bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Widespread dysregulation of the circadian clock in human cancer

Guanhua Chen[1], Jacob J. Hughey*[,2]

[1] Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, United States
[2] Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, United States
* Correspondence: jakejhughey@gmail.com

Abstract

The mammalian circadian clock is a critical regulator of metabolism and cell division. Work in model systems indicates that disrupting the circadian clock systemically by environmental or genetic means promotes cancer, and that multiple oncogenes can in turn disrupt the circadian clock. However, whether the circadian clock is disrupted in primary human cancers is unknown. Here we used transcriptome data from mice to define a signature of the mammalian circadian clock based on the co-expression of 12 genes that form the core clock or are directly controlled by the clock. Our approach can be applied to samples that are not labeled with time of day and were not acquired over the entire circadian (24-h) cycle. We validated the clock signature in circadian transcriptome data from humans, then developed a metric we call the delta clock correlation distance (ΔCCD) to describe the extent to which the signature is perturbed in samples from one condition relative to another. We calculated the ΔCCD comparing human tumor and non-tumor samples from The Cancer Genome Atlas and six independent datasets, discovering widespread dysregulation of clock gene co-expression in tumor samples. Subsequent analysis of data from mouse clock knockouts suggested that clock dysregulation in human cancer is not caused solely by loss of activity of clock genes. Our findings suggest that dysregulation of the circadian clock is a common mechanism by which human cancers achieve unrestrained growth and division. In addition, our approach opens the door to using publicly available transcriptome data to quantify clock function in a multitude of human phenotypes.

Background

Daily rhythms in mammalian physiology are guided by a system of oscillators called the circadian clock [1]. The core clock consists of feedback loops between several genes and proteins, and based on work in mice, is active in nearly every tissue in the body [2,3]. The clock aligns itself to environmental cues, particularly cycles of light-dark and food intake [4–6]. In turn, the clock regulates various aspects of metabolism [7–9] and is tightly linked to the cell cycle [10–15].

Consistent with the tight connections between the circadian clock, metabolism, and the cell cycle, multiple studies have found that systemic disruption of the circadian system can promote

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cancer. In humans, long-term rotating shift work and night shift work, which perturb sleep-wake and circadian rhythms, have been associated with breast, colon, and lung cancer [16–18], although a recent meta-analysis found no effect of long-term shift work on breast cancer incidence [19]. In mice, environmental disruption of the circadian system (e.g., through severe and chronic jet lag) increases the risk of breast cancer and hepatocellular carcinoma [20,21]. Furthermore, both environmental and genetic disruption of the circadian system promote tumor growth and decrease survival in a mouse model of human lung adenocarcinoma [22]. Finally, pharmacological stimulation of circadian clock function slows tumor growth in a mouse model of melanoma [23]. While these studies support the link from the clock to cancer, complementary work has established a link in the other direction, namely that multiple components of a tumor, including the RAS and MYC oncogenes, can induce dysregulation of the circadian clock [24–26]. Despite this progress, however, whether the clock is disrupted in human tumors has remained unclear.

When the mammalian circadian clock is functioning normally, clock genes and clock-controlled genes show characteristic rhythms in expression throughout the body and in vitro [2,3,27]. Measurements of these rhythms through time-course experiments have revealed that the clock is altered or perturbed in some human breast cancer cell lines [28,29]. Existing computational methods for this type of analysis require that samples be labeled with time of day (or time since start of experiment) and acquired throughout the 24-h cycle [30–32]. Unfortunately, existing data from resected human tumors meet neither of these criteria.

A common approach to analyze cancer transcriptome data is to look for associations between levels of gene expression and other biological and clinical variables. For example, in human breast cancer, the expression levels of several clock genes have been associated with metastasis-free survival (with the direction of association depending on the gene) [33]. However, because a functional circadian clock is marked not by the actual levels of gene expression, but by periodic variation in gene expression, this type of analysis cannot necessarily be used to determine whether the clock is functional.

To account for this periodic variation, one approach to detect a functional clock might be to examine the correlations in expression between clock genes. Indeed, a previous study found different levels of co-expression between a few clock genes in different subtypes and grades of human breast cancer [33]. Although this finding was an important first step, its generalizability has been limited because the correlations in expression were not examined (1) for all clock genes, (2) in other human cancer types, or (3) in healthy tissues where the circadian clock is known to be functional. Thus, a definitive answer to whether the circadian clock is functional across the spectrum of human cancers is still lacking.

The goal of this study was to determine whether the circadian clock is functional in human cancer. Using transcriptome data from mice, we defined a robust signature of the mammalian circadian clock based on the co-expression of clock genes. We validated the signature in circadian transcriptome data from humans, then examined the extent to which the signature was

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

perturbed in tumor compared to non-tumor samples from The Cancer Genome Atlas (TCGA) and from multiple independent datasets. Our findings suggest that the circadian clock is dysfunctional in a wide range of human cancers.

Results

Consistent correlations in expression of clock genes in mice

The progression of the mammalian circadian clock is marked by characteristic rhythms in gene expression throughout the body [3]. We hypothesized that the relative phasing of the rhythms of different genes would give rise to a characteristic pattern of correlations between genes. Such a pattern could be used to infer the activity of the clock, even in datasets in which samples are not labeled with time of day (Fig. 1A). To investigate this hypothesis, we first collected eight publicly available datasets of genome-wide, circadian gene expression from various mouse organs under both constant darkness and alternating light-dark cycles [3,12,34–38] (Table S1). We focused on 12 genes that are part of the core circadian clock or are directly controlled by the clock and that exhibit strong, consistently phased rhythms in expression across organs [32]. For the rest of the manuscript, we will refer to these 12 genes as “clock genes.”

For each dataset, we calculated the Spearman correlation between expression values (over all samples) of each pair of genes. The pattern of correlations was highly similar across datasets and revealed two groups of genes, where the genes within a group tended to be positively correlated with each other and negatively correlated with genes in the other group (Fig. 1B). Genes in the first group (Arntl, Npas2, and Clock), which are known to form the positive arm of the clock [39], peaked in expression shortly before zeitgeber time 0 (ZT0, which corresponds to time of lights on or sunrise; Fig. S1). Genes in the second group (Cry2, Nr1d1, Nr1d2, Per1, Per2, Per3, Dbp, and Tef), which are known to form the negative arms of the clock, peaked in expression near ZT10. Cry1, which appeared to be part of the first group in some datasets and the second group in others, tended to peak in expression around ZT18. These results indicate that the progression of the circadian clock in mice produces a consistent pattern of correlations in expression between clock genes. The pattern does not depend on the absolute phasing of clock gene expression relative to time of day. Consequently, the pattern is not affected by phase shifts, such as those caused by temporally restricted feeding [40] (Fig. S2).

Most computational methods for quantifying circadian rhythmicity and inferring the status of the clock require that samples be acquired over the entire 24-h cycle. Because our approach does not attempt to infer oscillations, we wondered how robust it would be to partial coverage of the 24-h cycle. We therefore examined clock gene expression in three of the previous datasets, in samples acquired during the first 8 h of the day (or subjective day) or the first 8 h of the night (or subjective night). In each dataset, the correlation pattern was preserved in both daytime and nighttime samples (Fig. S3). These results suggest that our approach can detect an active

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

circadian clock in groups of samples without using time of day information, even if the samples' coverage of the 24-h cycle is incomplete.

Validation of the correlation pattern in humans

We next applied our approach to nine publicly available datasets of circadian transcriptome data from human tissues: one from skin [41], two from brain [42,43], three from blood [44–46], and three from cells cultured in vitro [36,47,48] (Table S1). The dataset from human skin consisted of samples taken at only three time-points for each of 19 subjects (9:30am, 2:30pm, and 7:30pm). The datasets from human brain were based on postmortem tissue from multiple anatomical areas, and zeitgeber time for each sample was based on the respective donor's date and time of death and geographic location (subject and sample info). The datasets from human blood consisted of ~8 samples taken throughout the 24-h cycle for each subject (~20 subjects per study). The datasets from cells cultured in vitro were based on time-courses following synchronization by dexamethasone, serum, or alternating temperature cycles.

The patterns of clock gene co-expression in human tissues and cells were similar to the patterns in mice (Fig. 2 and Fig. S4), which is consistent with our previous findings of similar relative phasing of clock gene expression in mice and humans [49]. The pattern was less distinct in human blood (Fig. S5), likely because several clock genes show weak or no rhythmicity in expression in blood cells [49]. The strong pattern in human skin was due to clock gene co-expression both between the three time-points and between individuals at a given time-point (Fig. S6). Compared to data from mouse organs and human skin, the correlations in human brain were noisier, which is consistent with the relatively weaker circadian rhythmicity for clock genes in those two, brain-specific datasets [49]. Overall, these results suggest that our approach can detect the signature of a functional circadian clock in human tissues in vitro and in vivo.

Aberrant patterns of clock gene expression in human cancer

To examine patterns of clock gene expression in human cancer, we applied our approach to RNA-seq data collected by The Cancer Genome Atlas (TCGA) and reprocessed using the Rsubread package [50]. TCGA samples are from surgical resections performed prior to neoadjuvant treatment. The times of day of surgery are not available and the surgeries were likely only performed during part of the day. We analyzed data from the 12 cancer types that included at least 30 samples from adjacent non-tumor tissue (Table S1). For each cancer type, we calculated the Spearman correlations in expression between clock genes across all tumor samples and all non-tumor samples.

In non-tumor samples from most cancer types, we observed a similar pattern of clock gene co-expression as in the mouse and human circadian datasets (Fig. 3 and Fig. S7). In contrast, in tumor samples from each cancer type, the pattern was weaker or absent. We observed the same trend when we restricted our analysis to only matched samples, i.e., samples from patients from whom both non-tumor and tumor samples were collected (Fig. S8). To confirm

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

these findings, we analyzed six additional datasets of gene expression in human cancer, three from liver and three from lung, each of which included matched tumor and adjacent non-tumor samples [51–56] (Table S1). As in the TCGA data, clock gene co-expression in tumor samples was perturbed relative to non-tumor samples (Fig. S9).

To quantify the dysregulation of clock gene co-expression in human cancer, we first combined the eight mouse datasets in a fixed-effects meta-analysis (Fig. 4A and Methods) in order to construct a single "reference" correlation pattern (Fig. S10 and Table S2). For each of the 12 TCGA cancer types and each of the six additional datasets of human cancer, we then calculated the Euclidean distances between the reference pattern and the non-tumor pattern and between the reference pattern and the tumor pattern. We refer to each of these distances as a clock correlation distance (CCD), and we refer to the difference between the tumor and non-tumor CCDs as the delta clock correlation distance (ΔCCD). A positive ΔCCD indicates that the correlation pattern of the non-tumor samples is more similar to the reference than is the correlation pattern of the tumor samples.

Consistent with the visualizations of clock gene co-expression, every TCGA cancer type and additional cancer dataset had a positive ΔCCD (Fig. 4B), as did the individual tumor grades in the TCGA data (Fig. S11). Among the three TCGA cancer types with the lowest ΔCCD, prostate adenocarcinoma had a relatively high non-tumor CCD (suggesting dysregulated clock gene co-expression even in non-tumor samples), whereas renal clear cell carcinoma and thyroid carcinoma had relatively low tumor CCDs (Fig. S12). To evaluate the statistical significance of the ΔCCD, we permuted the sample labels (non-tumor or tumor) in each dataset and re-calculated the ΔCCD 1000 times. Based on this permutation testing, the observed ΔCCD for 11 of the 18 datasets had a one-sided $P < 0.001$ (Fig. 4B). Taken together, these results suggest that the circadian clock is dysregulated in a wide range of human cancers.

Tumors are a complex mixture of cancer cells and various non-cancerous cell types. The proportion of cancer cells in a tumor sample is called the tumor purity and is an important factor to consider in genomic analyses of bulk tumors [57]. We therefore examined the relationship between ΔCCD and tumor purity in the TCGA data. With the exception of thyroid carcinoma and prostate adenocarcinoma, ΔCCD and median tumor purity in TCGA cancer types were positively correlated (Fig. S13; Spearman correlation = 0.67, $P = 0.059$ by exact test). These findings suggest that at least in some cancer types, dysregulation of the circadian clock is stronger in cancer cells than in non-cancer cells.

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

Finally, we investigated whether the clock dysregulation in human cancer resembled that caused by genetic mutations to core clock genes. We assembled seven datasets of circadian gene expression that included samples from wild-type mice and from mice in which at least one core clock gene was knocked out, either in the entire animal or in a specific cell type

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[8,40,58–62] (Table S1). For each dataset, we calculated the correlations in expression between pairs of clock genes in wild-type and mutant samples and calculated the ΔCCD (Fig. S14 and Fig. 5A).

The two datasets with the highest ΔCCD (>50% higher than any ΔCCD we observed in human cancer) were those in which the mutant mice lacked not one, but two components of the clock (Cry1 and Cry2 in GSE13093; Nr1d1 and Nr1d2 in GSE34018). The ΔCCDs for the other five mutants were similar to or somewhat lower than the ΔCCDs we observed in human cancer. Given the smaller sample sizes compared to the human cancer datasets, the ΔCCDs for those five mutants were not significant (one-sided $P > 0.05$ by permutation test). In addition, we observed no strong trends in the expression levels of individual clock genes in either human cancer or mouse clock gene knockouts (Fig. S15).

In each of the clock gene knockouts, rhythmic expression of the clock genes was reduced or lost (Fig. S16). Although it was not possible to quantify the rhythmicity of clock gene expression in the human cancer datasets directly, we reasoned that a proxy for rhythmicity could be the magnitude of variation in expression. Therefore, for each TCGA cancer type and each additional human cancer dataset, we calculated the median absolute deviation (MAD) in expression of the clock genes in non-tumor and tumor samples. We then compared the $\log_2$ ratios of MAD between tumor and non-tumor samples to the $\log_2$ ratios of MAD between mutant and wild-type samples from the clock gene knockout data (Fig. 5B). As expected, samples from clock gene knockouts showed widespread reductions in MAD compared to samples from wild-type mice. In contrast, human tumor samples tended to show similar or even somewhat higher MAD compared to non-tumor samples. These results suggest that the dysregulation of the clock in human cancer is not due solely to loss of activity of one or more core clock genes.

Discussion

Increasing evidence has suggested that systemic disruption of the circadian clock can promote tumor development and that components of a tumor can disrupt the circadian clock. Until now, however, the question of whether the clock is functional in primary human cancers has lacked a clear answer. Here we developed a simple method to probe clock function based on the co-expression of a small set of clock genes. By applying the method to cancer transcriptome data, we uncovered widespread dysregulation of the clock in human cancer tissue.

Our approach for detecting a functional circadian clock is based on three principles. First, we rely on prior knowledge of clock genes and clock-controlled genes. Second, we account for the fact that the clock is defined not by a static condition, but by a dynamic cycle. Our approach thus exploits the co-expression of clock genes that arises from (1) different genes having different circadian phases and (2) different samples being taken from different points in the cycle. Finally, our method does not attempt to infer an oscillatory pattern, but instead uses only the statistical correlations between pairs of genes. The assumption is that if the clock is perturbed, the correlations will be altered. Although the correlation matrix only partially captures

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the complex relationship between genes (to fully capture the relationship would require modeling the joint probability density of gene expression), it is intuitive and straightforward to calculate. Altogether, these principles enable our method to detect the signature of the circadian clock in groups of samples whose times of day of acquisition are unknown and whose coverage of the 24-h cycle is incomplete.

One limitation of our method is that it is insensitive to the alignment of the circadian clock to the time of day, and so cannot detect phase differences between conditions. This limitation, however, allowed us to readily construct a reference pattern using data from mice and compare it to data from humans, despite the circadian phase difference between the two species [49]. A second limitation is that because our method relies on co-expression across samples, it does not immediately lend itself to quantifying clock function in single samples. In the future, it may be possible to complement the ΔCCD and assess clock function in some datasets by directly comparing matched samples from the same patient.

In healthy tissues in vivo, the circadian clocks of individual cells are entrained and oscillating together, which is what allows bulk measurements to contain robust circadian signals. Consequently, the loss of a circadian signature in human tumor samples could result from dysfunction in either entrainment, the oscillator, or both. Dysfunction in entrainment would imply that the clocks in at least some of the cancer cells are out of sync with each other and therefore free running, i.e. ignoring zeitgeber signals. Dysfunction in the oscillator would imply that the clocks in at least some of the cancer cells are no longer "ticking" (albeit in such a way that variation in clock gene expression, at least across patients, is not diminished). Given the current data, which are based on averaged clock gene expression from many cells, these scenarios cannot be distinguished. Furthermore, the moderate correlation between ΔCCD and tumor purity across cancer types leads us to speculate that the circadian clocks in stromal and/or infiltrating immune cells may be operating normally. In the future, these issues may be resolved through a combination of mathematical modeling [63,64] and single-cell measurements. A separate matter not addressed here is how the cancer influences circadian rhythms in the rest of the body [65], which may be relevant for optimizing the daily timing of anticancer treatments [66].

Based on the current data alone, which are observational, it is not possible to determine whether dysregulation of the circadian clock is a driver of the cancer or merely a passenger. However, when taken together with the clock's established role in regulating metabolism and a recent finding that stimulation of the clock inhibits tumor growth in melanoma [23], our findings suggest that clock dysregulation may be a cancer driver in multiple solid tissues. On the other hand, a functional circadian clock seems to be required for growth of acute myeloid leukemia cells [67], so further work is necessary to clarify this issue.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Conclusions

Our findings suggest that dysregulation of the circadian clock is a common mechanism by which human cancers achieve unrestrained growth and division. Thus, restoring clock function could be a viable therapeutic strategy in a wide range of cancer types. In addition, given the practical challenges of studying circadian rhythms at the cellular level in humans, our method offers the possibility to quantify clock function in a wide range of human phenotypes using publicly available transcriptome data.

Methods

Selecting the datasets

We selected the datasets of circadian gene expression in mice (both for defining the reference pattern and for comparing clock gene knockouts to wild-type) to represent multiple organs, light-dark regimens, and microarray platforms. For circadian gene expression in humans, we included three datasets from blood, two from brain, and one from skin. The samples from blood and skin were obtained from living volunteers, whereas the samples from brain were obtained from postmortem donors who had died rapidly. For GSE45642 (human brain), we only included samples from control subjects (i.e., we excluded subjects with major depressive disorder). Zeitgeber times for samples from GSE56931 (human blood) were calculated as described previously [68]. For the TCGA data, we analyzed all cancer types that had at least 30 non-tumor samples (all of which also had at least 291 tumor samples). When analyzing clock gene expression in human cancer, unless otherwise noted, we included all tumor and non-tumor samples, not just those from patients from whom both non-tumor and tumor samples were collected. For details of the datasets, all of which are publicly available, see Table S1.

Processing the gene expression data

For TCGA samples, we obtained the processed RNA-seq data (in units of transcripts per million, TPM, on a gene-level basis) and the corresponding metadata (cancer type, patient ID, etc.) from GSE62944 [50]. For E-MTAB-3428, we downloaded the RNA-seq read files from the European Nucleotide Archive, used kallisto to quantify transcript-level abundances in units of TPM [69], then used the mapping between Ensembl Transcript IDs and Entrez Gene IDs to calculate gene-level abundances.

For the remaining datasets, raw (in the case of Affymetrix) or processed microarray data were obtained from NCBI GEO and processed using MetaPredict, which maps probes to Entrez Gene IDs and performs intra-study normalization and log-transformation [70]. MetaPredict processes raw Affymetrix data using RMA and customCDFs [71,72]. As in our previous study, we used ComBat to reduce batch effects between anatomical areas in human brain and between subjects in human blood [49,73].

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Analyzing the gene expression data

We focused our analysis on the expression of 12 genes that are considered part of the core clock or are directly controlled by the clock and that show strong, consistently phased rhythms in a wide range of mouse organs [3,32]. We calculated times of peak expression and strengths of circadian rhythmicity of expression in wild-type and mutant mice using ZeitZeiger [32], with three knots for the periodic smoothing splines [74].

We quantified the relationship between expression values of pairs of genes using the Spearman correlation (Spearman's rho), which is rank-based and therefore invariant to monotonic transformations such as the logarithm and less sensitive to outliers than the Pearson correlation. All heatmaps of gene-gene correlations in this paper have the same mapping of correlation value to color, so they are directly visually comparable.

We calculated the reference Spearman correlation for each pair of genes (Table S2) using a fixed-effects meta-analysis of the eight mouse datasets shown in Fig. 1 [75]. First, we applied the Fisher z-transformation ($arctanh$) to the correlations from each dataset. Then we calculated a weighted average of the transformed correlations, where the weight for dataset i was $n_i - 3$ (corresponding to the inverse variance of the transformed correlation), where n_i is the number of samples in dataset i. Finally, we applied the inverse transformation ($tanh$) to the weighted average.

To quantify the similarity in clock gene expression between two groups of samples (e.g., between the mouse reference and human tumor samples), we calculated the Euclidean distance between the respective Spearman correlation vectors, which contains all values in the strictly lower (or strictly upper) triangular part of the correlation matrix. Given a reference and a dataset with samples from two conditions, we calculated the Euclidean distances between the reference and each condition, which we call the clock correlation distances (CCDs). We then calculated the difference between these two distances, which indicates how much more similar to the reference one condition is than the other and which we refer to as the delta clock correlation distance (ΔCCD). Although here we used Euclidean distance, other distance metrics could be used as well.

To evaluate the statistical significance of the ΔCCD for a given dataset, we conducted the permutation test as follows: First, we permuted the relationship between the sample labels (e.g., non-tumor or tumor) and the gene expression values and recalculated the ΔCCD 1000 times, always keeping the reference fixed. We then calculated the one-sided p-value as the fraction of permutations that gave a ΔCCD greater than or equal to the observed ΔCCD. Since we used the one-sided p-value, the alternative hypothesis was that non-tumor (or wild-type) is more similar to the reference than is tumor (or mutant).

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

To compare ΔCCD and tumor purity, we used published consensus purity estimates for TCGA tumor samples [57]. The estimates are based on DNA methylation, somatic copy number variation, and the expression of immune genes and stromal genes (none of which are clock genes).

We quantified the variation in expression of clock genes in each dataset and condition using the median absolute deviation (MAD), which is less sensitive to outliers than the standard deviation.

Abbreviations

ΔCCD: delta clock correlation distance; MAD: median absolute deviation; TCGA: The Cancer Genome Atlas; ZT: zeitgeber time; BRCA: breast invasive cell carcinoma; COAD: colon adenocarcinoma; HNSC: head and neck squamous cell carcinoma; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; PRAD: prostate adenocarcinoma; STAD: stomach adenocarcinoma; THCA: thyroid carcinoma; UCEC: uterine corpus endometrial carcinoma

Declarations

Acknowledgments

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

Funding

This work was supported by start-up funds from the Vanderbilt University School of Medicine (to JJH) and by NIH grants 1U2COD023196-01 and U01HG009086-01 (to GC).

Availability of data and materials

All data and code to reproduce this study are available at https://figshare.com/s/2eaf11e88642418f7e81. The original gene expression data and metadata for all datasets are available from NCBI GEO or Array Express. A web application to calculate ΔCCD for one's own gene expression data is available at https://jakejh.shinyapps.io/deltaccd.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Author contributions

GC conceived and designed the analysis and reviewed drafts of the paper. JJH conceived and designed the analysis, performed the analysis, wrote the paper, and reviewed drafts of the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Additional files

Table S1
Figures S1-S15
Table S2

References

1. Dibner C, Schibler U, Albrecht U. The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks. Annu. Rev. Physiol. 2010;72:517–49.

2. Yoo S-H, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, et al. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc. Natl. Acad. Sci. U. S. A. 2004;101:5339–46.

3. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. Proc. Natl. Acad. Sci. U. S. A. 2014;111:16219–24.

4. Damiola F, Minh NL, Preitner N, Kornmann B, Fleury-Olela F, Schibler U. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. Genes Dev. 2000;14:2950–61.

5. Asher G, Reinke H, Altmeyer M, Gutierrez-Arcelus M, Hottiger MO, Schibler U. Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. Cell. 2010;142:943–53.

6. Eckel-Mahan KL, Patel VR, de Mateo S, Orozco-Solis R, Ceglia NJ, Sahar S, et al. Reprogramming of the circadian clock by nutritional challenge. Cell. 2013;155:1464–78.

7. Nakahata Y, Sahar S, Astarita G, Kaluzova M, Sassone-Corsi P. Circadian control of the NAD+ salvage pathway by CLOCK-SIRT1. Science. 2009;324:654–7.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

8. Cho H, Zhao X, Hatori M, Yu RT, Barish GD, Lam MT, et al. Regulation of circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β. Nature. 2012;485:123–7.

9. Neufeld-Cohen A, Robles MS, Aviram R, Manella G, Adamovich Y, Ladeuix B, et al. Circadian control of oscillations in mitochondrial rate-limiting enzymes and nutrient utilization by PERIOD proteins. Proc. Natl. Acad. Sci. U. S. A. 2016;113:E1673–82.

10. Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H. Control mechanism of the circadian clock for timing of cell division in vivo. Science. 2003;302:255–9.

11. Gréchez-Cassiau A, Rayet B, Guillaumond F, Teboul M, Delaunay F. The circadian clock component BMAL1 is a critical regulator of p21WAF1/CIP1 expression and hepatocyte proliferation. J. Biol. Chem. 2008;283:4535–42.

12. Geyfman M, Kumar V, Liu Q, Ruiz R, Gordon W, Espitia F, et al. Brain and muscle Arnt-like protein-1 (BMAL1) controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the epidermis. Proc. Natl. Acad. Sci. U. S. A. 2012;109:11758–63.

13. Bieler J, Cannavo R, Gustafson K, Gobet C, Gatfield D, Naef F. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Mol. Syst. Biol. 2014;10:739.

14. Feillet C, Krusche P, Tamanini F, Janssens RC, Downey MJ, Martin P, et al. Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. Proc. Natl. Acad. Sci. U. S. A. 2014;111:9828–33.

15. Matsu-Ura T, Dovzhenok A, Aihara E, Rood J, Le H, Ren Y, et al. Intercellular Coupling of the Cell Cycle and Circadian Clock in Adult Stem Cell Culture. Mol. Cell. 2016;64:900–12.

16. Schernhammer ES, Laden F, Speizer FE, Willett WC, Hunter DJ, Kawachi I, et al. Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. J. Natl. Cancer Inst. 2001;93:1563–8.

17. Schernhammer ES, Laden F, Speizer FE, Willett WC, Hunter DJ, Kawachi I, et al. Night-shift work and risk of colorectal cancer in the nurses' health study. J. Natl. Cancer Inst. 2003;95:825–8.

18. Schernhammer ES, Feskanich D, Liang G, Han J. Rotating night-shift work and lung cancer risk among female nurses in the United States. Am. J. Epidemiol. 2013;178:1434–41.

19. Travis RC, Balkwill A, Fensom GK, Appleby PN, Reeves GK, Wang X-S, et al. Night Shift Work and Breast Cancer Incidence: Three Prospective Studies and Meta-analysis of Published Studies. J. Natl. Cancer Inst. [Internet]. 2016;108. Available from: http://dx.doi.org/10.1093/jnci/djw169

20. Van Dycke KCG, Rodenburg W, van Oostrom CTM, van Kerkhof LWM, Pennings JLA, Roenneberg T, et al. Chronically Alternating Light Cycles Increase Breast Cancer Risk in Mice. Curr. Biol. 2015;25:1932–7.

21. Kettner NM, Voicu H, Finegold MJ, Coarfa C, Sreekumar A, Putluri N, et al. Circadian

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Homeostasis of Liver Metabolism Suppresses Hepatocarcinogenesis. Cancer Cell [Internet]. 2016; Available from: http://dx.doi.org/10.1016/j.ccell.2016.10.007

22. Papagiannakopoulos T, Bauer MR, Davidson SM, Heimann M, Subbaraj L, Bhutkar A, et al. Circadian Rhythm Disruption Promotes Lung Tumorigenesis. Cell Metab. 2016;24:324–31.

23. Kiessling S, Beaulieu-Laroche L, Blum ID, Landgraf D, Welsh DK, Storch K-F, et al. Enhancing circadian clock function in cancer cells inhibits tumor growth. BMC Biol. 2017;15:13.

24. Relógio A, Thomas P, Medina-Pérez P, Reischl S, Bervoets S, Gloc E, et al. Ras-mediated deregulation of the circadian clock in cancer. PLoS Genet. 2014;10:e1004338.

25. Michael AK, Harvey SL, Sammons PJ, Anderson AP, Kopalle HM, Banham AH, et al. Cancer/Testis Antigen PASD1 Silences the Circadian Clock. Mol. Cell. 2015;58:743–54.

26. Altman BJ, Hsieh AL, Sengupta A, Krishnanaiah SY, Stine ZE, Walton ZE, et al. MYC Disrupts the Circadian Clock and Metabolism in Cancer Cells. Cell Metab. 2015;22:1009–19.

27. Balsalobre A, Damiola F, Schibler U. A serum shock induces circadian gene expression in mammalian tissue culture cells. Cell. 1998;93:929–37.

28. Rossetti S, Esposito J, Corlazzoli F, Gregorski A, Sacchi N. Entrainment of breast (cancer) epithelial cells detects distinct circadian oscillation patterns for clock and hormone receptor genes. Cell Cycle. 2012;11:350–60.

29. Xiang S, Mao L, Duplessis T, Yuan L, Dauchy R, Dauchy E, et al. Oscillation of clock and clock controlled genes induced by serum shock in human breast epithelial and breast cancer cells: regulation by melatonin. Breast Cancer . 2012;6:137–50.

30. Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB. MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics. 2016;32:3351–3.

31. Thaben PF, Westermark PO. Differential rhythmicity: detecting altered rhythmicity in biological data. Bioinformatics. 2016;32:2800–8.

32. Hughey JJ, Hastie T, Butte AJ. ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. Nucleic Acids Res. 2016;44:e80.

33. Cadenas C, Sandt L van de, Edlund K, Lohr M, Hellwig B, Marchan R, et al. Loss of circadian clock gene expression is associated with tumor progression in breast cancer. Cell Cycle. 2014;13:3282–91.

34. Oster H, Damerow S, Hut RA, Eichele G. Transcriptional profiling in the adrenal gland reveals circadian regulation of hormone biosynthesis genes and nucleosome assembly genes. J. Biol. Rhythms. 2006;21:350–61.

35. Hoogerwerf WA, Sinha M, Conesa A, Luxon BA, Shahinian VB, Cornélissen G, et al. Transcriptional profiling of mRNA expression in the mouse distal colon. Gastroenterology. 2008;135:2019–29.

36. Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, et al. Harmonics

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of Circadian Gene Transcription in Mammals. PLoS Genet. 2009;5:e1000442.

37. Negoro H, Kanematsu A, Doi M, Suadicani SO, Matsuo M, Imamura M, et al. Involvement of urinary bladder Connexin43 and the circadian clock in coordination of diurnal micturition rhythm. Nat. Commun. 2012;3:809.

38. Haspel JA, Chettimada S, Shaik RS, Chu J-H, Raby BA, Cernadas M, et al. Circadian rhythm reprogramming during lung inflammation. Nat. Commun. 2014;5:4753.

39. Partch CL, Green CB, Takahashi JS. Molecular architecture of the mammalian circadian clock. Trends Cell Biol. 2014;24:90–9.

40. Vollmers C, Gill S, DiTacchio L, Pulivarthy SR, Le HD, Panda S. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc. Natl. Acad. Sci. U. S. A. 2009;106:21453–8.

41. Spörl F, Korge S, Jürchott K, Wunderskirchner M, Schellenberg K, Heins S, et al. Krüppel-like factor 9 is a circadian transcription factor in human epidermis that controls proliferation of keratinocytes. Proc Natl Acad Sci USA. 2012;109:10903–8.

42. Li JZ, Bunney BG, Meng F, Hagenauer MH, Walsh DM, Vawter MP, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc Natl Acad Sci USA. 2013;110:9950–5.

43. Chen C-Y, Logan RW, Ma T, Lewis DA, Tseng GC, Sibille E, et al. Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. Proc Natl Acad Sci USA. 2016;113:206–11.

44. Möller-Levet CS, Archer SN, Bucca G, Laing EE, Slak A, Kabiljo R, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. Proc Natl Acad Sci USA. 2013;110:E1132–41.

45. Archer SN, Laing EE, Möller-Levet CS, Veen DR van der, Bucca G, Lazar AS, et al. Mistimed sleep disrupts circadian regulation of the human transcriptome. Proc Natl Acad Sci USA. 2014;111:E682–91.

46. Arnardottir ES, Nikonova EV, Shockley KR, Podtelezhnikov AA, Anafi RC, Tanis KQ, et al. Blood-Gene Expression Reveals Reduced Circadian Rhythmicity in Individuals Resistant to Sleep Deprivation. Sleep. 2014;37:589–600.

47. Janich P, Toufighi K, Solanas G, Luis NM, Minkwitz S, Serrano L, et al. Human epidermal stem cell function is regulated by circadian oscillations. Cell Stem Cell. 2013;13:745–53.

48. Hoffmann J, Symul L, Shostak A, Fischer T, Naef F, Brunner M. Non-circadian expression masking clock-driven weak transcription rhythms in U2OS cells. PLoS One. 2014;9:e102238.

49. Hughey JJ, Butte AJ. Differential Phasing between Circadian Clocks in the Brain and Peripheral Organs in Humans. J. Biol. Rhythms. 2016;31:588–97.

50. Rahman M, Jackson LK, Johnson WE, Li DY, Bild AH, Piccolo SR. Alternative preprocessing of RNA-Sequencing data in The Cancer Genome Atlas leads to improved

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

analysis results. Bioinformatics. 2015;31:3666–72.

51. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One. 2008;3:e1651.

52. Roessler S, Jia H-L, Budhu A, Forgues M, Ye Q-H, Lee J-S, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res. 2010;70:10202–12.

53. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One. 2010;5:e10312.

54. Lamb JR, Zhang C, Xie T, Wang K, Zhang B, Hao K, et al. Predictive genes in adjacent normal tissue are preferentially altered by sCNV during tumorigenesis in liver cancer and may rate limiting. PLoS One. 2011;6:e20090.

55. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. Genome Res. 2012;22:1197–211.

56. Villa E, Critelli R, Lei B, Marzocchi G, Cammà C, Giannelli G, et al. Neoangiogenesis-related genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Results from a prospective study. Gut. 2016;65:861–9.

57. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat. Commun. 2015;6:8971.

58. Nikolaeva S, Pradervand S, Centeno G, Zavadova V, Tokonami N, Maillard M, et al. The circadian clock modulates renal sodium handling. J. Am. Soc. Nephrol. 2012;23:1019–26.

59. Paschos GK, Ibrahim S, Song W-L, Kunieda T, Grant G, Reyes TM, et al. Obesity in mice with adipocyte-specific deletion of clock component Arntl. Nat. Med. 2012;18:1768–77.

60. Dyar KA, Ciciliot S, Wright LE, Biensø RS, Tagliazucchi GM, Patel VR, et al. Muscle insulin sensitivity and glucose metabolism are controlled by the intrinsic muscle clock. Mol Metab. 2014;3:29–41.

61. Young ME, Brewer RA, Peliciari-Garcia RA, Collins HE, He L, Birky TL, et al. Cardiomyocyte-specific BMAL1 plays critical roles in metabolism, signaling, and maintenance of contractile function of the heart. J. Biol. Rhythms. 2014;29:257–76.

62. Dudek M, Gossan N, Yang N, Im H-J, Ruckshanthi JPD, Yoshitane H, et al. The chondrocyte clock gene Bmal1 controls cartilage homeostasis and integrity. J. Clin. Invest. 2016;126:365–76.

63. Kim JK, Forger DB. A mechanism for robust circadian timekeeping via stoichiometric balance. Mol. Syst. Biol. 2012;8:630.

64. Lück S, Thurley K, Thaben PF, Westermark PO. Rhythmic degradation explains and unifies

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

circadian transcriptome and proteome data. Cell Rep. 2014;9:741–51.

65. Masri S, Papagiannakopoulos T, Kinouchi K, Liu Y, Cervantes M, Baldi P, et al. Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis. Cell. 2016;165:896–909.

66. Dulong S, Ballesta A, Okyar A, Lévi F. Identification of Circadian Determinants of Cancer Chronotherapy through In Vitro Chronopharmacology and Mathematical Modeling. Mol. Cancer Ther. 2015;14:2154–64.

67. Puram RV, Kowalczyk MS, de Boer CG, Schneider RK, Miller PG, McConkey M, et al. Core Circadian Clock Genes Regulate Leukemia Stem Cells in AML. Cell. 2016;165:303–16.

68. Hughey JJ. Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. Genome Med. 2017;9:19.

69. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 2016;34:525–7.

70. Hughey JJ, Butte AJ. Robust meta-analysis of gene expression using the elastic net. Nucleic Acids Res. 2015;43:e79.

71. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249–64.

72. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005;33:e175.

73. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.

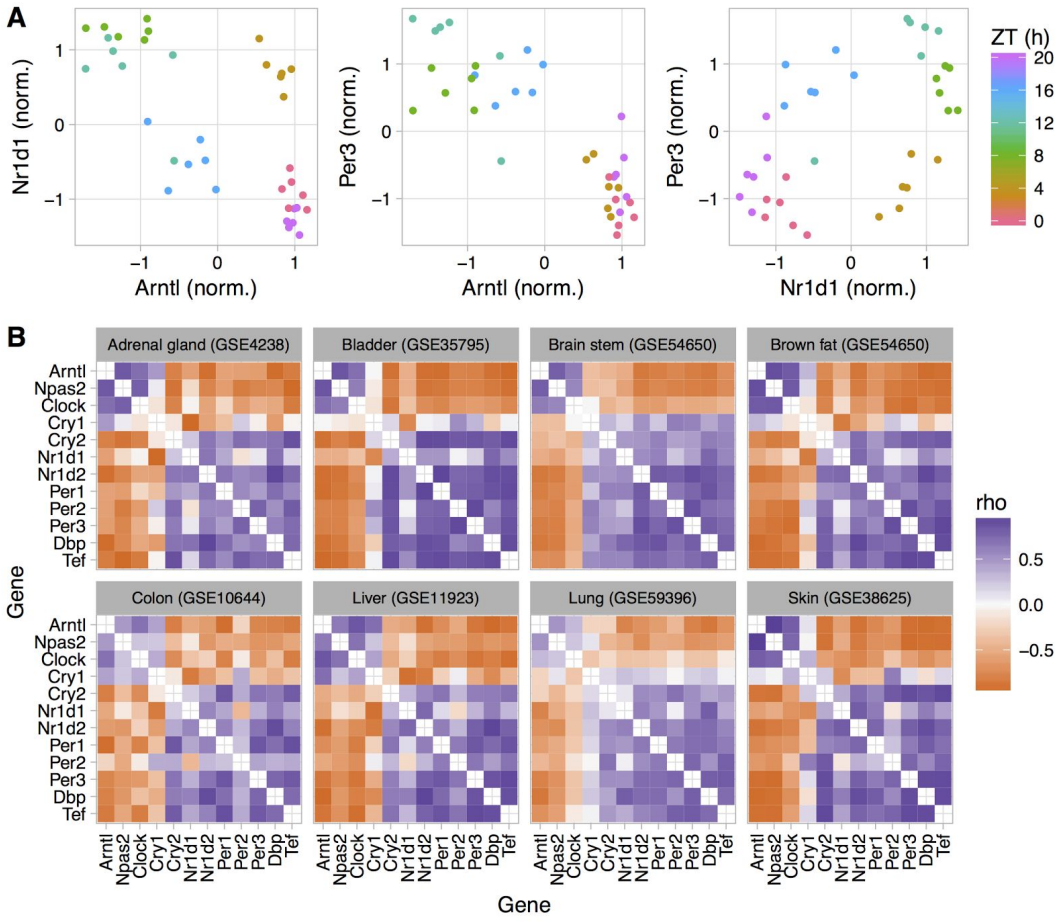
74. Helwig NE, Ma P. Fast and stable multiple smoothing parameter selection in smoothing spline analysis of variance models with large samples. J. Comput. Graph. Stat. 2014;24:715–32.

75. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Academic Press; 1985.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures

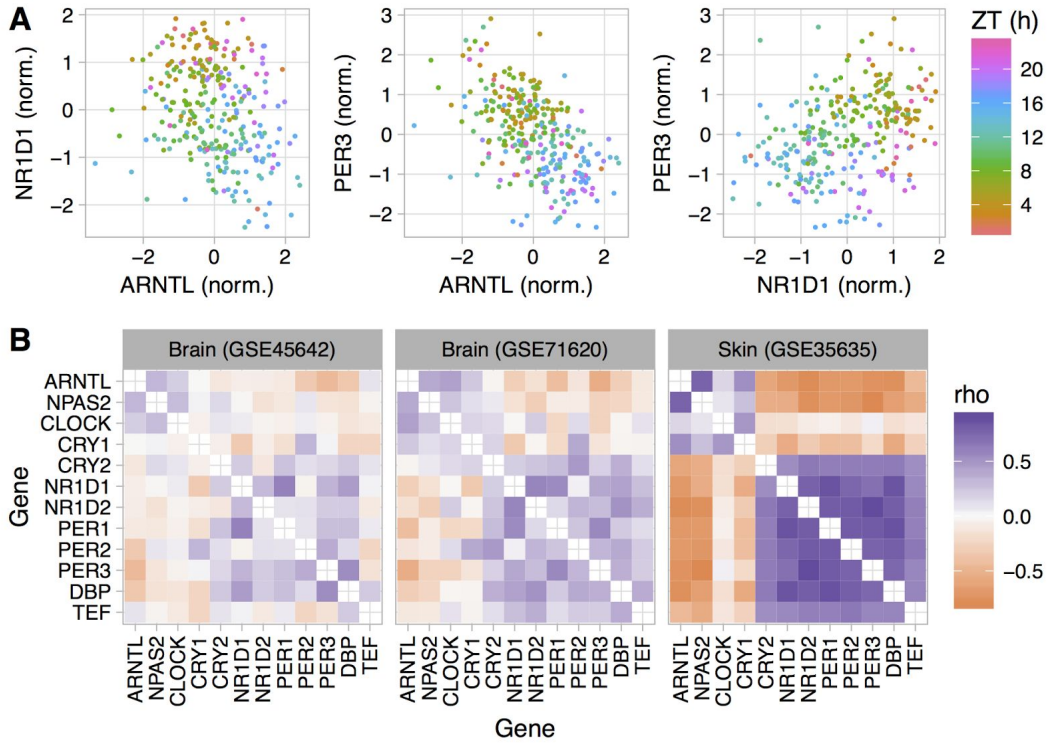
Figure 1



Consistent patterns of clock gene co-expression in various mouse organs. (A) Scatterplots of expression for three clock genes in lung (GSE59396). Each point is a sample, and the color indicates zeitgeber time, where ZT0 corresponds to “lights on.” Expression values of each gene were normalized to have mean zero and standard deviation one. (B) Heatmaps of Spearman correlation (“rho”) between each pair of clock genes in each dataset. Genes are ordered manually by a combination of name and known function in the clock. GSE54650 includes gene expression from 12 organs, however to maintain diversity in datasets, we used data from only two organs.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

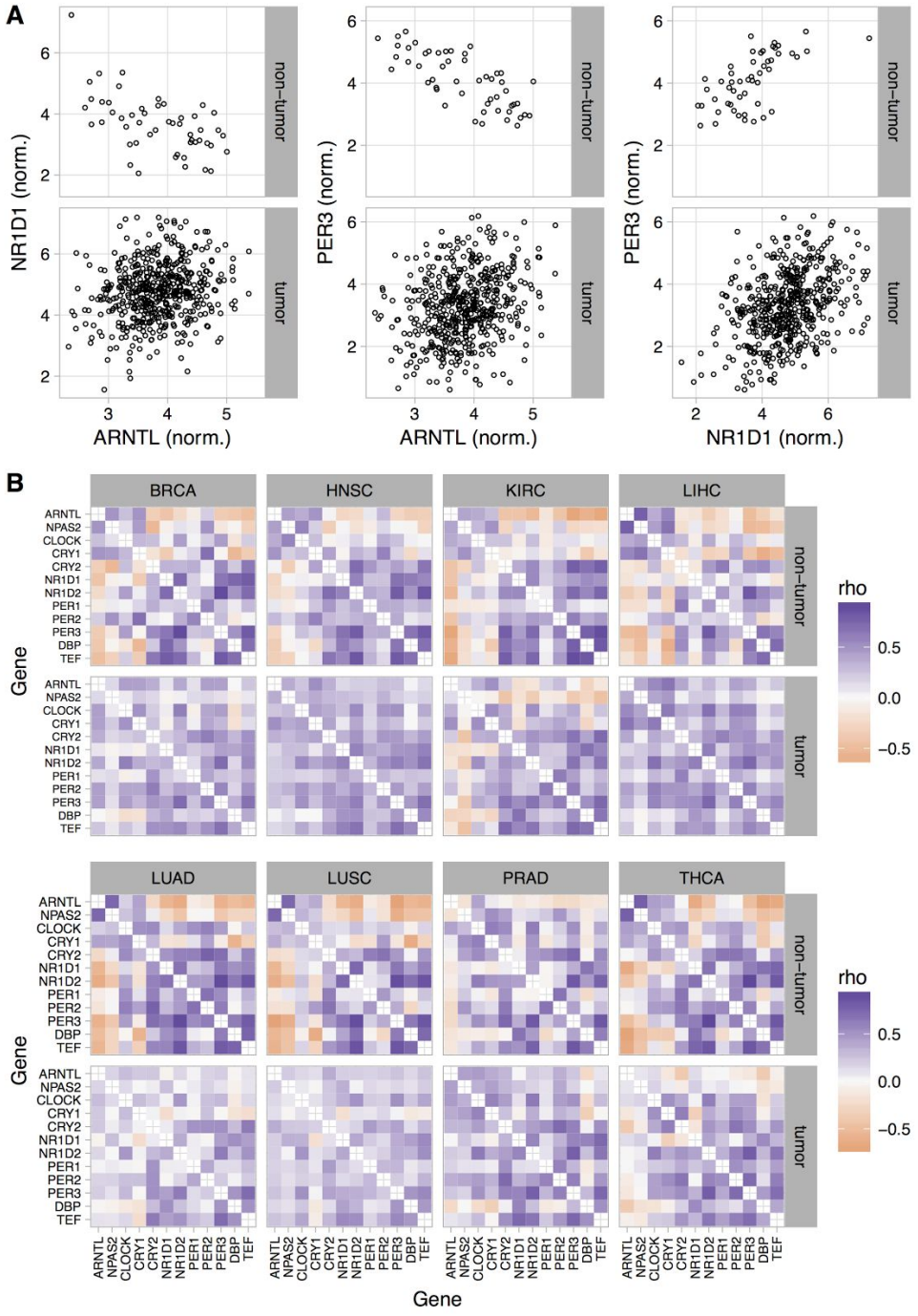
Figure 2



Consistent patterns of clock gene co-expression in human skin and brain, based on transcriptome data designed to quantify circadian variation in gene expression. (A) Scatterplots of expression for three clock genes in GSE71620 (gene expression measured in postmortem tissue, with zeitgeber time based on time of death). Each point is a sample, and the color corresponds to zeitgeber time, where ZT0 corresponds to sunrise. Expression values of each gene were normalized to have mean zero and standard deviation one. (B) Heatmaps of Spearman correlation between each pair of clock genes in each dataset. Genes are in the same order as their mouse orthologs in Fig. 1.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3



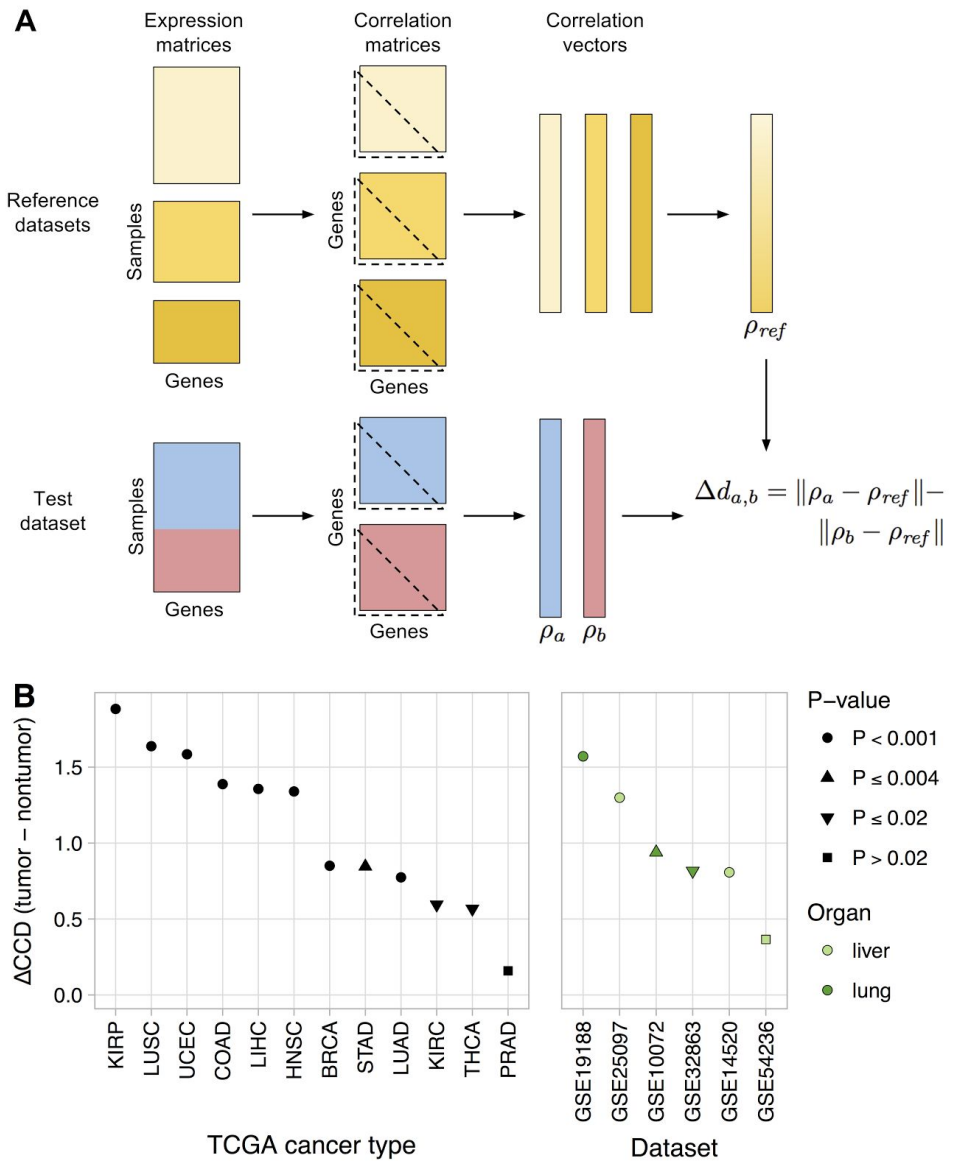
Loss of normal clock gene co-expression in tumor samples from various TCGA cancer types. (A) Pairwise scatterplots of expression for three clock genes in tumor and adjacent non-tumor samples from lung squamous cell carcinoma (LUSC). Each point is a sample. Expression values from RNA-seq data are shown in units of $\log_2(\text{tpm} + 1)$. For ease of visualization, two tumor samples with very high normalized expression of ARNTL are not shown. (B) Heatmaps of

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Spearman correlation between each pair of clock genes in non-tumor and tumor samples in each cancer type. The eight cancer types with at least 44 non-tumor samples are shown here. Four other cancer types with at least 30 non-tumor samples are shown in Fig. S6. Cancer types shown are breast invasive cell carcinoma (BRCA), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), and thyroid carcinoma (THCA). In each cancer type, all tumor samples are used, not only those with matched non-tumor samples.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

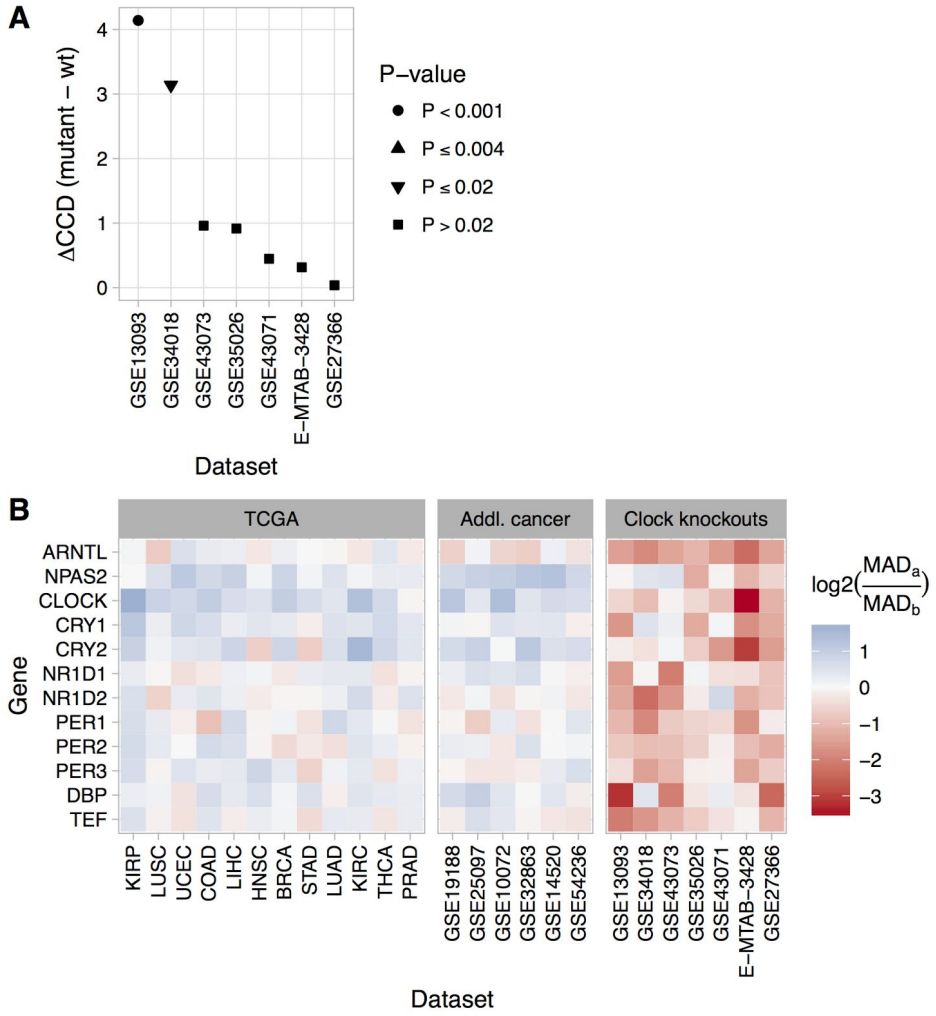
Figure 4



Quantifying dysregulation of clock gene co-expression in human cancer. (A) Schematic of procedure for comparing patterns of clock gene expression between two conditions in samples lacking time of day information. For details of procedure, see Methods. (B) Delta clock correlation distance (ΔCCD) between non-tumor and tumor samples in 12 TCGA cancer types and six additional datasets. Positive ΔCCD indicates that the correlation pattern of the non-tumor samples is more similar to the mouse reference than is the correlation pattern of the tumor samples. P-values are one-sided and are based on 1000 permutations between the sample labels (non-tumor or tumor) and the gene expression values, and have not been adjusted for multiple hypotheses.

bioRxiv preprint doi: https://doi.org/10.1101/130765; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 5



Dysregulation of the circadian clock in human cancer is distinct from that caused by knockout of the clock genes in mice. (A) ΔCCD between wild-type and mutant samples in seven datasets of mouse clock gene knockouts. Positive ΔCCD indicates that the correlation pattern of the wild-type samples is more similar to the mouse reference than is the correlation pattern of the mutant samples. P-values are one-sided and are based on 1000 permutations between the sample labels and the gene expression values, and have not been adjusted for multiple hypotheses. (B) Heatmaps of the $\log_2$ ratio of the median absolute deviation (MAD) of expression in tumor compared to non-tumor samples and mutant compared to wild-type samples. In the legend, MAD_a refers to tumor or mutant 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 mutant) samples than in non-tumor (or wild-type) samples. For RNA-seq data (TCGA and E-MTAB-3428), expression values were based on $\log_2(\text{tpm} + 1)$. For microarray data, expression values were based on log-transformed, normalized intensity. Datasets are ordered by descending ΔCCD.