Time series-free rhythm profiling using COFE reveals multi-omic circadian rhythms in in-vivo human cancers

Bharath Ananthasubramaniam[1,*] and Ramji Venkataramanan[2]

[1]Institute for Theoretical Biology, Humboldt Universität zu Berlin, Berlin, Germany
[2]Department of Engineering, University of Cambridge, Cambridge, United Kingdom
[*]Corresponding Author: bharath.ananthasubramaniam@hu-berlin.de

Abstract

The study of the ubiquitous circadian rhythms in human physiology typically requires regular measurements across time. Repeated sampling of the different internal tissues that house circadian clocks is both practically and ethically infeasible. Here, we present a novel unsupervised machine learning approach (COFE) that can use single high-throughput (omics) samples from individuals to reconstruct circadian rhythms across cohorts. COFE can both assign time-labels to samples, and identify rhythmic data features used for temporal reconstruction. With COFE, we discovered widespread de novo circadian gene expression rhythms in eleven different human adenocarcinomas using data from the TCGA database. The arrangement of peak times of core clock gene expression was conserved across the cancers and resembled a healthy functional clock except for the mistiming of few key genes. The commonly rhythmic genes were involved in metabolism and the cell cycle. Although these rhythms were synchronized with the cell cycle in many cancers, they were uncoupled with clocks in healthy matched tissue. Moreover, rhythms in the transcriptome were strongly associated with the cancer-relevant proteome. The targets of most of FDA-approved and potential anti-cancer drugs were rhythmic in tumor tissue with different amplitudes and peak times, emphasizing the utility of considering "time" in cancer therapy. Our approach thus creates new opportunities to repurpose data without time-labels to study circadian rhythms.

Introduction

Near 24h biological rhythms, called circadian rhythms, are essential to healthy human physiology and their disruption is associated with several pathologies [1]. These rhythms are orchestrated by a symphony of biological clocks in most tissues that drive tissue-specific temporal programs [2]. Characterization of these rhythms in a human subject under health and disease (in the relevant tissue) requires measurements at regular time intervals over one or more cycles, also known as a time-series [3]. Would *you* want to give repeated samples of internal tissues to quantify your circadian rhythm? It is highly unlikely that you do, and such a procedure would also contradict sound medical and ethical practice.

Therefore, we and others recently developed assays to quantify internal clock time or the *phase* of human rhythms from a single biosample [4–8]. These approaches involve discovering a small panel of circadian biomarkers using machine learning (ML) and then predicting the clock phase from levels of the circadian biomarkers in a single sample [9]. However, the discovery of biomarkers for these assays still requires time series sampling of humans, albeit a small cohort under highly-controlled conditions. This makes these assays

feasible only for determining rhythm parameters in or from accessible peripheral tissue, such as blood, skin, oral mucosa or hair follicle cells [10].

How then can we quantify rhythms in human clock tissue, such as liver or kidney, that are not easily accessible? The above single sample approach, though useable in principle, still requires time series measurements to discover biomarkers. One possible source of time series data are the rhythms reconstructed from post-mortem tissue samples from a large cohort of humans [11, 12]. These studies reconstruct a rhythm across a population (a *population* rhythm) rather than a rhythm in each subject. Nevertheless, population rhythms are still valuable as long as the circadian variation exceeds the inter-individual variation, which appears to be the case [13].

This reconstruction relies critically on algorithms to assign time-labels to samples based only on the sample data, since time-labels for samples might be unavailable or unreliable (e.g., time of death). To reconstruct time-labels from samples, each sample must consist of multiple features (mRNA, proteins, etc.), typically measured using high-throughput methods. Patterns among these different features (or biomarkers) encode time. Scholz et al. [14] first proposed a type of nonlinear principal component analysis called circular PCA implemented using an autoencoder with circular units, which served as the basis for the popular method CYCLOPS and its variants [11, 15, 16]. Talamanca et al. [12] used an expectation-maximization-based algorithm that leverages the availability of multiple post-mortem tissues from a single donor for reconstruction. Finally, very recently, Larriba et al. [17] leveraged the known order of peak times of a small set of core features to reorder samples.

Analogously, single cell technologies also produce data where the state of the clock in each cell is not known, i.e., each cell is not time-labelled. In this context, Oscope [18] aggregated pairs of oscillating mRNA (features) in single-cell RNA-sequencing (scRNA-seq) data using a paired sine model. Recently, another method Tempo [19] used a variational Bayes approach that models the statistical properties of scRNA-seq data to achieve the same end.

These approaches, however, typically need to be seeded with a set of features that are known or expected to cycle for the reconstruction to function well, i.e., combined with some biological insight. Yet, this assumption on rhythmic features (based on studies on model organisms and limited human studies) may or may not hold under different conditions, e.g., when the clock is disrupted in disease. So, an approach that assigns time-labels using de novo (data-driven) rhythmic features is highly desirable. In addition, current methods rely on a two-step procedure consisting of time-label assignment followed by rhythm detection that is statistically suboptimal [20].

Here, we present a data-driven approach that simultaneously reconstructs the time ordering of data and identifies the list of rhythmic features that contribute to the reordering. These rhythmic features (circadian biomarkers) can be later used to assign time-labels to new samples not seen during training. Our approach, Cyclic Ordering with Feature Extraction (COFE), relies on a new variant of nonlinear PCA, called Sparse Cyclic PCA (scPCA), combined with a novel unsupervised cross-validation technique to fix the optimal sparsity level. We used COFE to reveal and compare for the first time in vivo circadian rhythms in 11 different human cancers based on the multi-omic data made available from The Cancer Genome Atlas (TCGA) database. Our study finds widespread rhythms at the transcript and protein levels in tumor tissue that are uncoupled with healthy clocks, but coupled with the cell cycle.

Results

COFE reconstructs time-labels and predicts rhythmic features with high accuracy

To develop a suitable approach to assign time-labels to samples, we relied on our observation that high-dimensional rhythmic time-series data (with time-labels) projected on a suitable plane produce a closed curve,

where the samples are ordered in time (within one cycle) [5, 7, 13, 21]. Sparse Cyclic PCA (scPCA) finds this projection from data without time-labels, while also enforcing a limit on the number of features used in the projection. We formulate this problem as a search for a low-dimensional elliptical approximation of the data (dimensionality reduction) that can be solved efficiently using a simple and fast iterative algorithm (Figure 1A, Material and Methods). We implemented scPCA with a novel cross-validation scheme for unsupervised learning in a Python package named COFE (https://github.com/bharathananth/COFE, Figure 1A) and benchmarked it on synthetic and biological (transcriptomic) data (SI Methods).

On synthetic data, COFE performed consistently well and its performance improved with increasing signal-to-noise ratios, larger number of samples and higher fraction of rhythmic features in the data (Figure S1). We quantified both COFE's ability to predict time-labels of samples and its ability to accurately (precision) and exhaustively (recall) identify the true rhythmic features in the data. COFE predicted the time-labels of samples in the data with high accuracy that improved with lower noise, higher number of samples and rhythmic features (Figure S1A). Although COFE achieved perfect recall only with "better" and "larger" data (like the reordering performance), COFE always achieved perfect precision (Figure S1B). In comparison, the state-of-the-art autoencoder-based method (CYCLOPS [15]) predicted time-labels with similar accuracy (Figure S1C), but was unable to achieve perfect precision-recall performance under any data configuration; CYCLOPS was sequentially combined with cosinor analysis [22] to identify rhythmic features.

On biological time-series (labelled) data, COFE's performance was only a little worse than the corresponding supervised approach (Figure S2). In these data, we only evaluated COFE's ability to predict time-labels, since the ground truth on true rhythmic features is unavailable. COFE was able to find the two-dimensional projection for mouse liver time-series data (Figure S2A,B) and predicted time-labels to within the sampling resolution of the data (Figure S2C). Time-labels of longitudinal human time-series data from blood monocytes [21](Figure S2D) and skin [13](Figure S2E) could also be predicted almost as accurately as the corresponding supervised approach (using Zeitzeiger) [5] with accuracies of 0.75h and 1.2h, respectively. Finally, COFE was able to predict both the time-labels and the underlying period of the clock on non-mammalian time-series data, such as from different strains of the malaria parasite [23] (Figure S2F).

Therefore, we are confident in COFE's capability to reconstruct underlying rhythms in high-dimensional biological datasets.

Widespread *in vivo* population rhythms in human adenocarcinomas

To investigate human rhythms under heretofore inaccessible conditions, we leveraged the high-quality transcriptomic data from human cancer biopsies assembled in the TCGA database [24]. These data are a collection of patient-derived single samples from a variety of human cancers. We restricted our attention to adenocarcinomas (ACs), which are cancers of epithelial origin that develop in organs and glands and account for 80 to 90 percent of all cancers. We further limited our study to ACs with sufficient large number of samples (> 250) within a single histological type, which ensured homogeneity of the samples and comparability of the different ACs. This resulted in a collection of 11 different ACs – ductal breast (BRCA), clear cell (KIRC) and papillary (KIRP) kidney, prostate (PRAD), ovarian (OV), uterine (UCEC), lung (LUAD), colon (COAD), thyroid (THCA), bladder (BLCA) and liver (LIHC) cancers. Samples from each AC were evenly balanced between sexes, except for the four sex-specific ACs (BRCA, PRAD, OV, UCEC) (Figure 1B).

COFE predicted time-labels for the samples from each AC independently and reproducibly. In addition to the time-labels, COFE also identifies the rhythmic features (here genes) in the data with high precision. Surprisingly, COFE reconstructed rhythmic gene expression profiles of 1000s of genes in each AC with the peak time of expression distributed throughout the day (Figures S3, S4A). The number of rhythmic genes identified by COFE varied greatly between the different ACs, ranging from about 2000 in uterine AC to over

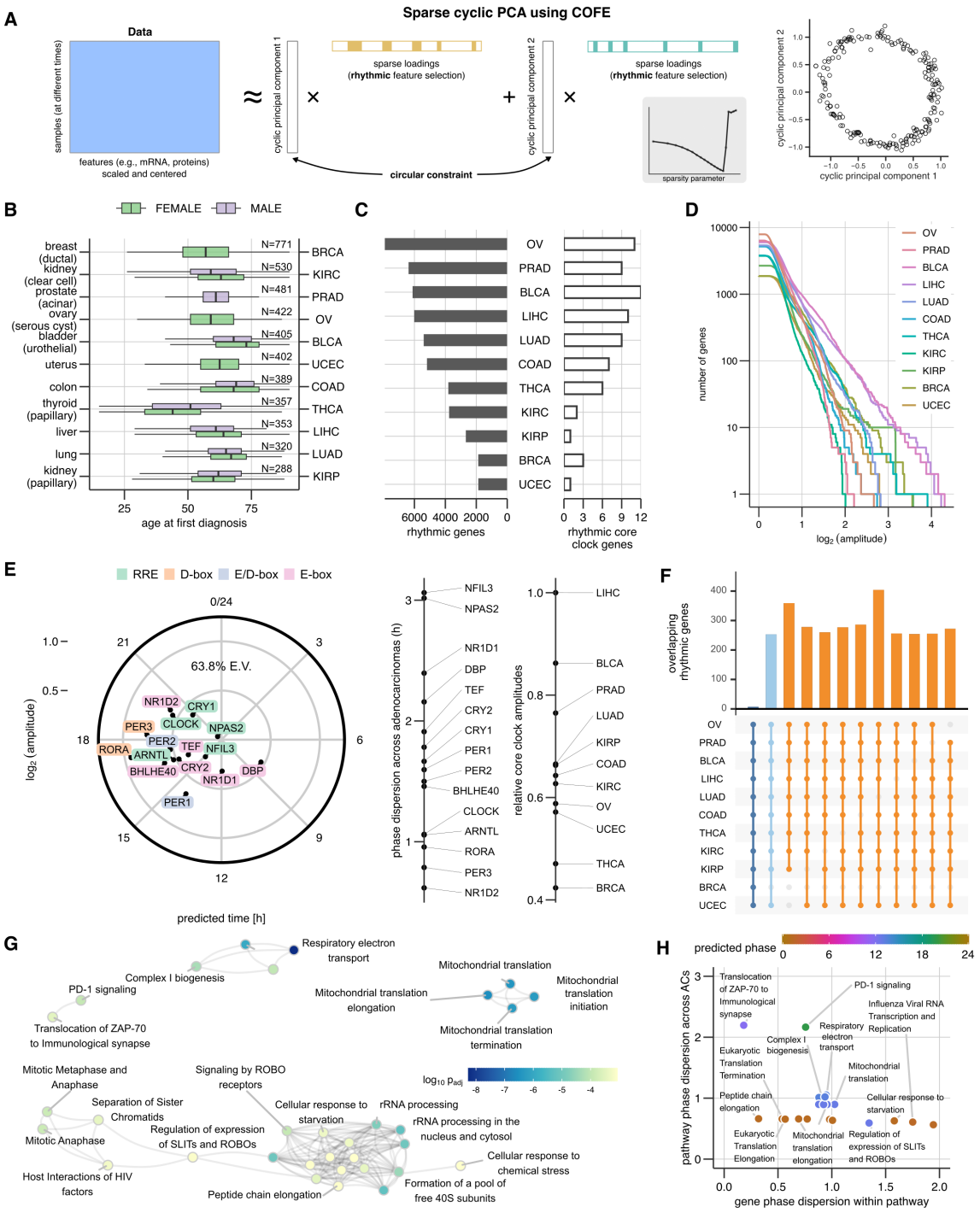

Figure 1: **Circadian population rhythms in human adenocarcinomas (ACs).** (A) Schematic of the sparse cyclic principal component analysis implemented in COFE. (B) Demographics of patients, whose biosamples provided data for training COFE. (C) Number of rhythmic genes and rhythmic core clock genes identified in each AC. (D) Cumulative distribution of the rhythmic gene amplitudes in different ACs. (E) Consensus amplitude and peak times of 15 commonly expressed core clock genes across the ACs (left). The dispersion of peak times of clock genes from the consensus arrangement across the ACs (middle). The relative strength of core clock network in different ACs. (F) Intersection of rhythmic gene sets across 11, 10 and 9 (selected) different ACs. (G) Reactome-based gene-set enrichment of the rhythmic genes. (H) Variability of peak times of genes within a pathway compared against the variability of peak times with respect to the core clock in different ACs.

6000 in ovarian AC out of about 13000 expressed genes in each AC (Figure 1C). When restricted only to the rhythmic genes with at least a two-fold peak-to-trough amplitude, the variability in the number of rhythmic genes across ACs was greatly reduced (Figure S4B). The number of rhythmic genes were independent of the number of samples in the AC ($r_s(9) = 0.04, p = 0.92$). In general, ACs with more rhythmic genes not only had larger median amplitudes, but also achieved higher amplitudes of gene expression (Figure 1D). The exception to that was BRCA, where the number of high-amplitude rhythmic genes (>4-fold peak-to-trough) was comparable to the ACs with numerous rhythmic genes.

Thus, COFE revealed widespread and heterogeneous rhythms in human cancers.

Core clock is rhythmic in human adenocarcinomas

Due to the absence of an inherent timescale in the biopsy-derived transcriptomic data, COFE alone cannot ascribe a period to the reconstructed rhythms. To infer this period, we investigated a set of 20 core clock genes that constitute the fundamental rhythm generating mechanism of the cell autonomous clock. Many of the 15 expressed core clock genes were also identified as rhythmic in ACs (Figure 1C, S4C), although the expressed core clock genes were not statistically more rhythmic than other output genes. This is indicative of a rhythmic core clock network in several ACs.

COFE's time-label assignments are consistent within each AC, but time-labels from different ACs do not share a common time reference. We therefore constructed the consensus relationships between the peak times of the 15 commonly expressed core clock genes across 11 ACs in the form of an 'eigen-core clock' using complex SVD [12]. We used this consensus core clock arrangement to define a time reference for each AC. The arrangement of peak times was reasonably well conserved across ACs (and explained 64% of inter-AC variation) (Figure 1E). This de novo analysis arranged peak times of most core clock genes in the expected order based on the response elements in their promoters (E-box $\rightarrow$ E-box/D-box $\rightarrow$ D-box $\rightarrow$ RRE) and the peak times spanned half a cycle like in a healthy functional clock [7, 12]. The core clock genes with atypical (disrupted) peak times (circadian phase) were E-box targets NR1D2, TEF, BHLHE40 and E-box/D-box target PER2, which all peaked later in the day than expected (phase delayed).

Simultaneously, the analysis also revealed the strength (amplitude) of different clock genes across the different ACs. The D-box targets RORA and PER3 had the largest amplitudes (more than threefold peak-to-trough), while RRE targets NPAS2, NFIL3 and CRY1 had the smallest amplitudes (approximately twofold peak-to-trough) (Figure 1E). The amplitudes in the consensus core clock combine the amplitudes in each AC and degree of peak time consistency of the genes across ACs. The median amplitude of each clock gene across ACs correlated strongly with consensus amplitudes ($r_p(13) = 0.86, p < 0.001$). NR1D2 not only peaked aberrantly about 8h late, but had extremely consistent peak times in the ACs. Core clock genes with higher amplitudes deviated less from the peak time in the consensus arrangement (Figure 1E, $r_s(13) = -0.86, p < 0.001$). Thus, clock genes with larger consensus amplitudes both had higher amplitudes in each individual AC and generally adhered to the consensus peak time across ACs.

We finally evaluated the relative strength of this consensus core clock arrangement in the different ACs. The core clock strength ranged from the strongest in the liver AC and weakest in the breast cancer AC, which was only half as strong. Interestingly, the ordering based on core clock network strength correlated only weakly with an ordering of ACs based on the number of rhythmic genes ($r_s(13) = 0.58, p = 0.066$).

In summary, there is a functional yet disrupted core clock that is conserved across different cancers, but the similarity of the conserved peak times of the core clock in the cancers to healthy tissue lets us conclude that these rhythms are indeed circadian.

Key cellular processes are enriched for rhythmic genes

To identify the processes enriched for rhythmic genes in different ACs, we performed gene set enrichment analysis [25] using the Reactome database [26] with genes scored based on the number of ACs in which the gene was rhythmic. Although only 7 rhythmic genes were shared between all the ACs, the ACs excluding BRCA shared 250 rhythmic genes (Figure 1F). Moreover, the overlap across different 9 out of 11 ACs again excluding BRCA was also 200–300 genes, which were all highly significant (Fisher Exact Test, $p < 0.001$). The genes with higher scores (i.e., rhythmic in more ACs) were highly enriched in Reactome pathways that fell into four Reactome pathway clusters – mitochondrial translation, respiratory electron transport, mitotic cell cycle, and adaptive immune system (Figure 1G). Respiratory electron transport and mitochondrial translation were the most highly enriched categories ($p < 10^{-6}$) making cellular metabolism the most rhythmic process in ACs.

We next quantified the consistency in peak times of rhythmic genes within the enriched Reactome pathways across different ACs. We computed two metrics that measured coherence of gene expression peak times within a pathway and the consistency of the peak arrangement across ACs (Figure 1H). This resulted in three types of pathways – pathways with highly synchronized rhythmic genes peaking at the same time with respect to the core clock network (e.g., translation), pathways with synchronized rhythmic genes but the gene expression peaks in each AC at different times with respect to the core clock (e.g., programmed cell death) and pathways with neither synchronized rhythmic gene expression nor fixed times with respect to the core clock in different ACs. The peak activity of genes in respiratory electron transport (with respect to the core clock) varied by utmost 6h across ACs and the genes within category were synchronized except for a few subunits of respiratory complex I and complex IV.

The rhythmic processes we identified are therefore fundamental and highly conserved across the cancers.

Clocks in tumors are uncoupled from clocks in healthy tissue

We thus far presented the results using the core clock in each AC as a reference to define the peak time of rhythmic gene expression. Since the core clock in different tissues in the circadian network typically are not phase-aligned, we would like to determine gene expression with respect to the network (at the organismal level). We therefore first considered the distribution of predicted time-labels in each AC over one circadian cycle. If clocks in the ACs are phase-locked to the circadian system, which in turn is entrained to the light-dark cycle, and assuming these tissue biopsies were gained from procedures conducted during the day (standard in most countries), we expect the predicted time-labels to be restricted to a fraction of the circadian cycle. Surprisingly, predicted time-labels lay throughout the circadian cycle in all ACs (Figure 2A). Although the biosamples in TCGA were obtained from several countries, the predicted time-labels were only weakly associated with the origin of the samples (MANOVA effect size $\eta^2 < 0.06$, Figure S5A), except for a strong effect ($\eta^2 = 0.26$) in COAD, and medium effect in UCEC and LIHC ($0.06 < \eta^2 < 0.14$). Simultaneously, we also verified that the sex of the subjects was not associated with time-labels in the seven non-sex-specific ACs (effect size $\eta^2 < 0.01$) and the staging of the cancers in these patients also only had an effect on the time-labels in BRCA, UCEC and KIRP, where the effect was not strong ($0.07 \leq \eta^2 \leq 0.1$).

Next, we compared the predicted time-labels for tumor tissue samples with the patient-matched healthy tissue samples that were available for 10 of the 11 ACs for a small fraction of the original patients (Figure 2B). The difference between the predicted time-labels for healthy and tumor tissue was distributed throughout the circadian cycle in most ACs suggesting that the time differences are almost random. As a negative control, we compared the difference between predicted time-labels for patient-matched tumor tissue not used to train COFE (Figure S5B). As expected, these differences clustered around zero. Based on both these observations,

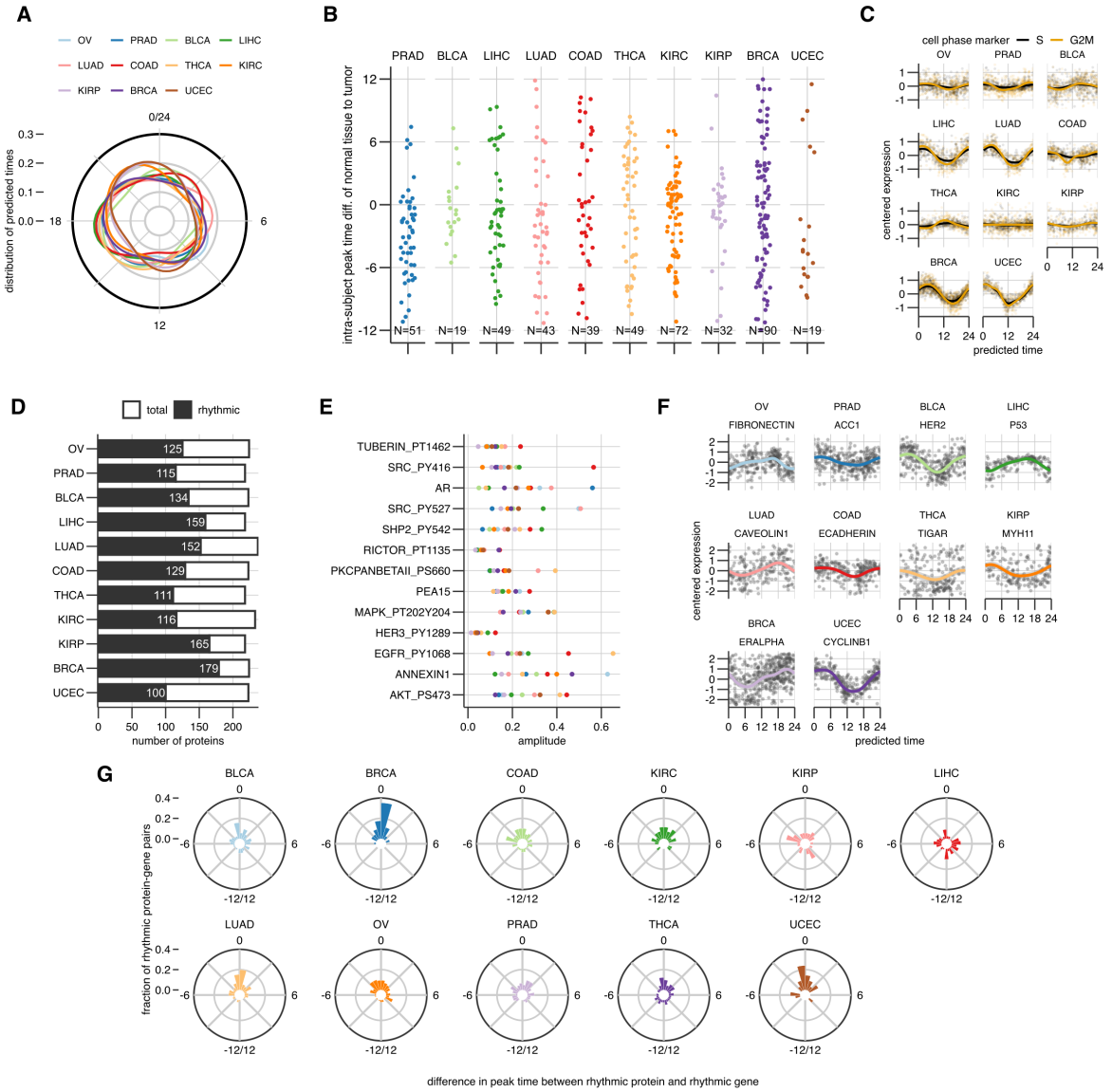

Figure 2: **Coupling of the circadian clock to the cell-cycle and the proteome.** (A) Distribution of predicted time-labels for the samples in each AC. (B) Difference in time-labels predicted by COFE between cancer and patient-matched healthy samples where both are available. (C) Scores for cell cycle phases (S and G2/M) in the different ACs. The raw and LOESS-smoothed scores per sample are shown. (D) The number of rhythmic and total measured proteins in each AC. (E) The amplitude of rhythmic protein expression (color coded as in (A)) for proteins including modified forms that are rhythmic in at least 10 of the 11 ACs. (F) The raw data and LOESS-smoothed means for the rhythmic protein with the highest amplitude in each AC. (G) Distribution of difference in peak times of expression of rhythmic proteins and their corresponding rhythmic gene.

we conclude that the clocks in tumors do not have a fixed timing relationship with clocks in healthy tissue.

Circadian clock is coupled with the cell-cycle in several ACs

Our previous analysis suggested that tumor clocks are not synchronized with the healthy clocks within the same tissue. We wondered whether the tumor circadian clock might be instead coupled to other rhythmic processes in the cell, such as the cell-cycle, which was enriched in our Reactome analysis (Figure 1G). To test this, we computed gene expression-based scores for cell-cycle phases (S and G2M) (Figure 2C)[27]. Both the scores oscillated over the circadian cycle in all ACs (cosinor, $p < 0.003$) except KIRC ($p = 0.39$). The largest amplitude rhythms in the scores were in BRCA, UCEC and LUAD. Surprisingly, the two least rhythmic ACs (BRCA and UCEC) both showed robustly rhythmic scores. The predominant cell-cycle phase of cells in these (bulk) tissue samples can be predicted using these marker scores. The ACs with rhythmic scores indeed show distinct time-of-day dependent cell-cycle phases of the constituent cells, with G1 and G2/M phases separated in time (Figure S5C). So, The two important cyclic processes in the cell are synchronized in several human cancers.

Protein expression is rhythmic, even in ACs with few rhythmic genes

Since COFE assigns time-labels to samples, we can also assign time-labels to measurements from other assays run on the same samples. The Cancer Proteome Atlas (TCPA) database contains the protein expression of a subset of the proteome (258 proteins and modified forms) important in cancer measured using Reverse Phase Protein Array on the same tumor samples in TCGA. We used a cosinor approach [22] to find the rhythmic proteins, including modified forms, in the different ACs (Table S3). 237 proteins were rhythmic in at least one AC. Moreover, 13 proteins were rhythmic in only one AC, while three were rhythmic in all. Unlike the transcriptome, at least half the measured proteins were rhythmic in all ACs (Figure 2D). Unexpectedly, BRCA had the most rhythmic proteins despite BRCA having among the fewest rhythmic genes and also the least overlap of rhythmic genes with other ACs. 13 proteins were rhythmic in 10 out of 11 ACs (Figure 2E). The nuclear receptor AR, phosphorylated tyrosine kinase SRC and phosphorylated Tuberin were rhythmic with robust amplitudes in all ACs. Activated forms of members of key signaling pathways, such as AKT, EGFR, and MAPK, were also rhythmic in almost all ACs. Nevertheless, the protein with the highest amplitude in each AC varied and sometimes also included proteins that were rhythmic in only a few of the ACs, e.g., P53 and HER2 (Figure 2F).

We next related the rhythmicity of the proteins to the rhythmicity of their underlying genes. The genes of rhythmic proteins were significantly more likely to be rhythmic than arrhythmic proteins ($p < 0.001$ for each AC, Fisher Exact test) except in the two most rhythmic ACs, OV and PRAD. Similarly, proteins of rhythmic genes were more likely to be rhythmic than proteins of arrhythmic genes ($p < 0.02$ for each AC, Fisher Exact test). We also compared the peak expression times of rhythmic proteins and their corresponding rhythmic genes (Figure 2G). The peak of rhythmic protein expression was generally synchronized with the peak of rhythmic gene expression for most ACs (average difference ± 1.5h). The notable exception here was LIHC, where the rhythmic proteins peaked on average 6h after their respective rhythmic gene. In summary, there was broad agreement between the independent analyses that identified transcript and protein rhythms.

A large, diverse array of small molecule drugs target rhythmic genes in tumor tissue

If the rhythmic genes include targets of approved and experimental cancer drugs, time-of-day changes in the target could be potentially leveraged to improve drug efficacy (chronotherapy)[28]. We identified rhythmic gene targets of a list of 481 FDA-approved, clinical candidates and putative small molecules that target key cancer

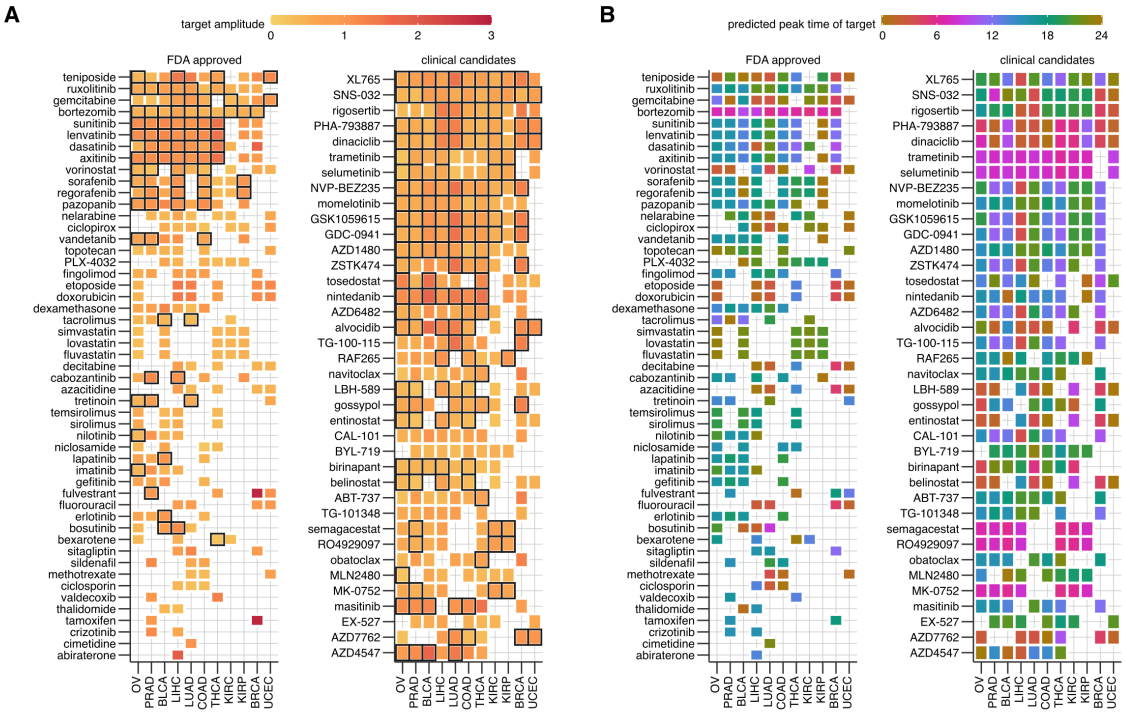

Figure 3: **Rhythmic targets of FDA-approved and clinical candidate cancer drugs** The $\log_2$-fold amplitude (A) and peak time relative to the core clock (B) of the largest rhythmic target gene in each AC among FDA approved and clinical candidate drugs that target cancer pathways or processes. All FDA-approved drugs and clinical candidates with rhythmic targets in at least 7 ACs are included. Drugs with multiple rhythmic gene targets in an AC are boxed in black.

pathways or processes from the Cancer Therapeutics Response Portal [29, 30]. 51 of the 53 FDA-approved drugs in the database targeted a rhythmic gene on average in around five different ACs and included several that had rhythmic targets in most ACs. The targets of two widely-used chemotherapeutic drugs, Teniposide and Gemcitabine, were rhythmic in all but one AC. Half the rhythmic FDA approved drug targets had better than 3-fold peak-to-trough amplitude. In 30% of cases, FDA approved drugs had more than one rhythmic target in an AC. Moreover, the distribution of FDA approved drug target amplitudes were similar across the different ACs with a few exceptions. Interestingly, targets of tamoxifen and fulvestrant in BRCA had the largest amplitudes, despite BRCA being one of the least rhythmic ACs. The peak time of the FDA approved drug target expression was in general similar across ACs in which they were rhythmic except for tacrolimus, microsporic, thalidomide and dasatinib that differed by up to 12h between ACs (Figure 3B).

91 of the 95 clinical candidates targeted rhythmic genes in some AC. Although clinical candidates also typically had rhythmic targets in about five ACs like the FDA approved drugs, five candidate drugs (PHA-793887, SNS-032, XL765, dinaciclib, rigosertib) had rhythmic targets in all ACs. Amplitudes of clinical candidate drug targets were generally smaller than the FDA approved drugs (Figure 3A). Similar to the FDA approved drugs, candidate drug targets in BRCA and UCEC (the least rhythmic ACs) had the highest median amplitude of 2-fold peak-to-trough. Clinical candidate drugs too could have targets with greatly different peak times (e.g., AZD4547, tosedostat), but the candidate drugs with rhythmic targets in all tissues had consistent phase across ACs.

Almost all putative cancer drugs in the database (248/267) targeted a rhythmic gene (Figure S6). These putative drugs were largely similar to the clinical candidates in the average number of ACs with rhythmic targets and number of drugs with rhythmic targets in all ACs. Combinations of drugs involving selumetinib

targeted rhythmic genes most broadly (Figure S6A). Putative drug targets in BRCA again and LIHC had the highest-median amplitude of 3-fold peak-to-trough. Compared to the other two drug classes, peak times of putative drugs were consistent across ACs, although they included some drugs with diversely peaking targets (selumetinib+navitoclax) (Figure S6B).

Discussion

In this work, we addressed the principal challenge to studying rhythms in humans, namely that repeated time-series sampling is necessary to study rhythms and these are difficult to obtain from most clock-relevant tissues in humans. We instead chose to reconstruct *population* rhythms from high-throughput (omics) snapshot data (one sample per individual) across a homogeneous cohort rather than quantifying rhythms in an individual, similar to previous studies [7, 12, 15]. Such snapshot data are much easier to obtain from databases or directly from biopsies and post-mortem biosamples, but lack time-labels specifying the sampling time or internal circadian time of the individual. Therefore, we developed an ML algorithm named COFE that can not only assign missing time-labels to the snapshots, but, unlike previous approaches [12, 15], can also identify the rhythmic biomarkers (features) de novo in the data. COFE identifies only truly rhythmic features as biomarkers and predicts time-labels with an accuracy comparable to supervised approaches, such as Zeitzeiger [5].

In humans, steadily accumulating epidemiological evidence has linked increased cancer risk to environmental disruption of the circadian clock in multiple cancers, including breast and prostate cancers [31, 32]. Investigations into the molecular relationships in humans between the circadian clock and cancer using cell lines have however yielded heterogeneous results [33]. Even whether cancerous cells in peripheral clock tissues possess cell autonomous rhythms is unclear and very dependent on cell lines (subtype, genetic background) and conditions [34, 35]. Moreover, studies on cell lines overlook the systemic circadian inputs that peripheral tissue receive in vivo. Further, translating insights from animal models to humans can be challenging. We therefore leveraged the snapshot transcriptomic data from large cohorts in the TCGA database to study in vivo rhythms in multiple human cancer tissues for the first time.

We discovered robust rhythms in 1000s of genes in all ACs, but to varying extents, which converged when restricted to only large amplitude rhythms. Interestingly, the hormone-dependent cancers, which are the most common cancers, accounted for the extremes in the number of rhythmic genes; with ovarian and prostate cancer containing more than 40% rhythmic genes, while breast and uterine cancer had less than 15%. A recent study [16] used CYCLOPS 2.0 to reorder the luminal A subset of BRCA data (193 samples) also found about 1100 rhythmic genes, and the high amplitude genes from that study overlapped highly significantly ($226/567, p < 0.001$) with our own. All data-driven rhythm reconstruction methods need to assume that an underlying rhythm exists. The bottleneck of time-series sampling human tissues unfortunately precludes direct validation of these rhythms. Nevertheless, the rhythms we found are bona fide patterns and not artifacts because of the following reasons: First, there are large statistically significant overlaps in the rhythmic genes discovered in each AC, which were all obtained from independent COFE analyses. Second, the predicted time-labels of the samples do not represent ordering of gene expression based on another factor, such as origin of the sample, cancer stage or sex. Third, we found consistency in expression phase between rhythmic genes and the rhythmic proteins they encode.

The TCGA data do not include direct temporal information to estimate the period of the discovered rhythms. Nevertheless, multiple lines of evidence support the conclusion that these are indeed *circadian* rhythms. First, COFE's de novo reconstruction identified many but not all known core clock genes rhythmic in all the ACs; luminal A BRCA samples too only exhibited rhythms in seven core clock genes [16]. Second, the peak expression of the core clock genes were arranged in the known sequence based on the response elements in their

promoters in all the ACs. Third, this arrangement of peak expression was mostly conserved across different independently-reordered ACs. The key difference between the core clock in these cancers and healthy tissue was a significant advance in the peak expression times of core clock genes NR1D2, PER2, TEF and DEC1. The nuclear receptor NR1D2 was almost 8h delayed compared to is homolog NR1D1, which is consistent with similar observations in live cell imaging of hepatocellular carcinomas in mouse liver (K. Padmanabhan, personal communication). This discordance between NR1D1 and NR1D2 is a possible source of the small amplitudes in RRE-targets NPAS2, NFIL3, and CRY1 in the cancers. The D-box targets, PER3 and RORA, had the largest amplitudes as observed in healthy human tissue [12] and simultaneously the smallest variability in peak times across ACs. This trend held across all clock genes that amplitudes and peak time variability were inversely correlated. Since peak time estimates are poorer with smaller amplitudes, this observation supports the view that the consensus peak time arrangement might be even more conserved. Previous work used significant changes in the correlation structure between a chosen set of core clock genes in TCGA data to claim that the clock is disrupted in cancer [36]. Another study associated differences in expression of core clock genes between tumor and healthy matched tissue in TCGA to different cancer phenotypes to infer relationship to the clock [37].

The commonly rhythmic genes in the different ACs were all involved in Reactome pathways associated with the Hallmarks of Cancer [38]. While it is known that processes, such as cellular metabolism, signalling and mitosis, are misregulated in cancer, it is unclear how robust rhythms in these processes might aid cancer progression. The peak times of some of these processes are also heterogeneously timed with respect to the core clock in each AC. Since we do not have access to the systemic circadian clock in these patients, we cannot assess whether these rhythms have evolved to be in-phase or out-of-phase with the circadian clock network. However, the lack of a stable relationship between predicted times of healthy matched tissue and tumor tissue suggests that, although the ACs appear to possess robust rhythms, these circadian rhythms are uncoupled from the systemic clock in the subject. A similar lack of correlations between predicted times of healthy and tumor samples was observed recently in a subset of BRCA samples [16]. We also inferred strong coupling between the cell cycle phases and circadian clock in many ACs, which predisposed cell cycle transitions to certain clock times. Previous in vitro studies have proposed uni- and bidirectional link between the two [39]. We can further speculate that this coupling between the cell cycle and the clock might override the coupling between the tissue clock and systemic clock leading to apparent free running clocks in tumors.

Assigning time-labels to samples based on transcriptomic data allows us to do much more than study gene expression rhythms. We also used the time-labels for multi-omic data integration to study temporal proteomic patterns. About 50% of measured protein and proteins forms were de novo rhythmic in each AC, with some proteins rhythmic in all ACs. In line with the Reactome enrichment of the rhythmic genes, activated protein forms in key cancer signalling pathways (AKT, EGFR, MAPK) were rhythmic in all ACs, which can also be inferred‘ from gene expression signatures [40]. The rhythmicity of genes and its protein product were also strongly associated. These observations further validate the reordering produced by COFE.

From a therapeutic perspective, the widespread rhythms we observed in the transcriptome and proteome of tumor tissue presents the possibility to leverage time to improve and optimize the effect of anti-cancer drugs. Remarkably, most analyzed small molecule drugs targeted a rhythmic gene in at least one AC, indicating that time of administration ought to be considered (depending on the half-life of the drug) not only in healthy tissue [11] but also in ACs. Half the currently used FDA-approved cancer drugs targeted genes with better than 3-fold peak-to-trough amplitude. The peak time of expression of the drug targets however could be quite drug- and AC-dependent, which must be considered for drugs used against multiple ACs and when treatment with drug combinations. The hormone-dependent cancers BRCA and UCEC possessed the highest amplitude in drug targets and might be prime candidates to investigate chronotherapies.

Our novel ML algorithm COFE that enabled the study of de novo rhythms in cancer has significant advan-

tages over other unsupervised methods [12, 15, 16]. These approaches first recover the time-labels and then identify rhythmic features by means of (cosinor) regression that uses the predicted time-labels. This two-step procedure using ML-predicted covariates is known to be statistically suboptimal [20]. In contrast, COFE directly provides interpretable outputs by simultaneously estimating time-labels and identifying biomarkers that contribute to the ordering. It does this via a novel variant of PCA called sparse cyclic PCA (scPCA), which can be viewed as a generalization of cosinor regression, where the sinusoidal features are estimated together with the regression coefficients. We combined this with a powerful cross-validation method for choosing the optimal sparsity level (Figure 1A), enabling robust automated identification of rhythmic biomarkers. The scPCA formulation also allows COFE to be enhanced using a range of PCA-related techniques to address issues such as confounders and missing values in the data.

It has become firmly established that daily temporal changes ought to be considered in all aspects of biology and medicine due to the ubiquity of circadian rhythms [1]. But, this endeavor is hampered by the need for time-labels on recorded data. Here we used ML to recover time-labels in a data-driven manner, which allows us to forgo the need for time-series measurements and to repurpose data without time-labels to also include this important covariate, i.e., "time". Although we showcased COFE on human data, our approach is equally effective in animal-model studies, where it can be used to better adhere to 3R principles and reduce the number of new animal studies that need to be performed. On the whole, our approach will greatly expand our ability to study circadian rhythms in humans and facilitate translation of experimental paradigms to the clinic.

Material and Methods

Theory

The p features measured at N points in time are assembled in a (data) matrix $\mathbf{X} \in \mathbb{R}^{N\times p}$. We approximate the matrix $\mathbf{X}$ along two principal components (PCs) $\mathbf{u}_1 \in \mathbb{R}^{N\times 1}$ and $\mathbf{u}_2 \in \mathbb{R}^{N\times 1}$ with corresponding loading vectors $\mathbf{v}_1 \in \mathscr{R}^{p\times 1}$ and $\mathbf{v}_2 \in \mathbb{R}^{p\times 1}$. To achieve a variance interpretation for the projection, the data matrix $\mathbf{X}$ is column-centered, i.e., $\sum_i \mathbf{X}_{ij} = 0$ for each j. As such, this is a standard rank-2 reduction of the matrix $\mathbf{X}$, like PCA:

$$\begin{aligned}\min_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1,d} &\|\mathbf{X} - d\mathbf{u}_1\mathbf{v}_1^T - d\mathbf{u}_2\mathbf{v}_2^T\|_F^2 \\ &= \max_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1} 2d\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + 2d\mathbf{u}_2^T\mathbf{X}\mathbf{v}_2 - d^2\|\mathbf{u}_1\|_2^2\|\mathbf{v}_1\|_2^2 - d^2\|\mathbf{u}_2\|_2^2\|\mathbf{v}_2\|_2^2 - 2d^2\mathbf{u}_1^T\mathbf{u}_2\mathbf{v}_1^T\mathbf{v}_2.\end{aligned}$$

To ensure that PCs represent circular processes, we enforce a circular constraint, ($u_{1i}^2 + u_{2i}^2 = 1\ \forall i \in \{1,\ldots,N\}$), i.e., to ensure that the samples in PC space lie on a circle. We term the resulting $\mathbf{u}_1$ and $\mathbf{u}_2$ *cyclic principal components*. To make this maximization well-defined, we impose l_2-regularization on the loading vectors ($\|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1$). This reduces the optimization to:

$$\begin{aligned}\max_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1,d} &\left\{2d\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + 2d\mathbf{u}_2^T\mathbf{X}\mathbf{v}_2 - d^2N - 2d^2\mathbf{u}_1^T\mathbf{u}_2\mathbf{v}_1^T\mathbf{v}_2\right\} \\ \text{s.t. } &\|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1,\ \ u_{1i}^2 + u_{2i}^2 = 1\ \ \forall i \in \{1,\ldots,N\}.\end{aligned} \tag{1}$$

The optimization in (1) is hard to solve due to the last term. Requiring orthogonality of the loading vectors or the cyclic principal components is one way to zero out that term, but that is not tractable, and potentially overly restrictive. So we simplify this problem by dropping the last term in the optimization. That is, we maximize an

upper bound on the objective function:

$$\max_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1,d} \left\{2d\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + 2d\mathbf{u}_2^T\mathbf{X}\mathbf{v}_2 - d^2N\right\} \text{ s.t. } \|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1,\ u_{1i}^2 + u_{2i}^2 = 1\ \forall i \in \{1,\ldots,N\} \tag{2}$$

This simplified equation (2) is bi-convex in **u**s and **v**s and can be solved to find a local minimum using alternating maximization [41].

Alternating maximization

We can maximize (2) by sequentially and iteratively updating the **u**s and **v**s from a suitable random initial choice of vectors until convergence using:

$$\mathbf{u}_{1i} = \frac{\mathbf{y}_{1i}}{\sqrt{\mathbf{y}_{1i}^2 + \mathbf{y}_{2i}^2}} \text{ and } \mathbf{u}_{2i} = \frac{\mathbf{y}_{2i}}{\sqrt{\mathbf{y}_{1i}^2 + \mathbf{y}_{2i}^2}}, \text{ where } \mathbf{y}_1 = \mathbf{X}\mathbf{v}_1,\ \ \mathbf{y}_2 = \mathbf{X}\mathbf{v}_2 \tag{3}$$

$$\mathbf{v}_1 = \frac{\mathbf{X}^T\mathbf{u}_1}{\|\mathbf{X}^T\mathbf{u}_1\|_2} \text{ and } \mathbf{v}_2 = \frac{\mathbf{X}^T\mathbf{u}_2}{\|\mathbf{X}^T\mathbf{u}_2\|_2} \tag{4}$$

$$d = \frac{\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + \mathbf{u}_2^T\mathbf{X}\mathbf{v}_2}{N}. \tag{5}$$

Alternating maximization with regularization

To identify rhythmic features, we constrain the loading vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ to be suitably sparse, so that non-zero loading coefficients correspond to rhythmic features. To keep the optimization tractable, sparsity is enforced via an l_1 constraint with a hyperparameter s. The resulting optimization, which we call *sparse cyclic PCA* (scPCA), is:

$$\begin{aligned}&\max_{\mathbf{u}_1,\mathbf{v}_1,\mathbf{u}_1,\mathbf{v}_1,d} \left\{2d\mathbf{u}_1^T\mathbf{X}\mathbf{v}_1 + 2d\mathbf{u}_2^T\mathbf{X}\mathbf{v}_2 - d^2N\right\} \\ &\quad \text{s.t. } \|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1,\ \|\mathbf{v}_1\|_1 \le s,\ \|\mathbf{v}_2\|_1 \le s,\ u_{1i}^2 + u_{2i}^2 = 1\ \forall i \in \{1,\ldots,N\}\end{aligned} \tag{6}$$

This optimization (6) is solved similarly to the non-regularized problem (2). Only the update equation (4) for the **v**s differs, as we have to jointly impose simultaneous ℓ_1 and ℓ_2 constraints, which is easily achieved in a single step [42].

Recovering the time-labels The time-labels for each sample $i \in \{1,\ldots,N\}$ in the interval $[0,1]$ are computed as $\frac{1}{2\pi}\arctan\left(\frac{u_{2i}}{u_{1i}}\right)$.

Implementation - Cyclic Ordering with Feature Extraction (COFE)

We implemented scPCA with the cross-validation in a Python package named COFE. The scPCA algorithm described above needs to address some additional issues before it can be used for rhythm profiling without a time series.

(i) We standardize the features to ensure that they are weighted equally with respect to the regularization constraints.

(ii) The solution to the optimization, found via alternating maximization, only guarantees a local maximum. We therefore solve scPCA from multiple initial conditions (*restarts*) for each level of sparsity and pick the solution with the best score.

(iii) The alternating maximization optimizes the simplified objective function in (6) rather than (1). We therefore score solutions with the complete objective in (1).

(iv) How do we choose a level of sparsity s for the regularized problem (6)? Since scPCA is an unsupervised learning technique, we do not have access to the true time labels to help us choose s. Hence, we implement a form of cross-validation for unsupervised learning [43]. We delete random sets of elements from the data matrix and apply scPCA on this data with missing values to obtain a low rank approximation. The error between these deleted values and their prediction from the low rank approximation is used as the cross-validation (CV) error metric, since it measures the ability to impute the missing values. We systematically perform repeated K-fold CV for a range of s values chosen by the user. The CV error shows a sharp phase transition at the optimal s when all the rhythmic features above the noise have been detected (Figure 1A).

Acknowledgements

B. A. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) grant AN 1553/2-1. The authors declare no competing interests.

References

1. Allada, R. & Bass, J. Circadian Mechanisms in Medicine. en. *New England Journal of Medicine* **384** (ed Longo, D. L.) 550–561. ISSN: 0028-4793, 1533-4406. http://www.nejm.org/doi/10.1056/NEJMra1802337 (2021) (Feb. 2021).

2. Koronowski, K. B. & Sassone-Corsi, P. Communicating clocks shape circadian homeostasis. en. *Science* **371,** eabd0951. ISSN: 0036-8075, 1095-9203. https://www.sciencemag.org/lookup/doi/10.1126/science.abd0951 (2021) (Feb. 2021).

3. Baum, L., Johns, M., Poikela, M., Möller, R., Ananthasubramaniam, B. & Prasser, F. Data integration and analysis for circadian medicine. en. *Acta Physiologica.* ISSN: 1748-1708, 1748-1716. https://onlinelibrary.wiley.com/doi/10.1111/apha.13951 (2023) (Feb. 2023).

4. Laing, E. E., Möller-Levet, C. S., Poh, N., Santhi, N., Archer, S. N. & Dijk, D.-J. Blood transcriptome based biomarkers for human circadian phase. en. *eLife* **6.** ISSN: 2050-084X. https://elifesciences.org/articles/20214 (2017) (Feb. 2017).

5. Hughey, J. J. Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. *Genome Medicine* **9,** 19. ISSN: 1756-994X. http://dx.doi.org/10.1186/s13073-017-0406-4 (2017) (2017).

6. Braun, R., Kath, W. L., Iwanaszko, M., Kula-Eversole, E., Abbott, S. M., Reid, K. J., Zee, P. C. & Allada, R. Universal method for robust detection of circadian state from gene expression. en. *Proceedings of the National Academy of Sciences* **115,** E9247–E9256. ISSN: 0027-8424, 1091-6490. http://www.pnas.org/lookup/doi/10.1073/pnas.1800314115 (2018) (Sept. 2018).

7. Wu, G., Ruben, M. D., Francey, L. J., Smith, D. F., Sherrill, J. D., Oblong, J. E., Mills, K. J. & Hogenesch, J. B. A population-based gene expression signature of molecular clock phase from a single epidermal sample. en. *Genome Medicine* **12,** 73. ISSN: 1756-994X. https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00768-9 (2021) (Dec. 2020).

8. Vlachou, D., Veretennikova, M., Usselmann, L., Vasilyev, V., Ott, S., Bjarnason, G. A., Dallmann, R., Levi, F. & Rand, D. A. TimeTeller: A tool to probe the circadian clock as a multigene dynamical system. en. *PLOS Computational Biology* **20** (ed Csikász-Nagy, A.) e1011779. ISSN: 1553-7358. https://dx.plos.org/10.1371/journal.pcbi.1011779 (2024) (Feb. 2024).

9. Dijk, D.-J. & Duffy, J. F. Novel approaches for assessing circadian rhythmicity in humans: a review. *Journal of Biological Rhythms* **35,** 421–438 (2020).

10. Crnko, S., Schutte, H., Doevendans, P. A., Sluijter, J. P. G. & van Laake, L. W. Minimally Invasive Ways of Determining Circadian Rhythms in Humans. en. *Physiology* **36,** 7–20. ISSN: 1548-9213, 1548-9221. https://journals.physiology.org/doi/10.1152/physiol.00018.2020 (2021) (Jan. 2021).

11. Ruben, M. D., Wu, G., Smith, D. F., Schmidt, R. E., Francey, L. J., Lee, Y. Y., Anafi, R. C. & Hogenesch, J. B. A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. en. *Science Translational Medicine* **10,** eaat8806. ISSN: 1946-6234, 1946-6242. http://stm.sciencemag.org/lookup/doi/10.1126/scitranslmed.aat8806 (2018) (Sept. 2018).

12. Talamanca, L., Gobet, C. & Naef, F. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans. en. *Science* **379,** 478–483. ISSN: 0036-8075, 1095-9203. https://www.science.org/doi/10.1126/science.add0846 (2023) (Feb. 2023).

13. Del Olmo, M., Spörl, F., Korge, S., Jürchott, K., Felten, M., Grudziecki, A., de Zeeuw, J., Nowozin, C., Reuter, H., Blatt, T., Herzel, H., Kunz, D., Kramer, A. & Ananthasubramaniam, B. Inter-layer and inter-subject variability of diurnal gene expression in human skin. en. *NAR Genomics and Bioinformatics* **4,** lqac097. ISSN: 2631-9268. https://academic.oup.com/nargab/article/doi/10.1093/nargab/lqac097/6965977 (2023) (Oct. 2022).

14. Scholz, M. en. in *Bioinformatics Research and Development* (eds Hochreiter, S. & Wagner, R.) Series Title: Lecture Notes in Computer Science, 38–47 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007). ISBN: 978-3-540-71232-9. http://link.springer.com/10.1007/978-3-540-71233-6_4 (2022).

15. Anafi, R. C., Francey, L. J., Hogenesch, J. B. & Kim, J. CYCLOPS reveals human transcriptional rhythms in health and disease. en. *Proceedings of the National Academy of Sciences* **114,** 5312–5317. ISSN: 0027-8424, 1091-6490. http://www.pnas.org/lookup/doi/10.1073/pnas.1619320114 (2017) (May 2017).

16. Li, S.-Y., Hammarlund, J. A., Wu, G., Lian, J.-W., Howell, S. J., Clarke, R. B., Adamson, A. D., Gonçalves, C. F., Hogenesch, J. B., Anafi, R. C. & Meng, Q.-J. Tumor circadian clock strength influences metastatic potential and predicts patient prognosis in luminal A breast cancer. en. *Proceedings of the National Academy of Sciences* **121,** e2311854121. ISSN: 0027-8424, 1091-6490. https://pnas.org/doi/10.1073/pnas.2311854121 (2024) (Feb. 2024).

17. Larriba, Y., Mason, I. C., Saxena, R., Scheer, F. A. J. L. & Rueda, C. CIRCUST: A novel methodology for temporal order reconstruction of molecular rhythms; validation and application towards a daily rhythm gene expression atlas in humans. en. *PLOS Computational Biology* **19** (ed Hong, C. I.) e1011510. ISSN: 1553-7358. https://dx.plos.org/10.1371/journal.pcbi.1011510 (2024) (Sept. 2023).

18. Leng, N., Chu, L.-F., Barry, C., Li, Y., Choi, J., Li, X., Jiang, P., Stewart, R. M., Thomson, J. A. & Kendziorski, C. Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. en. *Nature Methods* **12,** 947–950. ISSN: 1548-7091, 1548-7105. http://www.nature.com/articles/nmeth.3549 (2020) (Oct. 2015).

19. Auerbach, B. J., FitzGerald, G. A. & Li, M. Tempo: an unsupervised Bayesian algorithm for circadian phase inference in single-cell transcriptomics. en. *Nature Communications* **13,** 6580. ISSN: 2041-1723. https://www.nature.com/articles/s41467-022-34185-w (2022) (Nov. 2022).

20. Wang, S., McCormick, T. H. & Leek, J. T. Methods for correcting inference based on outcomes predicted by machine learning. en. *Proceedings of the National Academy of Sciences* **117,** 30266–30275. ISSN: 0027-8424, 1091-6490. https://pnas.org/doi/full/10.1073/pnas.2001238117 (2023) (Dec. 2020).

21. Wittenbrink, N., Ananthasubramaniam, B., Münch, M., Koller, B., Maier, B., Weschke, C., Bes, F., de Zeeuw, J., Nowozin, C., Wahnschaffe, A., Wisniewski, S., Zaleska, M., Bartok, O., Ashwal-Fluss, R., Lammert, H., Herzel, H., Hummel, M., Kadener, S., Kunz, D. & Kramer, A. High-accuracy determination of internal circadian time from a single blood sample. en. *Journal of Clinical Investigation* **128,** 3826–3839. ISSN: 0021-9738, 1558-8238. https://www.jci.org/articles/view/120874 (2018) (Aug. 2018).

22. Cornelissen, G. Applications of cosinor rhythmometry in pharmacology. en. *Journal of Pharmacokinetics and Pharmacodynamics* **48,** 339–359. ISSN: 1567-567X, 1573-8744. https://link.springer.com/10.1007/s10928-021-09748-x (2021) (June 2021).

23. Smith, L. M., Motta, F. C., Chopra, G., Moch, J. K., Nerem, R. R., Cummins, B., Roche, K. E., Kelliher, C. M., Leman, A. R., Harer, J., Gedeon, T., Waters, N. C. & Haase, S. B. An intrinsic oscillator drives the blood stage cycle of the malaria parasite Plasmodium falciparum. eng. *Science (New York, N.Y.)* **368,** 754–759. ISSN: 1095-9203 (May 2020).

24. *TCGA Research Network* https://www.cancer.gov/tcga.

25. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. en. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 15545–15550. ISSN: 0027-8424, 1091-6490. http://www.pnas.org/content/102/43/15545 (2014) (Oct. 2005).

26. Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. en. *Nucleic Acids Research* **52,** D672–D678. ISSN: 0305-1048, 1362-4962. https://academic.oup.com/nar/article/52/D1/D672/7369850 (2024) (Jan. 2024).

27. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. en. *Science* **352,** 189–196. ISSN: 0036-8075, 1095-9203. https://www.science.org/doi/10.1126/science.aad0501 (2024) (Apr. 2016).

28. Ballesta, A., Innominato, P. F., Dallmann, R., Rand, D. A. & Lévi, F. A. Systems Chronotherapeutics. en. *Pharmacological Reviews* **69** (ed Watts, S. W.) 161–199. ISSN: 0031-6997, 1521-0081. http://pharmrev.aspetjournals.org/lookup/doi/10.1124/pr.116.013441 (2018) (Apr. 2017).

29. *Cancer Therapeutics Response Portal v2.1* https://portals.broadinstitute.org/ctrp.v2.1/.

30. Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. en. *Nature Chemical Biology* **12,** 109–116. ISSN: 1552-4450, 1552-4469. https://www.nature.com/articles/nchembio.1986 (2024) (Feb. 2016).

31. Masri, S. & Sassone-Corsi, P. The emerging link between cancer, metabolism, and circadian rhythms. en. *Nature Medicine* **24,** 1795–1803. ISSN: 1078-8956, 1546-170X. https://www.nature.com/articles/s41591-018-0271-8 (2024) (Dec. 2018).

32. Shafi, A. A. & Knudsen, K. E. Cancer and the Circadian Clock. en. *Cancer Research* **79,** 3806–3814. ISSN: 0008-5472, 1538-7445. https://aacrjournals.org/cancerres/article/79/15/3806/638298/Cancer-and-the-Circadian-ClockCancer-and-the (2023) (Aug. 2019).

33. Sancar, A. & Van Gelder, R. N. Clocks, cancer, and chronochemotherapy. en. *Science* **371,** eabb0738. ISSN: 0036-8075, 1095-9203. https://www.science.org/doi/10.1126/science.abb0738 (2024) (Jan. 2021).

34. Relógio, A., Thomas, P., Medina-Pérez, P., Reischl, S., Bervoets, S., Gloc, E., Riemer, P., Mang-Fatehi, S., Maier, B., Schäfer, R., Leser, U., Herzel, H., Kramer, A. & Sers, C. Ras-Mediated Deregulation of the Circadian Clock in Cancer. en. *PLoS Genetics* **10** (ed Levi, F.) e1004338. ISSN: 1553-7404. https://dx.plos.org/10.1371/journal.pgen.1004338 (2024) (May 2014).

35. Ector, C., Schmal, C., Didier, J., De Landtsheer, S., Finger, A.-M., Müller-Marquardt, F., Schulte, J., Sauter, T., Keilholz, U., Herzel, H., Kramer, A. & Granada, A. E. *Time-of-day effects of drugs revealed by high-throughput deep phenotyping* en. preprint (Cancer Biology, Dec. 2023). http://biorxiv.org/lookup/doi/10.1101/2023.11.30.569380 (2024).

36. Shilts, J., Chen, G. & Hughey, J. J. Evidence for widespread dysregulation of circadian clock progression in human cancer. en. *PeerJ* **6,** e4327. ISSN: 2167-8359. https://peerj.com/articles/4327 (2018) (Jan. 2018).

37. Ye, Y., Xiang, Y., Ozguc, F. M., Kim, Y., Liu, C.-J., Park, P. K., Hu, Q., Diao, L., Lou, Y., Lin, C., Guo, A.-Y., Zhou, B., Wang, L., Chen, Z., Takahashi, J. S., Mills, G. B., Yoo, S.-H. & Han, L. The Genomic Landscape and Pharmacogenomic Interactions of Clock Genes in Cancer Chronotherapy. en. *Cell Systems* **6,** 314–328.e2. ISSN: 24054712. https://linkinghub.elsevier.com/retrieve/pii/S2405471218300504 (2024) (Mar. 2018).

38. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. en. *Cell* **144,** 646–674. ISSN: 00928674. https://linkinghub.elsevier.com/retrieve/pii/S0092867411001279 (2024) (Mar. 2011).

39. Farshadi, E., Van Der Horst, G. T. & Chaves, I. Molecular Links between the Circadian Clock and the Cell Cycle. en. *Journal of Molecular Biology* **432,** 3515–3524. ISSN: 00222836. https://linkinghub.elsevier.com/retrieve/pii/S0022283620302801 (2023) (May 2020).

40. Winkler, E., Ananthasubramaniam, B. & Herzel, H. *Gene expression signatures predict circadian rhythms in oncogenic pathways* en. preprint (Systems Biology, Nov. 2022). http://biorxiv.org/lookup/doi/10.1101/2022.11.02.514919 (2023).

41. Richtárik, P., Jahani, M., Ahipaşaoğlu, S. D. & Takáč, M. Alternating maximization: unifying framework for 8 sparse PCA formulations and efficient parallel codes. en. *Optimization and Engineering* **22,** 1493–1519. ISSN: 1389-4420, 1573-2924. https://link.springer.com/10.1007/s11081-020-09562-3 (2022) (Sept. 2021).

42. Guillemot, V., Beaton, D., Gloaguen, A., Löfstedt, T., Levine, B., Raymond, N., Tenenhaus, A. & Abdi, H. A constrained singular value decomposition method that integrates sparsity and orthogonality. en. *PLOS ONE* **14** (ed Peddada, S. D.) e0211463. ISSN: 1932-6203. https://dx.plos.org/10.1371/journal.pone.0211463 (2021) (Mar. 2019).

43. Patrick O. Perry. *Cross-Validation for Unsupervised Learning* Ph.D. (Stanford University, Sept. 2009). 10.48550/arXiv.0909.3052.

44. Wilks, C., Zheng, S. C., Chen, F. Y., Charles, R., Solomon, B., Ling, J. P., Imada, E. L., Zhang, D., Joseph, L., Leek, J. T., Jaffe, A. E., Nellore, A., Collado-Torres, L., Hansen, K. D. & Langmead, B. recount3: summaries and queries for large-scale RNA-seq expression and splicing. en. *Genome Biology* **22,** 323. ISSN: 1474-760X. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02533-6 (2024) (Dec. 2021).

45. Pan, Y., Ballance, H., Meng, H., Gonzalez, N., Kim, S.-M., Abdurehman, L., York, B., Chen, X., Schnytzer, Y., Levy, O., Dacso, C. C., McClung, C. A., O'Malley, B. W., Liu, S. & Zhu, B. 12-h clock regulation of genetic information flow by XBP1s. eng. *PLoS biology* **18,** e3000580. ISSN: 1545-7885 (Jan. 2020).