaller sample sizes compared to the human cancer datasets, the ΔCCDs
412   for four of seven datasets of clock gene knockouts were not significantly greater than zero (P >
413   0.05).

414   To further compare clock gene expression in human cancer and clock gene knockouts, we
415   calculated differential expression of the clock genes between non-tumor and tumor samples and
416   between wild-type and knockout samples (Fig. 3B). Differential expression in the knockouts was
417   largely consistent with current understanding of the core clock. For example, knockout of Arntl
418   (the primary transcriptional activator) tended to cause reduced expression of Nr1d1, Nr1d2,
419   Per1, Per2, Per3, Dbp, and Tef, and increased or unchanged expression of the other clock
420   genes. In the double knockout of Cry1 and Cry2 (two negative regulators of the clock), this
421   pattern was reversed. Interestingly, neither of these patterns of differential expression was
422   apparent in human cancer.

423   In the clock gene knockouts, rhythmic expression of most clock genes was reduced or lost
424   (calculated using each sample's time of day; Fig. S15). Without time of day information in the
425   human cancer datasets, it was not possible to directly quantify rhythms of clock genes in non-
426   tumor and tumor samples. However, we reasoned that a proxy for rhythmicity could be the
427   magnitude of variation in expression. Therefore, for each human cancer dataset, we calculated
428   the median absolute deviation (MAD) in expression of the clock genes in non-tumor and tumor
429   samples. We then compared the $\log_2$ ratios of MAD between tumor and non-tumor samples to
430   the $\log_2$ ratios of MAD between knockout and wild-type samples from the clock gene knockout
431   data (Fig. 3C). As expected, samples from clock gene knockouts showed widespread reductions
432   in MAD compared to samples from wild-type mice. In contrast, human tumor samples tended to
433   show similar or somewhat higher MAD compared to non-tumor samples. The distinct patterns of
434   differential expression and differential variability between tumor and non-tumor samples,
435   compared to knockout and wild-type samples, imply that dysregulation of clock progression in
436   human cancer is not due solely to the inactivation of one or more clock genes.

437   Dysregulation of broader circadian gene expression in human lung cancer
438   Finally, we explored how dysregulation of clock progression might affect the circadian
439   transcriptome. We obtained a set of 1,292 genes that were inferred by the CYCLOPS method to
440   be rhythmic in a dataset from healthy human lung (Anafi et al., 2017). By applying WGCNA
441   (Langfelder & Horvath, 2008) to the same dataset, we grouped the 1,292 genes into five
442   modules according to their co-expression (Fig. 4A). Based on DAVID (Huang, Sherman &
443   Lempicki, 2009) and consistent with the analysis of Anafi et al. (Anafi et al., 2017), the five

444 modules were enriched for genes involved in various biological processes, including
445 angiogenesis, phosphoprotein signaling, and alternative splicing (Table S3). We then used
446 NetRep (Ritchie et al., 2016) to determine the extent to which each module was preserved in
447 non-tumor and tumor samples from five datasets of human lung cancer (two from TCGA and
448 three from NCBI GEO). Notably, all five datasets had ≤65 non-tumor samples and the three
449 datasets from NCBI GEO had ≤91 tumor samples, too few for CYCLOPS to produce a reliable
450 ordering. Based on this analysis, four of five modules (comprising 1,256 of 1,292 rhythmic
451 genes) were significantly more strongly preserved in non-tumor compared to tumor samples
452 (Fig. 4B, $P \leq 0.001$ for at least two of seven preservation statistics in at least four of five
453 datasets).

454 We also quantified differential expression of the 1,292 normally rhythmic genes between non-
455 tumor and tumor samples in the five lung cancer datasets. We then examined the relationship
456 between each gene's $\log_2$ fold-change (tumor vs. non-tumor) and its phase of peak expression
457 (acrophase) in healthy lung as inferred by CYCLOPS. Remarkably, we observed a consistent
458 pattern, in which genes with an acrophase near π tended to show the strongest decrease in
459 expression in tumor samples (Fig. 4C, $P \leq 0.001$ of non-zero amplitude by permutation testing
460 for each dataset). Using Phase Set Enrichment Analysis (Zhang et al., 2016), we found that
461 genes with an acrophase near π were particularly enriched for involvement in G protein-
462 coupled receptor signaling (Table S4). Taken together, these findings suggest that perturbed
463 clock progression in human lung cancer is accompanied by systematic changes in broader
464 circadian gene expression.

Discussion

466 Increasing evidence has suggested that systemic disruption of the circadian clock can promote
467 tumor development and that components of a tumor can disrupt the circadian clock (reviewed in
468 (Lamia, 2017)). Until now however, whether the clock is progressing, i.e., oscillating, normally
469 within human tumors has been unclear. Here we developed a simple method to detect the
470 signature of clock progression based on the co-expression of a small set of genes. By applying
471 the method to publicly available cancer transcriptome data, we discovered that clock progression
472 may be perturbed in multiple types of human cancer. Our analysis also suggests that
473 dysregulation of clock progression in cancer is not caused solely by inactivation of core clock
474 genes and is accompanied by large-scale changes in circadian gene expression and co-
475 expression.

476 Our approach relies on three principles. First, we use prior knowledge of clock genes and clock-
477 controlled genes. Second, we account for the fact that clock progression is defined not by a
478 static condition, but by a dynamic cycle. Our approach thus exploits the co-expression of clock
479 genes that arises from (1) different genes having rhythms with different circadian phases and (2)
480 different samples being taken from different points in the cycle. Finally, our method does not
481 attempt to infer an oscillatory pattern, but instead uses only the statistical correlations in
482 expression between pairs of genes. The underlying assumption of the ΔCCD is that
483 perturbations to the clock will alter the relative phases and/or signal-to-noise ratios of rhythms in
484 clock gene expression and thereby alter the correlations. Although the correlation matrix
485 captures only part of the complex relationship between genes, it is intuitive, simple to calculate,

486   and requires relatively few samples (our results indicate that 30 in humans may be sufficient).
487   Altogether, these principles enable our method to detect perturbed clock progression in groups
488   of samples whose times of acquisition are unknown and whose coverage of the 24-h cycle is
489   incomplete. Consequently, the CCD and ΔCCD should be a valuable complement to methods
490   designed to infer rhythms in omics data, such as CYCLOPS (Anafi et al., 2017).

491   Despite these advantages, our method does have limitations. First, a large CCD does not
492   exclude the possibility that the clock genes are still rhythmic. Instead, it implies that if rhythms
493   are present, the phase relationships between genes are greatly altered relative to a normally
494   progressing clock. Similarly, despite altered expression and co-expression levels, some of the
495   genes rhythmic in healthy lung may still be rhythmic in lung cancer. Second, clock gene co-
496   expression is insensitive to the alignment of the circadian clock to time of day, and so cannot
497   detect a phase difference between conditions, such as that recently observed during manic
498   episodes of bipolar disorder (Moon et al., 2016). This limitation, however, allowed us to readily
499   compare clock gene co-expression between mice and humans, despite the circadian phase
500   difference between the two species (Hughey & Butte, 2016). Third, as shown by our analysis of
501   clock gene knockouts in mice, larger sample sizes are required to find differences in co-
502   expression than to find differences in rhythms (the former does not require time of day
503   information, the latter does). This is partly because differences in co-expression are invariant to
504   the relative levels of gene expression between conditions, which is why we also analyzed
505   differential expression and differential variability. Fourth, transcription is only one facet of the
506   core clock mechanism, and perturbations to post-translational modification or degradation of
507   clock proteins (if unaccompanied by changes in clock gene expression) would not be detected
508   by our approach. Finally, because co-expression is calculated for a group of samples, our
509   approach does not immediately lend itself to assessing clock progression in single samples.
510   Thus, although our results support dysregulated clock progression in tumors as a group, there
511   may be some tumors in which the clock is progressing normally. In the future, it may be possible
512   to complement the ΔCCD metric and infer clock progression in a given individual by directly
513   comparing matched healthy and disease samples.

514   In healthy tissues in vivo, the circadian clocks of individual cells are entrained and oscillating
515   together, which allows bulk measurements to contain robust circadian signals. Consequently, the
516   loss of a circadian signature in human tumors could result from dysfunction in either
517   entrainment, the oscillator, or both. Dysfunction in entrainment would imply that the clocks in at
518   least some of the cancer cells are out of sync with each other and therefore free running, i.e.
519   ignoring zeitgeber signals. Dysfunction in the oscillator would imply that the clocks in at least
520   some of the cancer cells are no longer progressing normally. Given the current data, in which
521   expression values are the averages across many cells, these scenarios cannot be distinguished.
522   However, the moderate correlation between ΔCCD and tumor purity across cancer types leads
523   us to speculate that the clocks in stromal and/or infiltrating immune cells may be progressing
524   normally. Notably, if clock gene co-expression in tumors were normal, then the moderate
525   increases in expression variability of clock genes could imply rhythms (synchronized across the
526   tumor) of increased amplitude. Because clock gene co-expression in tumors is strongly
527   perturbed, however, we find this scenario unlikely. We find it more likely that increased gene
528   expression variability in tumors is a result of non-circadian sources of variation, e.g., tumor grade
529   and subtype. In the future, the specific mechanisms of perturbed clock progression may be

530 unraveled through a combination of mathematical modeling (Kim & Forger, 2012; Lück et al.,
531 2014) and single-cell measurements. A separate matter not addressed here is how cancer
532 influences circadian rhythms in the rest of the body (Masri et al., 2016), which may be relevant
533 for optimizing the daily timing of anticancer treatments (Dulong et al., 2015).

534 An apparent contradiction to perturbed clock progression in tumors is the presence of normal
535 rhythms in U2OS cells (which are derived from an osteosarcoma) and some other cancer cell
536 lines (Relógio et al., 2014), as well as in human and mouse glioblastoma cells in vitro (Slat et al.,
537 2017). One explanation for this contradiction is that there could be a critical difference, in terms
538 of circadian clock progression, between tumors in vivo and cancer cells in vitro. If so, we would
539 speculate that the dysregulation of clock progression in some human tumors may, under certain
540 conditions, be reversible.

541 Based on this study alone, which is observational, it is not possible to determine the direction of
542 causality between tumor development and clock dysregulation, or to determine whether clock
543 dysregulation provides a selective advantage. Indeed, perturbations to co-expression in cancer
544 are not limited to clock genes or to rhythmic genes, as previous work has found changes in co-
545 expression and connectivity across the transcriptome (West et al., 2012; Anglani et al., 2014).
546 However, given the clock's established role in regulating metabolism and a recent finding that
547 stimulation of the clock inhibits tumor growth in melanoma (Kiessling et al., 2017), our findings
548 raise the possibility that clock dysregulation and manipulation of normal circadian physiology
549 may be a cancer driver in multiple solid tissues. On the other hand, a functional circadian clock
550 seems to be required for growth of acute myeloid leukemia cells (Puram et al., 2016), so further
551 work is necessary to clarify this issue.

Conclusions

553 Although the effect of restoring normal clock progression remains to be tested at the pre-clinical
554 and clinical stages, our findings raise the conjecture that this could be a viable treatment
555 strategy in a wide range of cancer types. Recent work on the effects of natural light exposure
556 (Stothard et al., 2017) and time-restricted feeding (Hatori et al., 2012; Chaix et al., 2014)
557 suggests that such a treatment strategy would not necessarily have to be pharmacological.
558 Finally, given the practical challenges of studying circadian rhythms at the cellular level in
559 humans, our method offers the possibility to infer clock progression in diverse human
560 phenotypes using publicly available transcriptome data.

Supplemental Information

562 Fig. S1. Clock gene co-expression for the eight mouse datasets used to build the reference.
563 Fig. S2. Estimated times of peak expression for the 12 clock genes in healthy mouse organs.
564 Fig. S3. Clock correlation distances between each of the eight mouse datasets used to build the
565 reference.
566 Fig. S4. Clock gene co-expression in mice subject to daytime feeding.
567 Fig. S5. Clock gene co-expression for samples collected only during the day or night.
568 Fig. S6. Clock gene co-expression in labeled human datasets not shown in Fig. 1D.
569 Fig. S7. Strength of clock gene co-expression for mouse and human datasets.

588   References

589   Altman BJ., Hsieh AL., Sengupta A., Krishnanaiah SY., Stine ZE., Walton ZE., Gouw AM.,

590   Venkataraman A., Li B., Goraksha-Hicks P., Diskin SJ., Bellovin DI., Simon MC., Rathmell

591   JC., Lazar MA., Maris JM., Felsher DW., Hogenesch JB., Weljie AM., Dang CV. 2015. MYC

592   Disrupts the Circadian Clock and Metabolism in Cancer Cells. *Cell metabolism* 22:1009–

593   1019.

594   Anafi RC., Francey LJ., Hogenesch JB., Kim J. 2017. CYCLOPS reveals human transcriptional

595   rhythms in health and disease. *Proceedings of the National Academy of Sciences of the*

596   *United States of America*. DOI: 10.1073/pnas.1619320114.

597   Anglani R., Creanza TM., Liuzzi VC., Piepoli A., Panza A., Andriulli A., Ancona N. 2014. Loss of

598   connectivity in cancer co-expression networks. *PloS one* 9:e87075.

599   Aran D., Sirota M., Butte AJ. 2015. Systematic pan-cancer analysis of tumour purity. *Nature*

600   *communications* 6:8971.

601   Archer SN., Laing EE., Möller-Levet CS., Veen DR van der., Bucca G., Lazar AS., Santhi N.,

602   Slak A., Kabiljo R., Schantz M von., Smith CP., Dijk D-J. 2014. Mistimed sleep disrupts

603   circadian regulation of the human transcriptome. *Proc Natl Acad Sci USA* 111:E682–E691.

604   Arnardottir ES., Nikonova EV., Shockley KR., Podtelezhnikov AA., Anafi RC., Tanis KQ., Maislin

605   G., Stone DJ., Renger JJ., Winrow CJ., Pack AI. 2014. Blood-Gene Expression Reveals

606   Reduced Circadian Rhythmicity in Individuals Resistant to Sleep Deprivation. *Sleep* 37:589–

607   600.

608   Asher G., Reinke H., Altmeyer M., Gutierrez-Arcelus M., Hottiger MO., Schibler U. 2010.

609   Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to

610   feeding. *Cell* 142:943–953.

611   Balsalobre A., Damiola F., Schibler U. 1998. A serum shock induces circadian gene expression

612   in mammalian tissue culture cells. *Cell* 93:929–937.

613   Bieler J., Cannavo R., Gustafson K., Gobet C., Gatfield D., Naef F. 2014. Robust

614   synchronization of coupled circadian and cell cycle oscillators in single mammalian cells.

615   *Molecular systems biology* 10:739.

616   Bonder MJ., Kasela S., Kals M., Tamm R., Lokk K., Barragan I., Buurman WA., Deelen P., Greve

617   J-W., Ivanov M., Rensen SS., van Vliet-Ostaptchouk JV., Wolfs MG., Fu J., Hofker MH.,

618   Wijmenga C., Zhernakova A., Ingelman-Sundberg M., Franke L., Milani L. 2014. Genetic

619   and epigenetic regulation of gene expression in fetal and adult human livers. *BMC*

620   *genomics* 15:860.

621   Bossé Y., Postma DS., Sin DD., Lamontagne M., Couture C., Gaudreault N., Joubert P., Wong

622   V., Elliott M., van den Berge M., Brandsma CA., Tribouley C., Malkov V., Tsou JA., Opiteck

623   GJ., Hogg JC., Sandford AJ., Timens W., Paré PD., Laviolette M. 2012. Molecular signature

624   of smoking in human lung tissues. *Cancer research* 72:3753–3763.

625   Cadenas C., Sandt L van de., Edlund K., Lohr M., Hellwig B., Marchan R., Schmidt M.,

626   Rahnenführer J., Oster H., Hengstler JG. 2014. Loss of circadian clock gene expression is

627   associated with tumor progression in breast cancer. *Cell cycle* 13:3282–3291.

628   Chaix A., Zarrinpar A., Miu P., Panda S. 2014. Time-restricted feeding is a preventative and

629   therapeutic intervention against diverse nutritional challenges. *Cell metabolism* 20:991–

630   1005.

631   Chen C-Y., Logan RW., Ma T., Lewis DA., Tseng GC., Sibille E., McClung CA. 2016. Effects of

632   aging on circadian patterns of gene expression in the human prefrontal cortex. *Proc Natl*

633   *Acad Sci USA* 113:206–211.

634   Cho H., Zhao X., Hatori M., Yu RT., Barish GD., Lam MT., Chong L-W., DiTacchio L., Atkins AR.,

635   Glass CK., Liddle C., Auwerx J., Downes M., Panda S., Evans RM. 2012. Regulation of

636   circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β. *Nature* 485:123–127.

637   Dai M., Wang P., Boyd AD., Kostov G., Athey B., Jones EG., Bunney WE., Myers RM., Speed

638   TP., Akil H., Watson SJ., Meng F. 2005. Evolving gene/transcript definitions significantly alter

639   the interpretation of GeneChip data. *Nucleic acids research* 33:e175.

640   Damiola F., Minh NL., Preitner N., Kornmann B., Fleury-Olela F., Schibler U. 2000. Restricted

641   feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in

642   the suprachiasmatic nucleus. *Genes & development* 14:2950–2961.

643   Dibner C., Schibler U., Albrecht U. 2010. The Mammalian Circadian Timing System:

644   Organization and Coordination of Central and Peripheral Clocks. *Annual review of*

645   *physiology* 72:517–549.

646   Dudek M., Gossan N., Yang N., Im H-J., Ruckshanthi JPD., Yoshitane H., Li X., Jin D., Wang P.,

647   Boudiffa M., Bellantuono I., Fukada Y., Boot-Handford RP., Meng Q-J. 2016. The

648   chondrocyte clock gene Bmal1 controls cartilage homeostasis and integrity. *The Journal of*

649   *clinical investigation* 126:365–376.

650   Dulong S., Ballesta A., Okyar A., Lévi F. 2015. Identification of Circadian Determinants of Cancer

651   Chronotherapy through In Vitro Chronopharmacology and Mathematical Modeling.

652   *Molecular cancer therapeutics* 14:2154–2164.

653   Dyar KA., Ciciliot S., Wright LE., Biensø RS., Tagliazucchi GM., Patel VR., Forcato M., Paz MIP.,

654   Gudiksen A., Solagna F., Albiero M., Moretti I., Eckel-Mahan KL., Baldi P., Sassone-Corsi P.,

655   Rizzuto R., Bicciato S., Pilegaard H., Blaauw B., Schiaffino S. 2014. Muscle insulin

656   sensitivity and glucose metabolism are controlled by the intrinsic muscle clock. *Mol Metab*

657   3:29–41.

658   Eckel-Mahan KL., Patel VR., de Mateo S., Orozco-Solis R., Ceglia NJ., Sahar S., Dilag-Penilla

659   SA., Dyar KA., Baldi P., Sassone-Corsi P. 2013. Reprogramming of the circadian clock by

660   nutritional challenge. *Cell* 155:1464–1478.

661   El-Athman R., Genov NN., Mazuch J., Zhang K., Yu Y., Fuhr L., Abreu M., Li Y., Wallach T.,

662   Kramer A., Schmitt CA., Relógio A. 2017. The Ink4a/Arf locus operates as a regulator of the

663   circadian clock modulating RAS activity. *PLoS biology* 15:e2002940.

664   Feillet C., Krusche P., Tamanini F., Janssens RC., Downey MJ., Martin P., Teboul M., Saito S.,

665   Lévi FA., Bretschneider T., van der Horst GTJ., Delaunay F., Rand DA. 2014. Phase locking

666   and multiple oscillating attractors for the coupled mammalian clock and cell cycle.

667   *Proceedings of the National Academy of Sciences of the United States of America*

668   111:9828–9833.

669   Geyfman M., Kumar V., Liu Q., Ruiz R., Gordon W., Espitia F., Cam E., Millar SE., Smyth P.,

670   Ihler A., Takahashi JS., Andersen B. 2012. Brain and muscle Arnt-like protein-1 (BMAL1)

671   controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the

672   epidermis. *Proceedings of the National Academy of Sciences of the United States of*

673   *America* 109:11758–11763.

674   Gréchez-Cassiau A., Rayet B., Guillaumond F., Teboul M., Delaunay F. 2008. The circadian

675   clock component BMAL1 is a critical regulator of p21WAF1/CIP1 expression and

676   hepatocyte proliferation. *The Journal of biological chemistry* 283:4535–4542.

677   Grundberg E., Small KS., Hedman ÅK., Nica AC., Buil A., Keildson S., Bell JT., Yang T-P.,

678   Meduri E., Barrett A., Nisbett J., Sekowska M., Wilk A., Shin S-Y., Glass D., Travers M., Min

679   JL., Ring S., Ho K., Thorleifsson G., Kong A., Thorsteindottir U., Ainali C., Dimas AS.,

680   Hassanali N., Ingle C., Knowles D., Krestyaninova M., Lowe CE., Di Meglio P., Montgomery

681   SB., Parts L., Potter S., Surdulescu G., Tsaprouni L., Tsoka S., Bataille V., Durbin R., Nestle

682   FO., O'Rahilly S., Soranzo N., Lindgren CM., Zondervan KT., Ahmadi KR., Schadt EE.,

683   Stefansson K., Smith GD., McCarthy MI., Deloukas P., Dermitzakis ET., Spector TD.,

684   Multiple Tissue Human Expression Resource (MuTHER) Consortium. 2012. Mapping cis-

685   and trans-regulatory effects across multiple tissues in twins. *Nature genetics* 44:1084–1089.

686   Guerrero-Vargas NN., Navarro-Espíndola R., Guzmán-Ruíz MA., Basualdo MDC., Espitia-

687   Bautista E., López-Bago A., Lascurain R., Córdoba-Manilla C., Buijs RM., Escobar C. 2017.

688   Circadian disruption promotes tumor growth by anabolic host metabolism; experimental

689   evidence in a rat model. *BMC cancer* 17:625.

690   Haspel JA., Chettimada S., Shaik RS., Chu J-H., Raby BA., Cernadas M., Carey V., Process V.,

691   Hunninghake GM., Ifedigbo E., Lederer JA., Englert J., Pelton A., Coronata A., Fredenburgh

692   LE., Choi AMK. 2014. Circadian rhythm reprogramming during lung inflammation. *Nature*

693   *communications* 5:4753.

694   Hatori M., Vollmers C., Zarrinpar A., DiTacchio L., Bushong EA., Gill S., Leblanc M., Chaix A.,

695   Joens M., Fitzpatrick JAJ., Ellisman MH., Panda S. 2012. Time-restricted feeding without

696   reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. *Cell*

697   *metabolism* 15:848–860.

698   Hedges LV., Olkin I. 1985. *Statistical Methods for Meta-Analysis*. Academic Press.

699   Helwig NE., Ma P. 2014. Fast and stable multiple smoothing parameter selection in smoothing

700   spline analysis of variance models with large samples. *Journal of computational and*

701   *graphical statistics: a joint publication of American Statistical Association, Institute of*

702   *Mathematical Statistics, Interface Foundation of North America* 24:715–732.

703   Hoffmann J., Symul L., Shostak A., Fischer T., Naef F., Brunner M. 2014. Non-circadian

704   expression masking clock-driven weak transcription rhythms in U2OS cells. *PloS one*

705   9:e102238.

706   Hoogerwerf WA., Sinha M., Conesa A., Luxon BA., Shahinian VB., Cornélissen G., Halberg F.,

707   Bostwick J., Timm J., Cassone VM. 2008. Transcriptional profiling of mRNA expression in

708   the mouse distal colon. *Gastroenterology* 135:2019–2029.

709   Hou J., Aerts J., den Hamer B., van Ijcken W., den Bakker M., Riegman P., van der Leest C., van

710   der Spek P., Foekens JA., Hoogsteden HC., Grosveld F., Philipsen S. 2010. Gene

711   expression-based classification of non-small cell lung carcinomas and survival prediction.

712   *PloS one* 5:e10312.

713   Huang DW., Sherman BT., Lempicki RA. 2009. Systematic and integrative analysis of large gene

714   lists using DAVID bioinformatics resources. *Nature protocols* 4:44–57.

715   Hughes ME., DiTacchio L., Hayes KR., Vollmers C., Pulivarthy S., Baggs JE., Panda S.,

716   Hogenesch JB. 2009. Harmonics of Circadian Gene Transcription in Mammals. *PLoS*

717   *genetics* 5:e1000442.

718   Hughey JJ. 2017. Machine learning identifies a compact gene set for monitoring the circadian

719   clock in human blood. *Genome medicine* 9:19.

720   Hughey JJ., Butte AJ. 2015. Robust meta-analysis of gene expression using the elastic net.

721   *Nucleic acids research* 43:e79.

722   Hughey JJ., Butte AJ. 2016. Differential Phasing between Circadian Clocks in the Brain and

723   Peripheral Organs in Humans. *Journal of biological rhythms* 31:588–597.

724   Hughey JJ., Hastie T., Butte AJ. 2016. ZeitZeiger: supervised learning for high-dimensional data

725   from an oscillatory system. *Nucleic acids research* 44:e80.

726   Innocenti F., Cooper GM., Stanaway IB., Gamazon ER., Smith JD., Mirkov S., Ramirez J., Liu

727   W., Lin YS., Moloney C., Aldred SF., Trinklein ND., Schuetz E., Nickerson DA., Thummel

728   KE., Rieder MJ., Rettie AE., Ratain MJ., Cox NJ., Brown CD. 2011. Identification,

729   replication, and functional fine-mapping of expression quantitative trait loci in primary human

730   liver tissue. *PLoS genetics* 7:e1002078.

731   Irizarry RA., Hobbs B., Collin F., Beazer-Barclay YD., Antonellis KJ., Scherf U., Speed TP. 2003.

732   Exploration, normalization, and summaries of high density oligonucleotide array probe level

733   data. *Biostatistics* 4:249–264.

734   Janich P., Toufighi K., Solanas G., Luis NM., Minkwitz S., Serrano L., Lehner B., Benitah SA.

735   2013. Human epidermal stem cell function is regulated by circadian oscillations. *Cell stem*

736   *cell* 13:745–753.

737   Johnson WE., Li C., Rabinovic A. 2007. Adjusting batch effects in microarray expression data

738   using empirical Bayes methods. *Biostatistics* 8:118–127.

739   Kettner NM., Voicu H., Finegold MJ., Coarfa C., Sreekumar A., Putluri N., Katchy CA., Lee C.,

740   Moore DD., Fu L. 2016. Circadian Homeostasis of Liver Metabolism Suppresses

741   Hepatocarcinogenesis. *Cancer cell*. DOI: 10.1016/j.ccell.2016.10.007.

742   Kiessling S., Beaulieu-Laroche L., Blum ID., Landgraf D., Welsh DK., Storch K-F., Labrecque N.,

743   Cermakian N. 2017. Enhancing circadian clock function in cancer cells inhibits tumor

744   growth. *BMC biology* 15:13.

745   Kim JK., Forger DB. 2012. A mechanism for robust circadian timekeeping via stoichiometric

746   balance. *Molecular systems biology* 8:630.

747   Laing EE., Möller-Levet CS., Poh N., Santhi N., Archer SN., Dijk D-J. 2017. Blood transcriptome

748   based biomarkers for human circadian phase. *eLife* 6. DOI: 10.7554/eLife.20214.

749   Lamb JR., Zhang C., Xie T., Wang K., Zhang B., Hao K., Chudin E., Fraser HB., Millstein J.,

750   Ferguson M., Suver C., Ivanovska I., Scott M., Philippar U., Bansal D., Zhang Z., Burchard

751   J., Smith R., Greenawalt D., Cleary M., Derry J., Loboda A., Watters J., Poon RTP., Fan ST.,

752   Yeung C., Lee NPY., Guinney J., Molony C., Emilsson V., Buser-Doepner C., Zhu J., Friend

753   S., Mao M., Shaw PM., Dai H., Luk JM., Schadt EE. 2011. Predictive genes in adjacent

754   normal tissue are preferentially altered by sCNV during tumorigenesis in liver cancer and

755   may rate limiting. *PloS one* 6:e20090.

756   Lamia KA. 2017. Ticking time bombs: connections between circadian clocks and cancer.

757   *F1000Research* 6:1910.

758   Landi MT., Dracheva T., Rotunno M., Figueroa JD., Liu H., Dasgupta A., Mann FE., Fukuoka J.,

759   Hames M., Bergen AW., Murphy SE., Yang P., Pesatori AC., Consonni D., Bertazzi PA.,

760   Wacholder S., Shih JH., Caporaso NE., Jen J. 2008. Gene expression signature of cigarette

761   smoking and its role in lung adenocarcinoma development and survival. *PloS one* 3:e1651.

762   Langfelder P., Horvath S. 2008. WGCNA: an R package for weighted correlation network

763   analysis. *BMC bioinformatics* 9:559.

764   Law CW., Chen Y., Shi W., Smyth GK. 2014. voom: Precision weights unlock linear model

765   analysis tools for RNA-seq read counts. *Genome biology* 15:R29.

766 Leng N., Chu L-F., Barry C., Li Y., Choi J., Li X., Jiang P., Stewart RM., Thomson JA.,

767   Kendziorski C. 2015. Oscope identifies oscillatory genes in unsynchronized single-cell RNA-

768   seq experiments. *Nature methods* 12:947–950.

769 Liberzon A., Subramanian A., Pinchback R., Thorvaldsdóttir H., Tamayo P., Mesirov JP. 2011.

770   Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27:1739–1740.

771 Li JZ., Bunney BG., Meng F., Hagenauer MH., Walsh DM., Vawter MP., Evans SJ., Choudary

772   PV., Cartagena P., Barchas JD., Schatzberg AF., Jones EG., Myers RM., Watson SJ., Akil

773   H., Bunney WE. 2013. Circadian patterns of gene expression in the human brain and

774   disruption in major depressive disorder. *Proc Natl Acad Sci USA* 110:9950–9955.

775 Lim H-Y., Sohn I., Deng S., Lee J., Jung SH., Mao M., Xu J., Wang K., Shi S., Joh JW., Choi YL.,

776   Park C-K. 2013. Prediction of disease-free survival in hepatocellular carcinoma by gene

777   expression profiling. *Annals of surgical oncology* 20:3747–3753.

778 Lück S., Thurley K., Thaben PF., Westermark PO. 2014. Rhythmic degradation explains and

779   unifies circadian transcriptome and proteome data. *Cell reports* 9:741–751.

780 Lu T-P., Tsai M-H., Lee J-M., Hsu C-P., Chen P-C., Lin C-W., Shih J-Y., Yang P-C., Hsiao CK.,

781   Lai L-C., Chuang EY. 2010. Identification of a novel biomarker, SEMA5A, for non-small cell

782   lung carcinoma in nonsmoking women. *Cancer epidemiology, biomarkers & prevention: a*

783   *publication of the American Association for Cancer Research, cosponsored by the American*

784   *Society of Preventive Oncology* 19:2590–2597.

785 Maier B., Wendt S., Vanselow JT., Wallach T., Reischl S., Oehmke S., Schlosser A., Kramer A.

786   2009. A large-scale functional RNAi screen reveals a role for CK2 in the mammalian

787   circadian clock. *Genes & development* 23:708–718.

788 Masri S., Papagiannakopoulos T., Kinouchi K., Liu Y., Cervantes M., Baldi P., Jacks T., Sassone-

789   Corsi P. 2016. Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis. *Cell*

790   165:896–909.

791 Matsuo T., Yamaguchi S., Mitsui S., Emi A., Shimoda F., Okamura H. 2003. Control mechanism

792   of the circadian clock for timing of cell division in vivo. *Science* 302:255–259.

793   Matsu-Ura T., Dovzhenok A., Aihara E., Rood J., Le H., Ren Y., Rosselot AE., Zhang T., Lee C.,

794   Obrietan K., Montrose MH., Lim S., Moore SR., Hong CI. 2016. Intercellular Coupling of the

795   Cell Cycle and Circadian Clock in Adult Stem Cell Culture. *Molecular cell* 64:900–912.

796   Michael AK., Harvey SL., Sammons PJ., Anderson AP., Kopalle HM., Banham AH., Partch CL.

797   2015. Cancer/Testis Antigen PASD1 Silences the Circadian Clock. *Molecular cell* 58:743–

798   754.

799   Möller-Levet CS., Archer SN., Bucca G., Laing EE., Slak A., Kabiljo R., Lo JCY., Santhi N.,

800   Schantz M von., Smith CP., Dijk D-J. 2013. Effects of insufficient sleep on circadian

801   rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci*

802   *USA* 110:E1132–E1141.

803   Moon J-H., Cho C-H., Son GH., Geum D., Chung S., Kim H., Kang S-G., Park Y-M., Yoon H-K.,

804   Kim L., Jee H-J., An H., Kripke DF., Lee H-J. 2016. Advanced Circadian Phase in Mania and

805   Delayed Circadian Phase in Mixed Mania and Depression Returned to Normal after

806   Treatment of Bipolar Disorder. *EBioMedicine* 11:285–295.

807   Nakahata Y., Sahar S., Astarita G., Kaluzova M., Sassone-Corsi P. 2009. Circadian control of the

808   NAD+ salvage pathway by CLOCK-SIRT1. *Science* 324:654–657.

809   Negoro H., Kanematsu A., Doi M., Suadicani SO., Matsuo M., Imamura M., Okinami T.,

810   Nishikawa N., Oura T., Matsui S., Seo K., Tainaka M., Urabe S., Kiyokage E., Todo T.,

811   Okamura H., Tabata Y., Ogawa O. 2012. Involvement of urinary bladder Connexin43 and

812   the circadian clock in coordination of diurnal micturition rhythm. *Nature communications*

813   3:809.

814   Neufeld-Cohen A., Robles MS., Aviram R., Manella G., Adamovich Y., Ladeuix B., Nir D.,

815   Rousso-Noori L., Kuperman Y., Golik M., Mann M., Asher G. 2016. Circadian control of

816   oscillations in mitochondrial rate-limiting enzymes and nutrient utilization by PERIOD

817   proteins. *Proceedings of the National Academy of Sciences of the United States of America*

818   113:E1673–82.

819 Nikolaeva S., Pradervand S., Centeno G., Zavadova V., Tokonami N., Maillard M., Bonny O.,

820   Firsov D. 2012. The circadian clock modulates renal sodium handling. *Journal of the*

821   *American Society of Nephrology: JASN* 23:1019–1026.

822 Oster H., Damerow S., Hut RA., Eichele G. 2006. Transcriptional profiling in the adrenal gland

823   reveals circadian regulation of hormone biosynthesis genes and nucleosome assembly

824   genes. *Journal of biological rhythms* 21:350–361.

825 Papagiannakopoulos T., Bauer MR., Davidson SM., Heimann M., Subbaraj L., Bhutkar A.,

826   Bartlebaugh J., Vander Heiden MG., Jacks T. 2016. Circadian Rhythm Disruption Promotes

827   Lung Tumorigenesis. *Cell metabolism* 24:324–331.

828 Partch CL., Green CB., Takahashi JS. 2014. Molecular architecture of the mammalian circadian

829   clock. *Trends in cell biology* 24:90–99.

830 Paschos GK., Ibrahim S., Song W-L., Kunieda T., Grant G., Reyes TM., Bradfield CA., Vaughan

831   CH., Eiden M., Masoodi M., Griffin JL., Wang F., Lawson JA., Fitzgerald GA. 2012. Obesity

832   in mice with adipocyte-specific deletion of clock component Arntl. *Nature medicine* 18:1768–

833   1777.

834 Patro R., Duggal G., Love MI., Irizarry RA., Kingsford C. 2017. Salmon provides fast and bias-

835   aware quantification of transcript expression. *Nature methods* 14:417–419.

836 Phipson B., Smyth GK. 2010. Permutation P-values should never be zero: calculating exact P-

837   values when permutations are randomly drawn. *Statistical applications in genetics and*

838   *molecular biology* 9:Article39.

839 Puram RV., Kowalczyk MS., de Boer CG., Schneider RK., Miller PG., McConkey M., Tothova Z.,

840   Tejero H., Heckl D., Järås M., Chen MC., Li H., Tamayo A., Cowley GS., Rozenblatt-Rosen

841   O., Al-Shahrour F., Regev A., Ebert BL. 2016. Core Circadian Clock Genes Regulate

842   Leukemia Stem Cells in AML. *Cell* 165:303–316.

843 Rahman M., Jackson LK., Johnson WE., Li DY., Bild AH., Piccolo SR. 2015. Alternative

844   preprocessing of RNA-Sequencing data in The Cancer Genome Atlas leads to improved

845   analysis results. *Bioinformatics* 31:3666–3672.

846   Relógio A., Thomas P., Medina-Pérez P., Reischl S., Bervoets S., Gloc E., Riemer P., Mang-

847   Fatehi S., Maier B., Schäfer R., Leser U., Herzel H., Kramer A., Sers C. 2014. Ras-

848   mediated deregulation of the circadian clock in cancer. *PLoS genetics* 10:e1004338.

849   Ritchie SC., Watts S., Fearnley LG., Holt KE., Abraham G., Inouye M. 2016. A Scalable

850   Permutation Approach Reveals Replication and Preservation Patterns of Network Modules

851   in Large Datasets. *Cell systems* 3:71–82.

852   Roessler S., Jia H-L., Budhu A., Forgues M., Ye Q-H., Lee J-S., Thorgeirsson SS., Sun Z., Tang

853   Z-Y., Qin L-X., Wang XW. 2010. A unique metastasis gene signature enables prediction of

854   tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer research* 70:10202–

855   10212.

856   Rossetti S., Esposito J., Corlazzoli F., Gregorski A., Sacchi N. 2012. Entrainment of breast

857   (cancer) epithelial cells detects distinct circadian oscillation patterns for clock and hormone

858   receptor genes. *Cell cycle* 11:350–360.

859   Schadt EE., Molony C., Chudin E., Hao K., Yang X., Lum PY., Kasarskis A., Zhang B., Wang S.,

860   Suver C., Zhu J., Millstein J., Sieberts S., Lamb J., GuhaThakurta D., Derry J., Storey JD.,

861   Avila-Campillo I., Kruger MJ., Johnson JM., Rohl CA., van Nas A., Mehrabian M., Drake TA.,

862   Lusis AJ., Smith RC., Guengerich FP., Strom SC., Schuetz E., Rushmore TH., Ulrich R.

863   2008. Mapping the genetic architecture of gene expression in human liver. *PLoS biology*

864   6:e107.

865   Schernhammer ES., Feskanich D., Liang G., Han J. 2013. Rotating night-shift work and lung

866   cancer risk among female nurses in the United States. *American journal of epidemiology*

867   178:1434–1441.

868   Schernhammer ES., Laden F., Speizer FE., Willett WC., Hunter DJ., Kawachi I., Colditz GA.

869   2001. Rotating night shifts and risk of breast cancer in women participating in the nurses'

870   health study. *Journal of the National Cancer Institute* 93:1563–1568.

871   Schernhammer ES., Laden F., Speizer FE., Willett WC., Hunter DJ., Kawachi I., Fuchs CS.,

872   Colditz GA. 2003. Night-shift work and risk of colorectal cancer in the nurses' health study.

873   *Journal of the National Cancer Institute* 95:825–828.

874   Selamat SA., Chung BS., Girard L., Zhang W., Zhang Y., Campan M., Siegmund KD., Koss MN.,

875   Hagen JA., Lam WL., Lam S., Gazdar AF., Laird-Offringa IA. 2012. Genome-scale analysis

876   of DNA methylation in lung adenocarcinoma and integration with mRNA expression.

877   *Genome research* 22:1197–1211.

878   Slat EA., Sponagel J., Marpegan L., Simon T., Kfoury N., Kim A., Binz A., Herzog ED., Rubin JB.

879   2017. Cell-intrinsic, Bmal1-dependent Circadian Regulation of Temozolomide Sensitivity in

880   Glioblastoma. *Journal of biological rhythms* 32:121–129.

881   Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential

882   expression in microarray experiments. *Statistical applications in genetics and molecular*

883   *biology* 3:Article3.

884   Song L., Langfelder P., Horvath S. 2012. Comparison of co-expression measures: mutual

885   information, correlation, and model based indices. *BMC bioinformatics* 13:328.

886   Spörl F., Korge S., Jürchott K., Wunderskirchner M., Schellenberg K., Heins S., Specht A., Stoll

887   C., Klemz R., Maier B., Wenck H., Schrader A., Kunz D., Blatt T., Kramer A. 2012. Krüppel-

888   like factor 9 is a circadian transcription factor in human epidermis that controls proliferation

889   of keratinocytes. *Proc Natl Acad Sci USA* 109:10903–10908.

890   Stothard ER., McHill AW., Depner CM., Birks BR., Moehlman TM., Ritchie HK., Guzzetti JR.,

891   Chinoy ED., LeBourgeois MK., Axelsson J., Wright KP Jr. 2017. Circadian Entrainment to

892   the Natural Light-Dark Cycle across Seasons and the Weekend. *Current biology: CB*. DOI:

893   10.1016/j.cub.2016.12.041.

894   Thaben PF., Westermark PO. 2016. Differential rhythmicity: detecting altered rhythmicity in

895   biological data. *Bioinformatics* 32:2800–2808.

896   Van Dycke KCG., Rodenburg W., van Oostrom CTM., van Kerkhof LWM., Pennings JLA.,

897   Roenneberg T., van Steeg H., van der Horst GTJ. 2015. Chronically Alternating Light Cycles

898   Increase Breast Cancer Risk in Mice. *Current biology: CB* 25:1932–1937.

899   Villa E., Critelli R., Lei B., Marzocchi G., Cammà C., Giannelli G., Pontisso P., Cabibbo G., Enea

900   M., Colopi S., Caporali C., Pollicino T., Milosa F., Karampatou A., Todesca P., Bertolini E.,

901   Maccio L., Martinez-Chantar ML., Turola E., Del Buono M., De Maria N., Ballestri S.,

902   Schepis F., Loria P., Enrico Gerunda G., Losi L., Cillo U. 2016. Neoangiogenesis-related

903   genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Results

904   from a prospective study. *Gut* 65:861–869.

905 Vollmers C., Gill S., DiTacchio L., Pulivarthy SR., Le HD., Panda S. 2009. Time of feeding and

906   the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the*

907   *National Academy of Sciences of the United States of America* 106:21453–21458.

908 Vollmers C., Panda S., DiTacchio L. 2008. A high-throughput assay for siRNA-based circadian

909   screens in human U2OS cells. *PloS one* 3:e3457.

910 Wegrzyn LR., Tamimi RM., Rosner BA., Brown SB., Stevens RG., Eliassen AH., Laden F., Willett

911   WC., Hankinson SE., Schernhammer ES. 2017. ROTATING NIGHT SHIFT WORK AND

912   RISK OF BREAST CANCER IN THE NURSES' HEALTH STUDIES. *American journal of*

913   *epidemiology*. DOI: 10.1093/aje/kwx140.

914 West J., Bianconi G., Severini S., Teschendorff AE. 2012. Differential network entropy reveals

915   cancer system hallmarks. *Scientific reports* 2:802.

916 Wu G., Anafi RC., Hughes ME., Kornacker K., Hogenesch JB. 2016. MetaCycle: an integrated R

917   package to evaluate periodicity in large scale data. *Bioinformatics* 32:3351–3353.

918 Xiang S., Mao L., Duplessis T., Yuan L., Dauchy R., Dauchy E., Blask DE., Frasch T., Hill SM.

919   2012. Oscillation of clock and clock controlled genes induced by serum shock in human

920   breast epithelial and breast cancer cells: regulation by melatonin. *Breast cancer: basic and*

921   *clinical research* 6:137–150.

922 Yoo S-H., Yamazaki S., Lowrey PL., Shimomura K., Ko CH., Buhr ED., Siepka SM., Hong H-K.,

923   Oh WJ., Yoo OJ., Menaker M., Takahashi JS. 2004. PERIOD2::LUCIFERASE real-time

924   reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral

925   tissues. *Proceedings of the National Academy of Sciences of the United States of America*

926   101:5339–5346.

927 Young ME., Brewer RA., Peliciari-Garcia RA., Collins HE., He L., Birky TL., Peden BW.,

928   Thompson EG., Ammons B-J., Bray MS., Chatham JC., Wende AR., Yang Q., Chow C-W.,

929   Martino TA., Gamble KL. 2014. Cardiomyocyte-specific BMAL1 plays critical roles in

930   metabolism, signaling, and maintenance of contractile function of the heart. *Journal of*

931   *biological rhythms* 29:257–276.

932 Zhang R., Lahens NF., Ballance HI., Hughes ME., Hogenesch JB. 2014. A circadian gene

933   expression atlas in mammals: implications for biology and medicine. *Proceedings of the*

934   *National Academy of Sciences of the United States of America* 111:16219–16224.

935 Zhang R., Podtelezhnikov AA., Hogenesch JB., Anafi RC. 2016. Discovering Biology in Periodic

936   Data through Phase Set Enrichment Analysis (PSEA). *Journal of biological rhythms* 31:244–

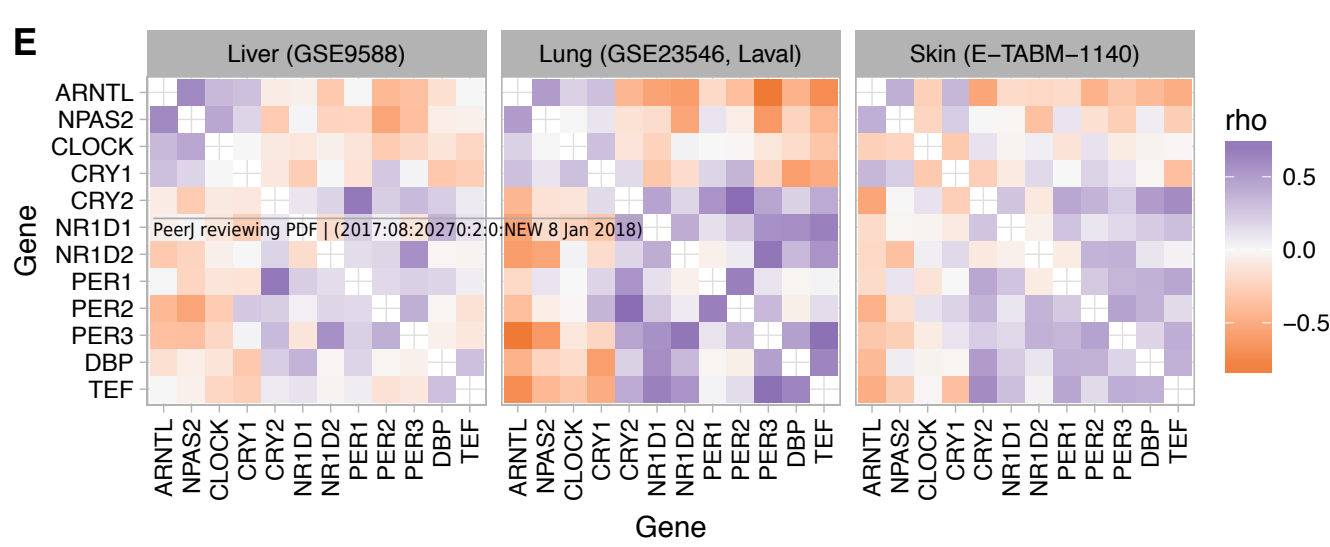
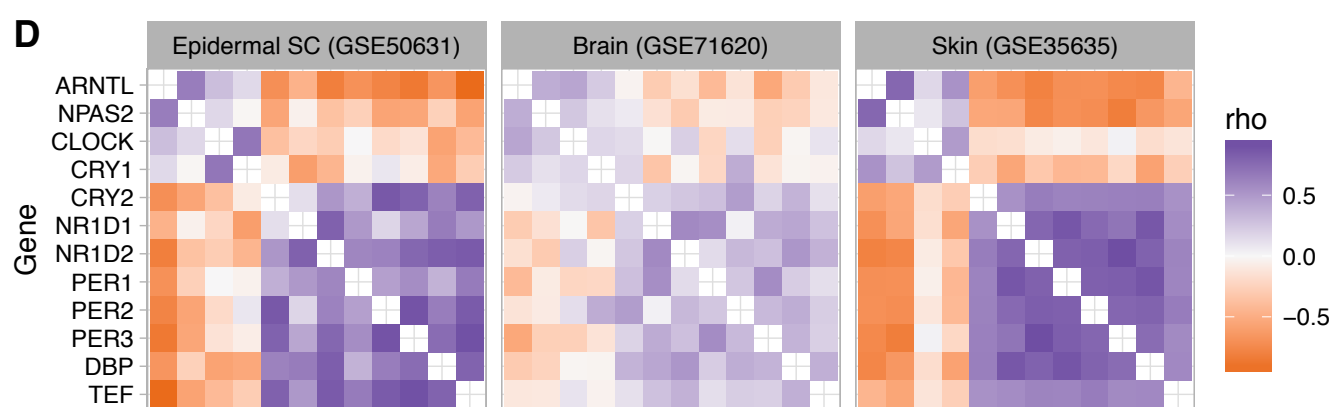
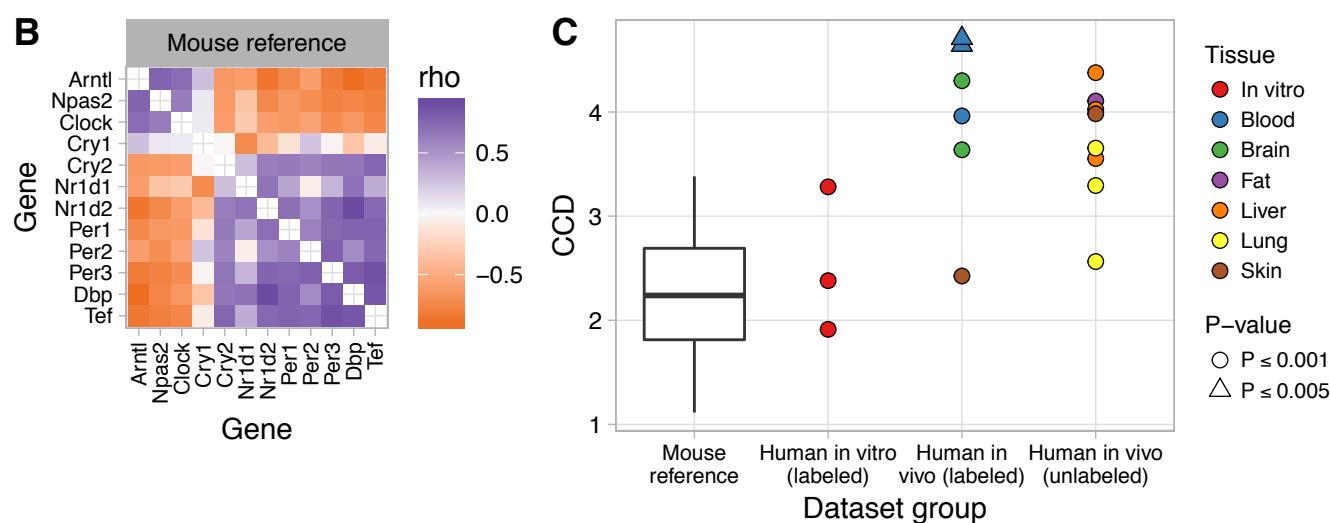
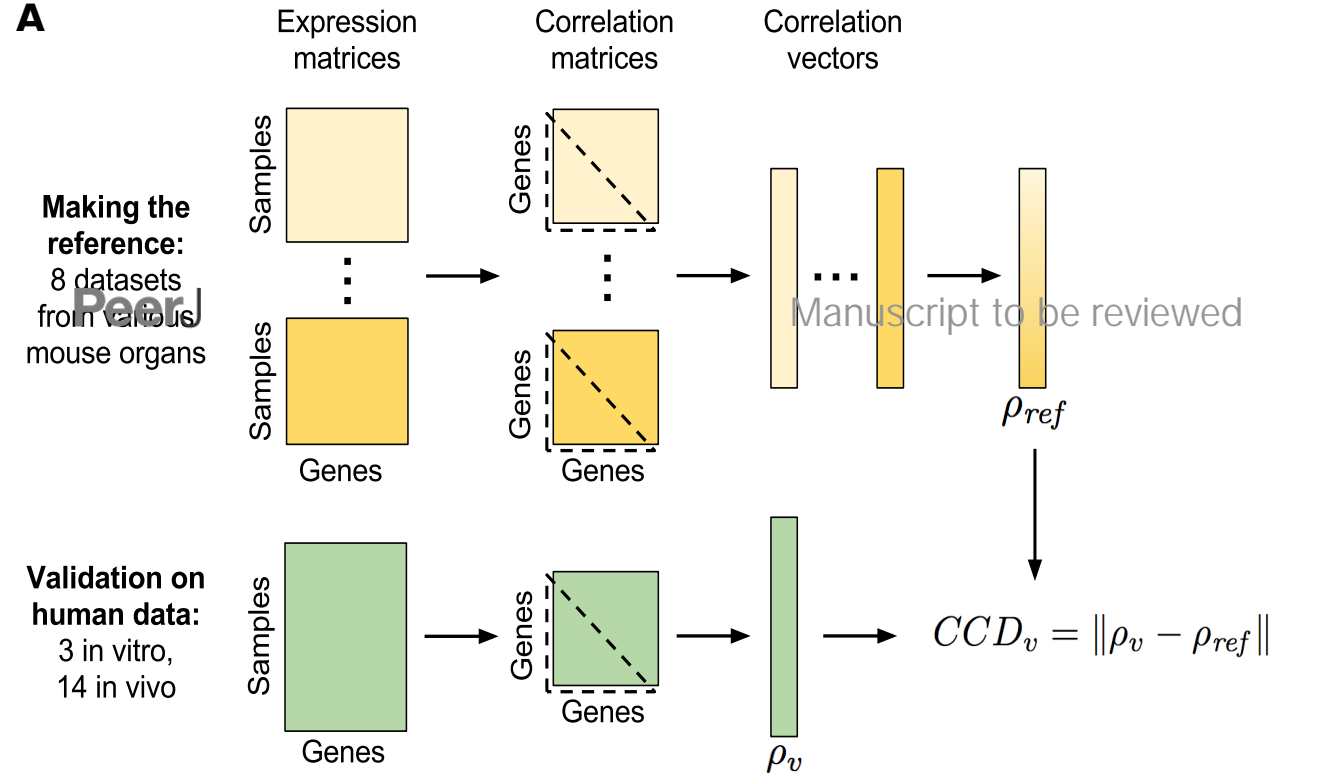
937   257.

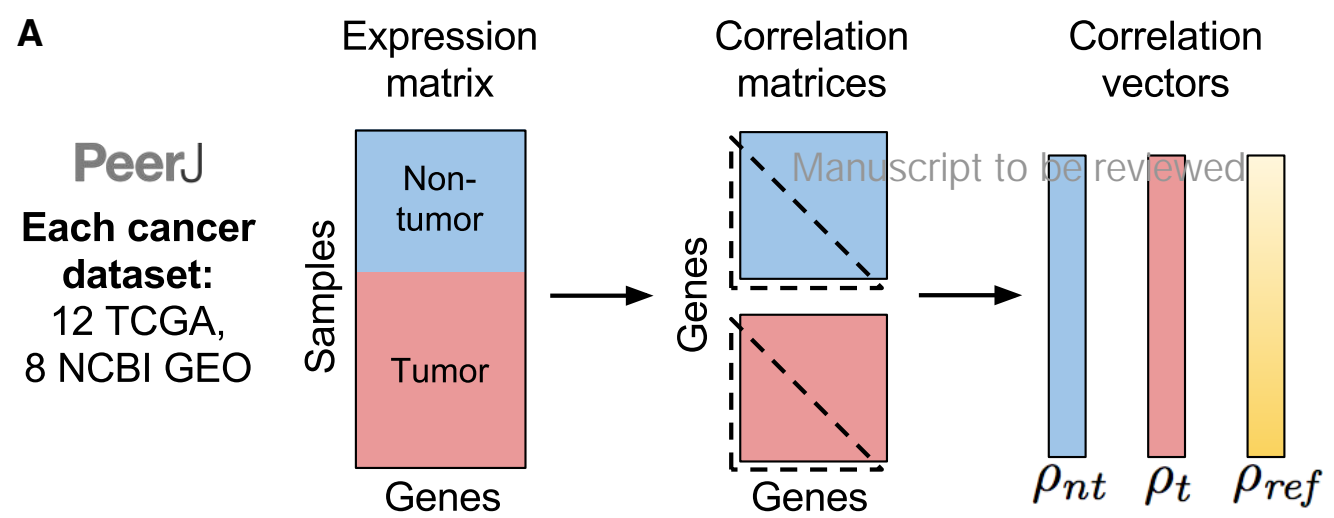
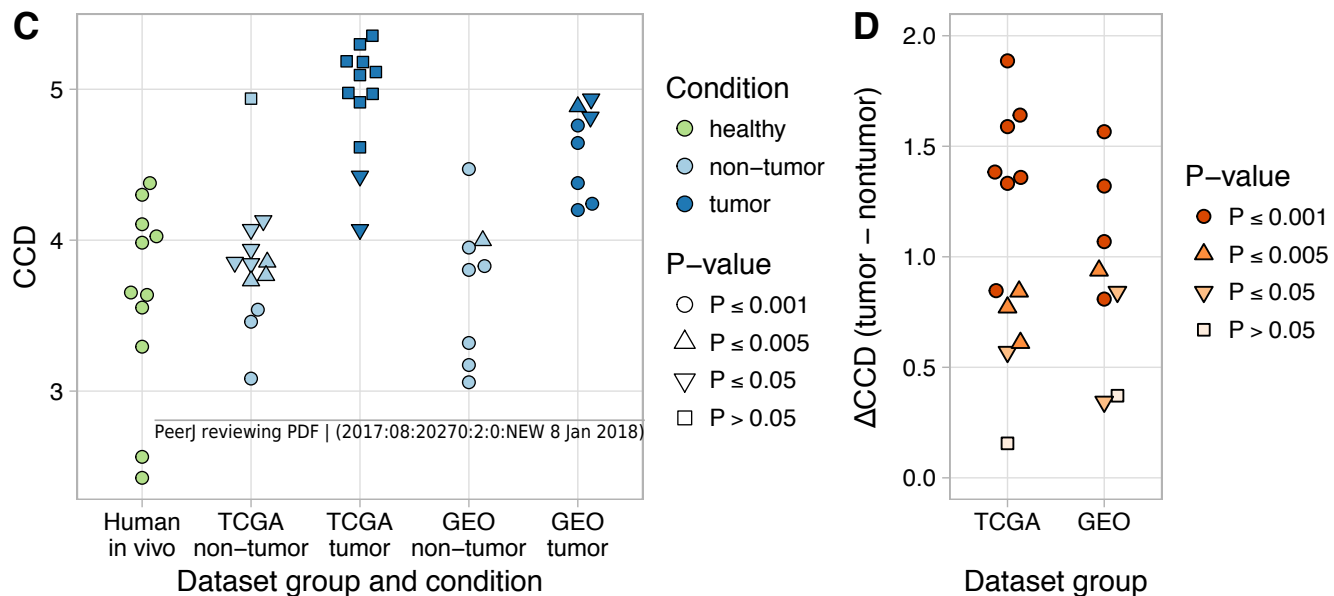
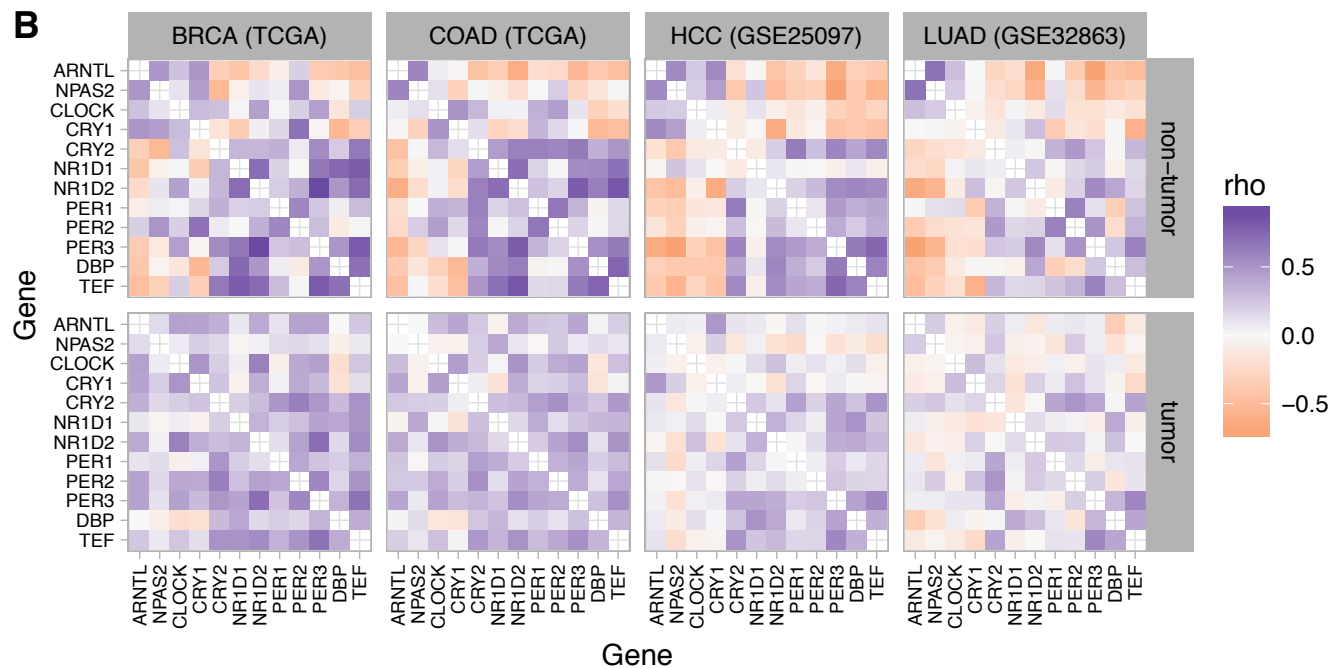
Acknowledgements

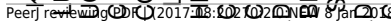
939 We thank Dvir Aran and the VUMC Editor's Club for helpful comments on the manuscript.

Data and materials availability
941 All data and code to reproduce this study are available at
942 https://figshare.com/s/2eaf11e88642418f7e81. The original gene expression data and metadata
943 for all datasets are available from NCBI GEO or Array Express. An R package to calculate the
944 CCD and/or ΔCCD for one's own gene expression data is available at
945 https://github.com/hugheylab/deltaccd. A web application is also available at
946 https://hugheylab.shinyapps.io/deltaccd.

Consistent patterns of clock gene co-expression in mice and humans.

(**A**) Schematic of procedure for constructing a reference pattern of clock gene co-expression from healthy mouse organs and for comparing clock gene co-expression in an independent dataset to the mouse reference using the clock correlation distance (CCD). (**B**) Heatmap of Spearman correlation (rho) between clock genes for the mouse reference, based on a fixed-effects meta-analysis as described in the Materials and Methods. (**C**) Clock correlation distance (CCD) for mouse and human datasets. For the boxplot corresponding to the mouse reference, distances were calculated between each pair of datasets (8 datasets give 28 unique pairs; Fig. S3). For human datasets, the CCD was calculated relative to the mouse reference. Example human (labeled) datasets are shown in (D), and example human (unlabeled) datasets are shown in (E). P-value corresponds to the probability that 12 randomly selected genes (instead of the 12 clock genes) could produce a CCD less than or equal to the one observed. (**D**) Heatmaps of Spearman correlation in three human datasets for which samples are labeled with time of day (or time since synchronization, for GSE50631). (**E**) Heatmaps for three datasets not designed to study circadian rhythms and for which samples are not labeled with time of day. Heatmaps not shown here are shown in Fig. S6. All heatmaps of clock gene co-expression in Fig. 1, Fig. 2, and the supplemental figures have the same mapping of correlation value to color, so they are directly visually comparable.

Loss of normal clock gene co-expression in human tumor samples from various cancer types.

(**A**) Schematic of procedure for comparing clock gene co-expression, relative to the mouse reference, between non-tumor and tumor samples from human cancer datasets. (**B**) Heatmaps of Spearman correlation between clock genes for non-tumor and tumor samples from two TCGA cancer types and two NCBI GEO datasets. Abbreviations: breast invasive cell carcinoma (BRCA), colon adenocarcinoma (COAD), hepatocellular carcinoma (HCC), lung adenocarcinoma (LUAD). Heatmaps for the other 16 human cancer datasets are shown in Fig. S9. (**C**) Clock correlation distance (CCD) for non-tumor and tumor samples relative to the mouse reference. Each point corresponds to one condition in one dataset. "Human in vivo" corresponds to the human in vivo datasets shown in Fig. 1C (labeled and unlabeled), minus the human blood datasets. P-value corresponds to the probability that 12 randomly selected genes (instead of the 12 clock genes) could produce a CCD less than or equal to the one observed. (**D**) Delta clock correlation distance (ΔCCD) between non-tumor and tumor samples in 12 TCGA cancer types and 8 datasets from NCBI GEO. Each point corresponds to one dataset. Positive ΔCCD indicates perturbed clock gene co-expression in tumor samples relative to non-tumor samples. P-value corresponds to the probability that a random permutation of the samples' condition labels could produce a ΔCCD greater than or equal to the one observed.

$$\Delta CCD_{t,nt} = \|\rho_t - \rho_{ref}\| - \|\rho_{nt} - \rho_{ref}\|$$
$$= CCD_t - CCD_{nt}$$

Changes in clock gene expression in human cancer are distinct from those caused by knockout of the clock genes in mice.

(**A**) Delta clock correlation distance (ΔCCD) for human cancer datasets (TCGA and GEO) and for clock gene knockout datasets from mice. (**B**) Heatmaps of the estimated $\log_2$ fold-change in expression between tumor and non-tumor samples or between knockout and wild-type samples. A positive value indicates higher expression in tumor samples or knockout samples, respectively. (**C**) Heatmaps of the $\log_2$ ratio of the median absolute deviation (MAD) of expression in tumor compared to non-tumor samples and knockout compared to wild-type samples. In the legend, MAD_a refers to tumor or knockout samples, MAD_b refers to non-tumor or wild-type samples. A positive value indicates that the variation in expression of that gene is greater in tumor (or knockout) samples than in non-tumor (or wild-type) samples. For RNA-seq data (TCGA and chondrocyte Arntl KO), expression values were based on $\log_2(\text{TPM} + 1)$, where TPM is transcripts per million. For microarray data, expression values were based on log-transformed, normalized intensity. Within each dataset group, datasets are ordered by descending ΔCCD.

Perturbed expression of normally rhythmic genes in human lung cancer.

(**A**) Hierarchical clustering of 1,292 genes inferred to have a circadian rhythm in healthy human lung. Gene modules were defined by following the procedure recommended by WGCNA. The number of genes in modules 1-5 are 62, 75, 828, 291, and 36. (**B**) Statistical significance of differential module preservation between non-tumor and tumor samples for seven preservation statistics for each module in each dataset. One-sided p-values are based on 1000 permutations of the sample labels in the respective test dataset. Abbreviations: lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), non-small cell lung cancer (NSCLC). (**C**) Scatterplots of $\log_2$ fold-change in each lung cancer dataset vs. the phase of peak expression (acrophase, in radians) in healthy lung. Acrophase π is defined to be the circular mean of the acrophases of the circadian clock-driven PAR bZip transcription factors DBP, TEF, and HLF. Each point corresponds to one of the 1,292 genes. Each blue curve indicates a fit to a periodic smoothing spline. The circular mean of the trough of the spline fits is 1.06π.