



Platform-independent estimation of human physiological time from single blood samples

Yitong Huang^{a,b} and Rosemary Braun^{a,b,c,d,e,f,1}

Edited by Joseph Takahashi, The University of Texas Southwestern Medical Center, Dallas, TX; received May 15, 2023; accepted October 25, 2023

Abundant epidemiological evidence links circadian rhythms to human health, from heart disease to neurodegeneration. Accurate determination of an individual's circadian phase is critical for precision diagnostics and personalized timing of therapeutic interventions. To date, however, we still lack an assay for physiological time that is accurate, minimally burdensome to the patient, and readily generalizable to new data. Here, we present TimeMachine, an algorithm to predict the human circadian phase using gene expression in peripheral blood mononuclear cells from a single blood draw. Once trained on data from a single study, we validated the trained predictor against four independent datasets with distinct experimental protocols and assay platforms, demonstrating that it can be applied generalizably. Importantly, TimeMachine predicted circadian time with a median absolute error ranging from 1.65 to 2.7 h, regardless of systematic differences in experimental protocol and assay platform, without normalizing the data or retraining the predictor. This feature enables it to be flexibly applied to both new samples and existing data without limitations on the transcriptomic profiling technology (microarray, RNAseq). We benchmark TimeMachine against competing approaches and identify the algorithmic features that contribute to its performance.

circadian rhythms | machine learning | transcriptomics | cross-platform prediction

The circadian rhythm is an endogenous, evolutionarily conserved timekeeping system that regulates biological processes in time with the Earth's daily cycle. In mammals, the rhythm is generated in each cell via a transcription–translation feedback loop, which in turn orchestrates 24-h transcriptional rhythms in nearly half of all genes in a tissue-specific manner (1). Numerous studies have demonstrated that dysregulation of circadian rhythms is associated with various health issues, including obesity, diabetes, cardiovascular disease, and cancer (2–4). In the past decade, it has become clear that aligning drug dosing time with the internal circadian clock can enhance the effectiveness of treatments and reduce side effects, particularly for cancer (5–7). However, the current gold standard measure of circadian phase, dim-light melatonin onset (DLMO), is time-consuming and expensive to obtain, requiring hourly saliva/plasma samples over a 24-h period. The expense of consecutive sample collection and the availability of a sleep clinic are major barriers to implementing circadian measures in research studies and clinical settings. The development of simple measurement of circadian phase is necessary to facilitate the translation of circadian-based medicine on a population scale.

Transcriptomic profiling technologies and machine learning algorithms offer a promising approach to assess the circadian phase by using gene expression in blood samples as a readout of the endogenous circadian rhythm. Prior to the availability of human circadian datasets, methods were developed to predict time of day using a single-time-point expression profile for mouse tissues. “Molecular Timetable,” developed by Ueda et al., used 168 time-indicating genes to estimate peak time of cycling genes via cosine fitting in mouse liver tissues (8). While this approach performed well in mice under laboratory conditions, the need for an accurate biomarker applicable to humans spurred the development of more sophisticated algorithms (9–11).

The first of these, Hughey et al.'s “ZeitZeiger” is a supervised machine learning method to predict time-of-day from gene expression data using smooth periodic splines and sparse principal components (12). Originally developed using the same mouse data as Molecular Timetable, it was later extended to human data (9) and optimized specifically using monocytes assayed with the NanoString platform (13). More recently, Laing et al. developed a multivariate model based on partial least squares regression (PLSR) that takes a single sample of gene expression profile as input to predict the melatonin phase in humans (10). This PLSR-based model was trained and validated on a mixture of two

Significance

The importance of circadian rhythm in health is evident in studies from metabolic disorders to Alzheimer's disease. However, translating these observations to the clinic remains stymied due to the burden of measuring physiological time. Methods to assess physiological time using blood biomarkers address this issue, but they must be accurate and generalizable across protocols and platforms. Here, we present TimeMachine, an algorithm that can estimate the circadian phase from a single blood sample. Validation on four distinct datasets shows TimeMachine accurately recovers phase estimates from a single-timepoint gene expression profile of human peripheral blood mononuclear cells across varied protocols and technologies, an important advance over current methods. This algorithm offers a feasible approach for incorporating circadian biomarkers in research and clinical care.

Author contributions: Y.H. and R.B. designed research; Y.H. performed research; Y.H. and R.B. contributed new reagents/analytic tools; Y.H. analyzed data; and Y.H. and R.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: rbraun@northwestern.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2308114120/-DCSupplemental>.

Published January 8, 2024.

human circadian gene expression datasets profiled by the same microarray platform and reported to predict 54% of the samples with an error under 2 h.

However, each of these methods has limitations that hampered their generalizability. For example, when ZeitZeiger is applied to a new batch of data, the training data and the new data are renormalized with batch correction, followed by retraining the machine after the batch correction is done. This can present challenges for interpretability, as the predictor itself may differ from batch to batch. Moreover, application of batch correction necessitates multiple samples from the same batch and thus cannot be used for a single new sample (as would be expected in clinical application). While the PLSR-based model developed by Laing et al. does not require batch correction, it instead relies on quantile normalization to the same reference array of training data. This means that new data must measure the same features as the training data ($\sim 10^4$ transcripts), including genes that are not part of the predictor itself (which comprises $\sim 10^2$ genes). It has also been shown that this normalization may not adequately address systematic batch differences, making the predictor inaccurate when applied across platforms (11).

To address the need for a generalizable predictor that could be used without batch correction or retraining, Braun et al. proposed “TimeSignature,” an elastic-net regression model with a within-subject normalization step (11, 14). TimeSignature was validated against three different studies using distinct microarrays and RNA-sequencing platforms and demonstrated accuracy across platforms and studies without the need for batch correction, retraining, or fitting platform-specific models. Its generalizability across datasets was shown to significantly outperform ZeitZeiger and the one sample PLSR-based method. However, TimeSignature requires two samples from each subject to compute the within-subject normalization, and the two samples need to be at least 8 h apart for peak accuracy. While still simpler than assaying DLMO, this requirement limits prospective application to settings where two samples can be collected and cannot be applied retrospectively to previously collected single samples.

In this study, we introduce a single-timepoint algorithm for predicting circadian phase in humans, emphasizing its simplicity and generalizability. The TimeMachine algorithm offers two variants—ratio TimeMachine (rTM) and Z-score TimeMachine (zTM). Both variants only require gene expression measurements for 37 genes from a single blood draw and achieve similar performance. No constraints are placed on the assay (microarray, RNAseq, etc). To evaluate its performance, we trained and tested the model on one human circadian gene expression dataset (divided subject-wise into training and testing subsets) and validated it on three independent datasets with different microarray and RNA-seq platforms. Our results indicate that, once trained, the TimeMachine predictor can be applied directly to new data without retraining the model or any batch corrections. TimeMachine achieves a median absolute error of ~ 2.5 h, performing as well or better than other single-sample methods. It utilizes fewer genes and is only around 20 to 40 min less accurate than methods requiring two draws. We also investigated factors associated with uncertainties in TimeMachine’s predictions, such as the amplitude of the regression model and sampling time, which offer valuable insights for future applications.

Results

TimeMachine Algorithm. The framework of TimeMachine starts with a feature selection and within-sample rescaling step, followed by fitting the predictor, and finally applying it to samples from separate studies. To demonstrate the performance of the

TimeMachine algorithm, we acquired four published datasets of human whole-blood transcriptome profiles from the NCBI Gene Expression Omnibus (GEO) repository: GSE39445 (TrTe) (15), GSE48113 (V1) (16), GSE56931 (V2) (17), and GSE113883 (V3) (11). Details of each dataset can be found in Materials and Methods. We will refer to these datasets as TrTe, V1, V2, and V3 throughout the manuscript. Only genes in common to all four datasets were retained for analysis, yielding a total of 7,615 genes.

Identifying predictor genes. The first step of the algorithm is to identify genes that contain phase information and would thus be promising biomarkers of time.

Existing methods, including ZeitZeiger (9), the PLSR-based method (10), and TimeSignature (11), offer distinct strategies for identifying genes with rhythmic patterns that can predict circadian phases in human blood transcriptomic data in PBMCs.

We applied the aforementioned algorithms using the provided code from the original authors to our training data, which comprises half of the subjects selected at random from TrTe. This process led to the identification of 24, 100, and 41 time-indicating genes, respectively (as shown in *SI Appendix, Table S1*). Because previous findings have demonstrated that the predictor genes identified by these algorithms rarely overlap, even when employed on the same dataset (10), we consider the union of these genes—135 genes in total (*SI Appendix, Table S1*)—as candidate phase markers.

Of these, we expect that genes that exhibit the strongest cycling patterns will better reflect circadian time. To this end, we identify those that show the most robust cycling via JTK_Cycle analysis (18, 19). Thirty-seven genes yield JTK_Cycle $P < 0.1$ (*SI Appendix, Table S2*). These 37 genes are used as the inputs to the TimeMachine predictor.

Sample-wise normalization. The diverse transcriptomic profiling platforms often generate gene expression data on different scales. This poses a significant challenge for developing a predictor that performs well independently of the specific profiling platform. Key techniques, such as batch correction, quantile normalization, or the subject-wise normalization used in TimeSignature, are implemented to ensure that the data are expressed on a comparable and consistent scale. Here, we propose two normalization approaches: 1) pairwise gene ratios and 2) Z-score transformation. Both approaches are based on the conjecture that the relative expression of the predictor genes, rather than their absolute magnitudes, are the biologically relevant features, and that these will be better preserved across platforms than the absolute expression values. As demonstrated further below, the two techniques offer similar performance. To distinguish between the normalization schemes, we will refer to the algorithm as ratio TimeMachine (rTM) when employing pairwise gene ratios, and as Z-score TimeMachine (zTM) when applying Z-score transformation in the subsequent sections.

Pairwise gene ratios. We denote the log₂ expression of gene j in sample i collected at time t_i as $x_{i,j}$, where $j = 1, \dots, 37$. For the given sample i , the difference between the log₂ expressions (i.e., the ratio of a pair of gene expressions) is:

$$z_{i,jk} = x_{i,j} - x_{i,k},$$

where $j = 1, \dots, 36$ and $k > j$. The resulting vector $\vec{z}_i$ is thus a 1×666 vector representing the ratio of gene expressions at time t_i .

Z-score transformation. For a given sample i collected at time t_i , the log₂ expression of 37 genes are denoted as $\vec{x}_i$ (a 1×37 vector). Calculating the Z-score transformation on the sample yields

$$\vec{z}_i = \frac{\vec{x}_i - \bar{\vec{x}}_i}{s_i},$$

where $i = 1, \dots, n$; $\bar{\vec{x}}_i$ is the sample mean; and s_i represents the sample SD.

Notably, no batch correction or other normalization is conducted, so that the predictor can be directly applied to future data simply by selecting these 37 genes and applying the ratio or Z-score sample-wise normalization.

Fitting the predictor. Both ratio TimeMachine and Z-score TimeMachine use a bivariate regression model with elastic net regularization to predict physiological time as a function of gene expression. Here, the target (dependent) variables are the Cartesian coordinates corresponding to the angle on a 24-h clock,

$$Y_{(N \times 2)} = \begin{bmatrix} \cos\left(\frac{2\pi}{24} t_i\right) \\ \sin\left(\frac{2\pi}{24} t_i\right) \end{bmatrix}^T = \beta_0 + Z_{(N \times P)} \beta_{(P \times 2)} + \eta_{(N \times 2)}, \quad [1]$$

where t_i is the physiological time (e.g. hours since DLMO, otherwise known as “melatonin phase”) for sample i ; Z is an N (observations) $\times$ P (features) matrix after the normalization described above ($P = 37$ for Z-score TimeMachine; $P = 666$ for ratio TimeMachine); and η is a Gaussian error term.

To obtain a more accurate and parsimonious model, we find coefficients β that minimize the cost function with elastic net regularization (20):

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{(P+1) \times 2}} \left[\frac{1}{2N} \sum_{i=1}^N \|\vec{y}_i - \beta_0 - \beta^T \vec{z}_i\|_F^2 + \lambda \left((1 - \alpha) \|\beta\|_F^2 / 2 + \alpha \sum_{j=1}^P \sqrt{\beta_{j1}^2 + \beta_{j2}^2} \right) \right], \quad [2]$$

where λ is the regularization penalty and α balances the L_1 and L_2 penalty terms. The value of α ranges from 0 to 1. Both α and λ are tuned by cross-validation within the training data.

Once the model coefficients are fit, the model can be applied to predict $\hat{Y} = [\hat{y}_{i1}, \hat{y}_{i2}]$ from a new sample. This output is then transformed back into a time measurement via

$$\hat{t}_i = \frac{24}{2\pi} \left(\text{atan2}(\hat{y}_{i1}, \hat{y}_{i2}) \bmod 2\pi \right). \quad [3]$$

The absolute prediction error for sample i can then be calculated as

$$\epsilon_i = \min \left(|\hat{t}_i - t_i|, 24 - |\hat{t}_i - t_i| \right). \quad [4]$$

Validation: Cross-Study and Cross-Platform Accuracy. To assess the performance of ratio TimeMachine and Z-score TimeMachine across varied datasets and platforms, we randomly selected half of the subjects from TrTe for training. We then used the remaining data from TrTe, along with the two independent datasets with melatonin data (V1 and V3), to compare algorithms’ accuracy with respect to the true melatonin phase—specifically, the time since DLMO (Fig. 1). (Here, we exclude dataset V2, as it did not include publicly accessible melatonin data; we return to it further below.)

Within the training set, we used 10-fold cross-validation to tune the penalty parameter. The optimal Elastic Net penalty resulted in a ratio-TimeMachine model with 71 (out of 666) predictive features and a Z-score TimeMachine with all 37 (out of

37) predictive features. We then applied the trained predictor to both the held-out data from TrTe and the independent datasets without any cross-platform normalization or batch correction, yielding results shown in Fig. 1. To assess performance, we compute the median absolute error and the normalized area under the error CDF curve (AUC). Here, a predictor with an AUC of 0.5 performs no better than random chance, while a perfect predictor would have an AUC of 1.

For the unseen TrTe data, ratio TimeMachine had a median absolute error of 1:39 h, with 55.7% of predictions within ± 2 h and 83.8% within ± 4 h. Despite the study employing two sleep protocols (control versus sleep restriction), sleep restriction did not significantly affect ratio TimeMachine’s transcriptomic phase estimates ($P = 0.1$; Wilcoxon rank-sum test).

We also evaluated ratio TimeMachine’s performance on two independent datasets where experimental conditions and profiling platforms differed. V1 used the same microarray platform as the training data, but had different experimental conditions. V3 not only had different experimental conditions, but also profiled samples using RNA-Seq rather than microarray. Importantly, we did not perform any batch correction or additional data manipulation beyond the sample-wise normalization described in the previous section.

In V1, ratio TimeMachine accurately predicted the melatonin phase (i.e., time since DLMO) within ± 2 h and ± 4 h in 39.5% and 65.9% of samples, respectively, yielding a median absolute error of 2:41 h. Comparison of the error distribution between the two sleep protocols in the V1 dataset suggests that predictions became less accurate when sleep misaligned with melatonin ($P = 0.002$; Wilcoxon rank-sum test), corroborating findings from prior studies (10).

Using V3, we showed that ratio TimeMachine can be generalized to gene expression data across technical variability. Despite being trained on microarray data, when applied to RNA-Seq data, it yielded a median absolute error of 1:53 h. Specifically, it predicted the melatonin phase within ± 2 h for 50.3% of samples and within ± 4 h for 78.4%, further emphasizing TimeMachine’s consistent performance across platforms.

The Z-score variant of TimeMachine displayed comparable performance on the validation datasets (SI Appendix, Table S3). There was no significant difference between ratio TimeMachine and Z-score TimeMachine across all datasets ($P = \{0.71; 0.77; 0.3\}$ for dataset TrTe, V1, and V3, respectively; Wilcoxon rank-sum test). This consistency suggests both normalization methods perform equivalently for the purpose of inferring circadian phase from blood transcriptomics across platforms.

Comparison with Competing Methods. We compared TimeMachine to the state-of-the-art machine-learning algorithm that to date provides the most accurate predictions of the circadian phase from a single blood draw. Laing et al. proposed an approach based on partial least squares regression (PLSR), which selects a set of 100 predictor genes from thousands of genes to predict the melatonin phase (10). In their work, the authors trained the PLSR model using a subset of combined samples from both Möller et al. and Archer et al., processed by batch correction and quantile normalization, and tested it on the remaining samples in the mixed datasets. With these data, the PLSR-based method outperformed both ZeitZeiger and Molecular Timetable. However, as both datasets were processed by identical microarray platforms and mixed during the training, it remained unclear whether the trained predictor is robust enough to infer circadian time using data from other high-throughput technologies. We adopted the authors’ code to train the PLSR model using the same training samples as TimeMachine, using all genes for the

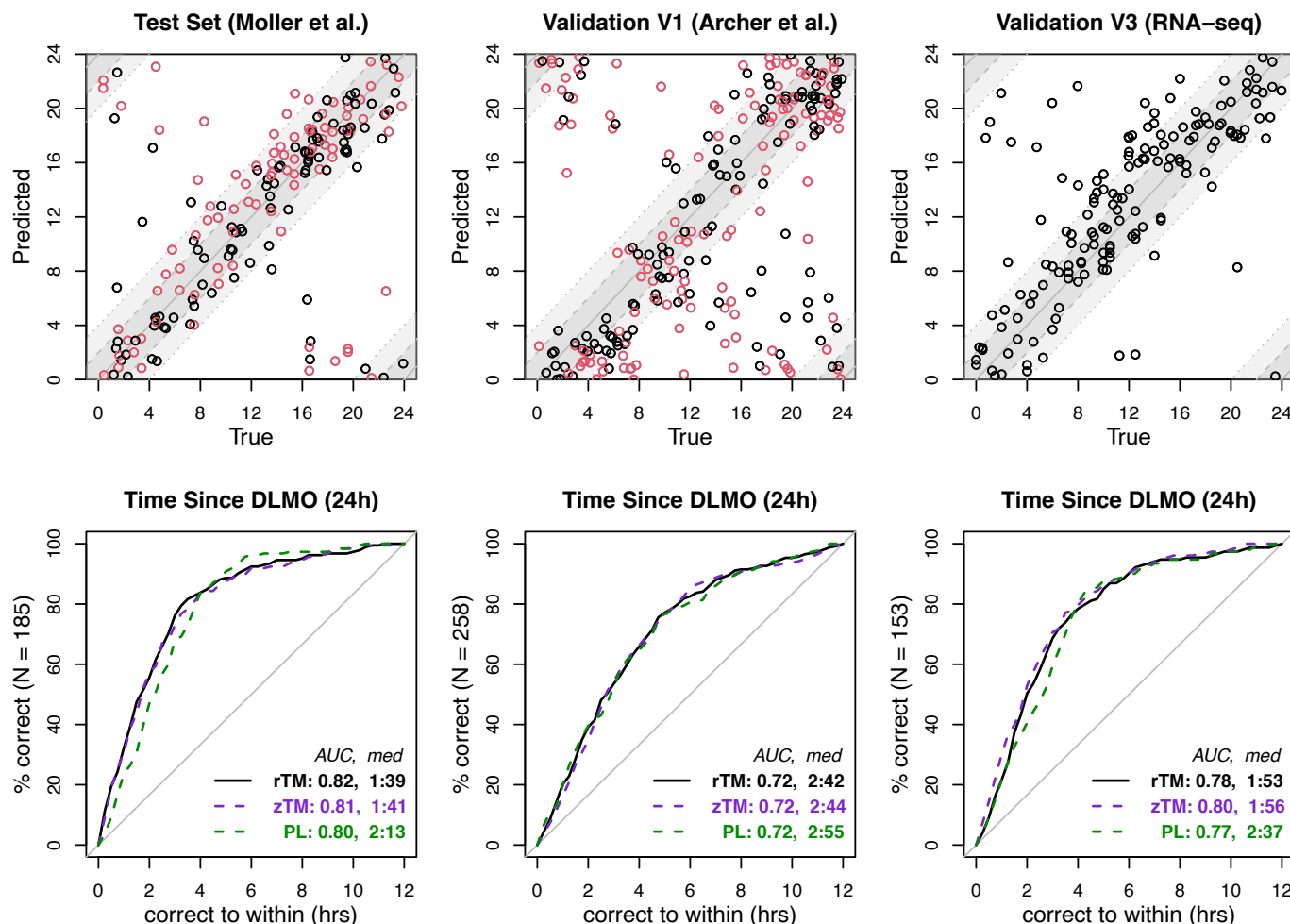


Fig. 1. TimeMachine predictions of melatonin phase (time since DLMO) on data from three distinct studies. Both variants of the TimeMachine algorithm, ratio TimeMachine (rTM) and Z-score TimeMachine (zTM), were trained on a subset of subjects from the Möller et al. study (TrTe) and then applied to the remaining test subjects in Möller et al. along with two independent datasets V1 (Archer et al.) and V3 (RNA-Seq) for validation. The top row shows the agreement of predictions from ratio TimeMachine with the measured melatonin phase (time since DLMO) for each sample. Dark and light gray bands indicate an error range of ± 2 and ± 4 h. The color of the point represents experimental protocols: Black denotes control condition, and red denotes sleep restriction (Möller et al.) and forced desynchrony (Archer et al.), respectively. In the bottom row, we plot the fraction of correctly predicted samples for each study vs. prediction errors for the ratio TimeMachine algorithm (solid black), in comparison to the other variant, Z-score TimeMachine algorithm (dashed purple), and the single-sample PLSR algorithm (dashed green), along with the normalized area under the curves (AUC) and median absolute errors for each algorithm.

quantile normalization as described in ref. 10. We then applied the trained PLSR model to the independent validation datasets.

As depicted in Fig. 1, TimeMachine's performance was as good or better than PLSR's, which exhibited a median absolute error of 2:13 h for the unseen TrTe testing data, 2:55 h for V1, and 2:37 h for V3 (Fig. 1). While the PLSR method comes close to the performance of TimeMachine, we note that both the quantile normalization step and the PLSR step itself require many more genes (all 7,615 genes and 100 genes, respectively) than are required for TimeMachine. In the testing subset of TrTe, ratio TimeMachine significantly outperforms the PLSR predictor ($P = 0.01$; Wilcoxon rank-sum test). However, its performance paralleled PLSRs in the V1 and V3 datasets ($P = \{1, 0.33\}$ for V1 and V3, respectively; Wilcoxon rank-sum test). Still, TimeMachine accurately predicted the melatonin phase within ± 2 h for $>50\%$ of the V3 samples, versus $\sim 40\%$ for PLSR (SI Appendix, Table S3). As noted above, similar accuracy was achieved by rTM and zTM. Comparison of Z-score TimeMachine and PLSR shows that the Z-score variant provided statistically better predictions than PLSR for datasets TrTe and V3, but not V1 ($P = \{0.035, 0.76, 0.047\}$ for TrTe, V1, and V3, respectively; Wilcoxon rank-sum test).

We further note that the performance of ratio and Z-score TimeMachine's performance is only slightly worse than that of reported state-of-the-art two-timepoint methods (11, 14), with a median absolute error that is ~ 20 min larger for the TrTe testing subset and ~ 40 min larger for the validation sets V1 and V3.

In order to further assess TimeMachine's generalizability, we apply it across all four distinct studies (including V2), covering two different microarray platforms and the RNA-Seq. Because melatonin data are unavailable for V2, the local time of the blood draw was used as a proxy. We justify this choice on the basis of the fact that the selection criteria for all four studies restricted participants to healthy, intermediate-chronotype individuals, for whom the circadian phase is expected to be well aligned to local time. Indeed, for the studies where both melatonin and local time data are available (TrTe, V1, and V3), it was previously demonstrated that melatonin phase is closely correlated with local time (14). Using local time gives us the opportunity to demonstrate TimeMachine's performance with a completely distinct sleep protocol and microarray platform from the other studies.

When trained on the same samples as in Fig. 1 to predict local time, ratio TimeMachine matched PLSR's accuracy for

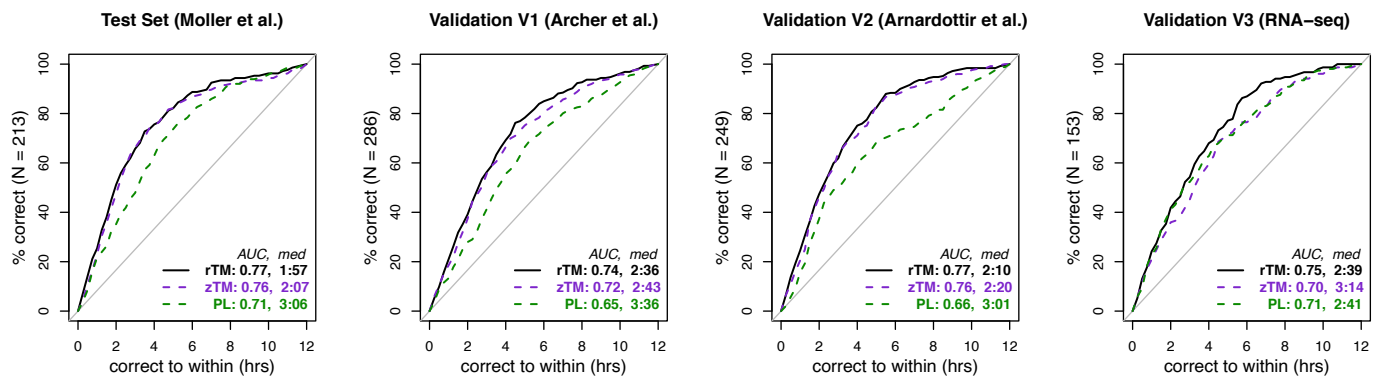


Fig. 2. Accuracy of one-timepoint methods on predicting the blood draw-time. Instead of predicting the melatonin phase as in Fig. 1, we trained and applied ratio TimeMachine (solid black), Z-score TimeMachine (dashed purple), and PLSR (dashed green) to obtain time-of-day predictions on four distinct datasets.

dataset V3 ($P = 0.41$; Wilcoxon rank-sum test). However, ratio TimeMachine outperformed PLSR across the three microarray datasets with statistical significance ($P < 0.001$ for TrTe, V1, and V2; Wilcoxon rank-sum test). Notably, ratio TimeMachine achieved a median absolute error that was ~ 1 h less than that of PLSR for data sourced from microarray platforms and predicted 10 to 15% more samples within the window of ± 2 h and ± 4 h (SI Appendix, Table S3). As shown in Fig. 2, ratio TimeMachine's predictions of local time had a median absolute error of ~ 2.5 h in all datasets, including V2 (a dataset from a different microarray platform than TrTe and V1). In contrast, PLSR's quantile normalized approach (labeled as PL in Fig. 2) exceeded a median absolute error of 3 h.

Furthermore, ratio TimeMachine and Z-score TimeMachine displayed similar error distribution when forecasting the local blood-draw times across all datasets ($P = \{0.51, 0.38, 0.59, 0.11\}$ for TrTe, V1, V2, and V3, respectively; Wilcoxon rank-sum test). Comparing Z-score TimeMachine with PLSR revealed that, akin to ratio TimeMachine, the Z-score variant generated statistically superior predictions for datasets TrTe, V1, and V2. However, the error distributions were indistinguishable for dataset V3 ($P = \{0.005, 0.002, <0.001, 0.48\}$ for TrTe, V1, V2 and V3, respectively; Wilcoxon rank-sum test). Overall, this indicates that either normalization technique—pairwise ratio or Z-score transformation—of the chosen 37 genes significantly enhanced prediction accuracy in predicting blood-draw timings.

Properties of TimeMachine. Beyond demonstrating ratio TimeMachine and Z-score TimeMachine's robustness across datasets, we aim to understand the factors that affect the algorithm's performance.

We expect that it may be harder to predict time in samples with a low amplitude of gene expression, simply due to the fact that the signal is low. In this case, we might expect that samples whose Cartesian prediction coordinates (Eq. 1) lie close to the origin will have larger errors in the time prediction (Eq. 4) than those that lie close to or beyond the unit circle. To investigate this, we calculate the "gene expression amplitude" as the magnitude of the mapping output, i.e., $\sqrt{\hat{y}_{i1}^2 + \hat{y}_{i2}^2}$. We find the resulting amplitude of predictions can easily deviate from the unit circle when the trained predictor is applied to the validation data. We observe that a smaller prediction amplitude yields a larger prediction error, as expected. Fig. 3 shows that for ratio TimeMachine, the absolute prediction error

is significantly larger when the associated output amplitude is below 0.5 ($P < 0.001$; Wilcoxon rank-sum test), suggesting that the output amplitude may serve as a proxy for quantifying the confidence of predictions. While ratio TimeMachine and Z-score TimeMachine take in different normalized data as the input, they both utilize a bivariate regression model. Thus, ratio TimeMachine and Z-score TimeMachine shared the same properties, where the absolute prediction error is significantly larger when the associated output amplitude is below 0.5 ($P < 0.001$; Wilcoxon rank-sum test).

Moreover, when we categorize samples by phase into 2-h intervals, a significant difference in amplitude over time becomes evident ($P < 0.01$, Kruskal–Wallis test). Visually examining the data suggests an inverse relationship between error distributions and predicted amplitude (SI Appendix, Figs. S1 and S2); that is, as amplitude decreases, error increases. Across all time bins, a predicted amplitude below 0.5 is associated with significantly higher error ($P < 0.001$). Furthermore, a lower predicted amplitude (below 0.5) is also associated with a significantly higher error even after adjusting for variation between individual subjects ($P < 0.001$). These findings are consistent for both ratio TimeMachine and Z-score TimeMachine, suggesting that the predicted amplitude serves as a useful measure of prediction confidence, irrespective of time or individual variation.

Discussion

We have introduced a machine-learning method called TimeMachine (with two variants, ratio TimeMachine and Z-score TimeMachine) that can predict physiological time using the expression of 37 genes from a single blood draw. In contrast to existing methods, our approach yields results that robustly generalize across platforms and study protocols without requiring batch correction or retraining. This feature is important for two reasons. First, it makes it possible to apply TimeMachine to future samples without constraining the transcriptomic profiling platform, and potentially devise small "kits" for its application. Second, it opens the possibility of applying TimeMachine to the wealth of existing, untimed data for secondary analysis, again without constraints on the platform.

The normalization step is crucial for achieving cross-platform accuracy, suggesting that normalization and machine learning algorithms may be optimized separately for best performance. Our algorithm presents two distinct within-sample normalization techniques: Z-score transformation (Z-score TimeMachine) and pairwise ratio of gene expression (ratio TimeMachine). No

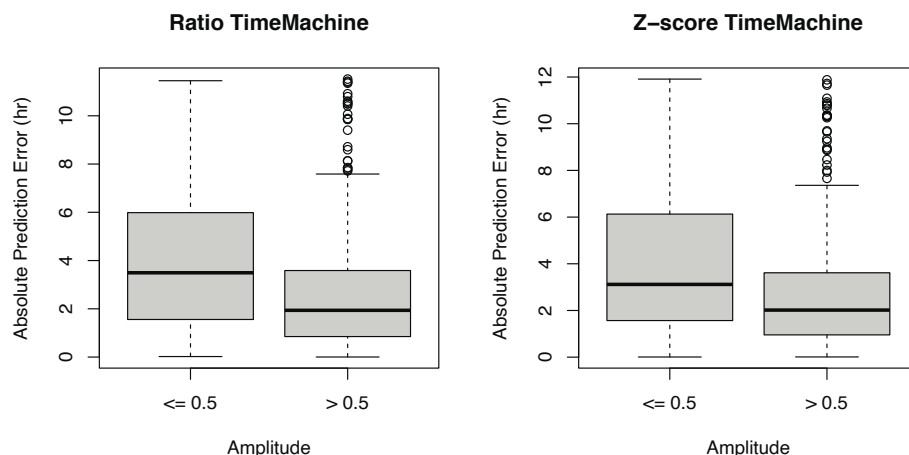


Fig. 3. Relationship between prediction accuracy and its predicted amplitude for ratio TimeMachine and Z-score TimeMachine. The amplitude here is defined as the magnitude of the predictor $\hat{Y}$. Predicted amplitudes ≤ 0.5 yield a significantly less accurate time prediction than higher amplitudes for both variants of TimeMachine ($P < 0.001$; Wilcoxon rank-sum test).

significant performance difference was found between ratio TimeMachine and Z-score TimeMachine on different datasets. Both the pairwise ratios and Z-score normalization operate on the assumption that relative gene abundances—rather than their absolute magnitudes—are both biologically relevant and faithfully captured across different platforms. We believe this is a reasonable assumption for both RNA-seq and PCR, and previous studies using pairwise gene ratios suggest that microarrays also preserve relative expression differences with enough fidelity to serve as robust predictors (21–23). Our empirical results with the 37 predictor genes corroborate this assumption. However, future studies should explore the strength of this assumption, particularly for older microarrays not considered in the current study.

TimeMachine performs at least as well as the state-of-the-art PLSR method and in some cases performs significantly better. However, although the PLSR-based approach yielded performance close to TimeMachine in several of our tests, we note that as described in ref. 10 it requires quantile normalization across the complete dataset, implying the need to assay all 7,615 genes in subsequent applications (not only the 100 PLSR predictor genes). In contrast, TimeMachine relies on just 37 genes, making it a more practical choice for future usage. Our findings suggest that assaying and normalizing the full complement of genes may not be necessary for predicting circadian time. In fact, our results indicate that the selection and preprocessing steps matter as much as the ML method (TimeMachine’s ElasticNet versus PLSR; *SI Appendix, Fig. S3*). This is unsurprising, considering that both ElasticNet and PLSR are regularized regression techniques, and should thus achieve comparable performance. It also underscores the benefits of thoughtful feature selection prior to any machine learning algorithm.

We also note the appearance of a recently proposed algorithm (tauFisher) (24) that employs a regression model on the principal components of the pairwise gene ratios of core clock genes (BMAL1, BDP, NR1D1, NR1D2, PER1, PER2, PER3, CRY1, and CRY2) along with the top 10 rhythmic genes, following smoothing and rescaling of their expression profiles (24). This approach has been shown to predict circadian time in single-cell scRNA-seq mouse SCN and skin samples, based on training with bulk microarray and RNA-seq mouse SCN and skin data. These findings independently corroborate our observation that the pairwise gene ratio likely maintains the circadian dynamics between genes, providing predictive features for circadian time, even across different assay platforms. However, at the time of

this writing, the extent to which tauFisher can be generalized to human data and across platforms or sleep conditions, remains unclear.

In addition to introducing a method for predicting circadian phase, we also investigated the factors that influence the accuracy of our predictions. We notice that low predicted amplitudes from the TimeMachine model (including both variants) are associated with a less accurate prediction of physiological time. This relationship suggests that the amplitude may be used as a rough measure of confidence in the prediction. Yet while there is some practical utility in the amplitude prediction, we must emphasize that it has no physiological interpretation, since the training data were assumed by fiat to have amplitude 1. Because only the circadian phase varied in the training data (and was the underlying prediction target), it may be of interest for future studies to consider instead a one-dimensional model using circular regression to predict time as a circular variable, rather than transforming time to a Cartesian plane as in Eqs. 1–3. In this case, the optimization would select β coefficients that minimize angular error alone, rather than the overall distance in the 2-d plane (as in Eq. 2). However, it will still be necessary to include a regularization penalty to inhibit overfitting. Solving the optimization problem for circular penalized regression remains an interesting avenue for future work.

From a practical perspective, TimeMachine’s single-sample requirement enhances its usability and feasibility, but greater accuracy can be achieved with two-sample methods such as TimeSignature (~20 to ~40 min improvement in the median absolute error). Using just one sample, TimeMachine predicted $>50\%$ of TrTe and V3 samples, and $\sim 40\%$ of V1 samples within a ± 2 h range. In contrast, TimeSignature, using two samples taken 12 h apart, predicted $\sim 70\%$ of TrTe and V3 samples and $\sim 45\%$ of V1 samples within the same time frame. Therefore, where possible, we still recommend using two blood draws, at least 8 and ideally 12 h apart (11), to assay circadian phase.

For applications requiring yet greater accuracy, we propose that TimeMachine’s performance could be used to identify the optimal sample collection window for melatonin studies, reducing the need for burdensome sampling. Until now, determining the gold-standard phase marker, DLMO, has been both costly and time-consuming, requiring hourly saliva or plasma samples for at least 6 to 8 h around the expected melatonin onset time. In cases where prior knowledge of melatonin onset is unknown, such as in people with sleep disorders, samples must be taken

throughout an entire circadian cycle. TimeMachine inferences could be used to optimize this collection to a smaller window.

It is important to highlight that the selection of the 37 predictor genes is not necessarily unique. As previous studies have discussed, many genes cycle in phase, and could be equally good predictors; a different training set might yield an alternate set of predictor genes (9–11). Moreover, the circadian transcriptomics datasets used in this study (11, 15–17) only include young, healthy, intermediate-chronotype individuals. Previous studies in animal models have shown that factors such as sex, sleep disorders, shift work, and aging can lead to significant changes in gene expression rhythms (25–30), and it is likely that similar changes occur in humans as well. Such populations may benefit the most from circadian-based treatments, but temporal gene expression data from diverse human populations (such as night shift workers and various disease groups) largely remains lacking. It remains critical to collect more data to explore the performance of all time inference methods (including TimeMachine) in diverse populations. In previous work, for instance, we demonstrated that the two-draw TimeSignature method is robust to changes in healthy aging (31), but sensitive to changes during critical illness (32). We expect that TimeMachine may have similar properties that could be explored in future studies.

Related to this, we find that being out of sync with melatonin (rather than mere sleep deprivation) can lead to a greater deviation between TimeMachine's predicted time and the actual melatonin phase. This trend parallels findings from predictions involving two-timepoint approaches (10). Research has demonstrated that being out of phase with melatonin leads to a reduction of rhythmic transcripts from 6.4% to a mere 1% (16), causing desynchrony between the peripheral circadian rhythm in blood transcriptome and the central circadian rhythm of melatonin.

By training the algorithm to predict melatonin phase, we use the transcriptional dynamics in blood as a proxy for rhythms in the brain. However, emerging evidence shows that peripheral clocks in other tissues may be affected by factors distinct from those influencing the central clock in the brain—for example, food and exercise for peripheral clocks as opposed to light for the central clock (33, 34). These peripheral clocks may also exhibit phase differences relative to those in the brain or blood. Therefore, additional studies are needed to draw inferences about phases in different organs, and to understand how the interplay between rhythms in the brain, blood, and other tissues impacts health outcomes.

Our study is an important step toward the use of circadian biomarkers in wide-ranging applications. While the importance of the circadian rhythm in human health has been highlighted in multiple studies, the inclusion of circadian biomarkers in medical research and clinical use has been hampered by the challenge of assaying physiological time. TimeMachine addresses this gap by providing predictions that show good agreement with in-lab DLMO data and blood draw time, along with an indicator of reliability (amplitude). TimeMachine requires only a single-timepoint gene expression profile of human peripheral blood mononuclear cells (PBMCs), making it a practical tool for real-world implementation. TimeMachine's generalizability is a significant advantage. It enables prospective application, wherein TimeMachine is applied to new samples (e.g., for the purpose of customizing the timing of light intervention based on an individual's circadian phase), without the need for data to be collected on a specific platform. Additionally, TimeMachine enables retrospective circadian analysis of existing data by providing a platform-free way to infer time label for existing untimed samples. By inferring these labels, secondary analyses of

vast quantities of existing PBMC expression data could be mined for associations with circadian time, shedding light on the role of circadian rhythms in diseases from neurodegeneration to cancer.

Materials and Methods

Data Sources. We analyzed published data of human whole-blood transcriptome profiles from four independent studies. Data were obtained from the NCBI Gene Expression Omnibus (GEO) repository under the following accession numbers and imported into R:

GSE39445 (training and testing [TrTe]) consists of 427 blood samples from 26 subjects in total (15). Each subject underwent two protocols in a crossover design: a sleep-restriction protocol (6 h of sleep opportunity per night) and a control protocol (10 h of sleep opportunity per night). Following the sleep opportunity, blood samples were collected every 3 h over a 27-h period under constant routine conditions. Samples were assayed using a Whole Human Genome 4×44K custom Agilent microarray platform (G2514F, AMADID 026817) as described in ref. 15. We split the data subject-wise at random into training and testing subsets for the purpose of the present study.

GSE48113 (validation dataset V1) consists of 287 blood samples from 22 subjects under a forced-desynchrony protocol, where the sleep-wake cycle of all subjects was scheduled for a 28-h day (16). Each subject went through two conditions, sleeping in phase and out of phase with their melatonin rhythm. Blood samples were collected every 4 h for both conditions. Samples from V1 were assayed using the same microarray platform as the training/testing data (Agilent; G2514F, AMADID 026817) as described in ref. 16.

GSE56931 (validation dataset V2) consists of 249 samples from 14 subjects. For each subject, blood samples were collected every 4 h during a 3-d study. Samples were processed using a different microarray platform (Rosetta/Merk Human RSTA Custom Affymetrix 2.0 microarray) from the previous two, as described in ref. 17.

GSE113883 (validation dataset V3) comprises 11 subjects whose whole blood samples were collected every 2 h over a 28-h constant routine, yielding a total of 165 samples (11). In contrast to the previous three datasets, transcriptional profiling of samples from V3 was processed by RNA sequencing.

The processing of microarray and RNA-Seq data, as well as transcript alignment, has been previously carried out by the original authors (11, 15–17), and these procedures remained unaltered for our present study. In the case of microarray data, gene expression was averaged when multiple probes represented the same gene. More details of the individual datasets can be found in refs. 11 and 15–17.

Availability of Melatonin Data. In conjunction with blood transcriptome, melatonin levels were also available for all datasets except V2. Dim-light melatonin onset (DLMO) is used in these studies the gold-standard reference marker for the human circadian phase. As in refs. 11 and 14, we also consider local time in an analysis that includes V2 (which lacks public melatonin data).

Data, Materials, and Software Availability. Code data have been deposited in GitHub (<https://github.com/pepperhuang/TimeMachine>) (35). All other data are included in the manuscript and/or *SI Appendix*. Previously published data were used for this work (GSE39445 (36), GSE48113 (37), GSE56931 (38), and GSE113883 (39)) were obtained from the NCBI Gene Expression Omnibus repository).

ACKNOWLEDGMENTS. Y.H. and R.B. were supported by NIH/NIA R01AG 068579, Simons Foundation 597491-RWC01, and NSF 1764421-01.

Author affiliations: ^aDepartment of Molecular Biosciences, Northwestern University, Evanston, IL 60208; ^bNational Institute for Theory and Mathematics in Biology, Northwestern University, Evanston, IL 60208; ^cDepartment of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208; ^dDepartment of Physics and Astronomy, Northwestern University, Evanston, IL 60208; ^eNorthwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208; and ^fSanta Fe Institute, Santa Fe, NM 87501

1. J. S. Takahashi, Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* **18**, 164–179 (2017).
2. R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, J. B. Hogenesch, A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16219–16224 (2014).
3. J. M. Lane *et al.*, Genetics of circadian rhythms and sleep in human health and disease. *Nat. Rev. Genet.* **24**, 1–17 (2022).
4. S. M. Abbott, R. G. Malkani, P. C. Zee, Circadian disruption and human health: A bidirectional relationship. *Euro. J. Neurosci.* **51**, 567–583 (2020).
5. M. D. Ruben, D. F. Smith, G. A. FitzGerald, J. B. Hogenesch, Dosing time matters. *Science* **365**, 547–549 (2019).
6. E. Filipinski *et al.*, Host circadian clock as a control point in tumor progression. *J. Natl. Cancer Inst.* **94**, 690–697 (2002).
7. F. Levi, U. Schibler, Circadian rhythms: Mechanisms and therapeutic implications. *Annu. Rev. Pharmacol. Toxicol.* **47**, 593–628 (2007).
8. H. R. Ueda *et al.*, Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11227–11232 (2004).
9. J. J. Hughey, Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. *Genome Med.* **9**, 1–11 (2017).
10. E. E. Laing *et al.*, Blood transcriptome based biomarkers for human circadian phase. *Elife* **6**, e20214 (2017).
11. R. Braun *et al.*, Universal method for robust detection of circadian state from gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9247–E9256 (2018).
12. J. J. Hughey, T. Hastie, A. J. Butte, Zeitzeiger: Supervised learning for high-dimensional data from an oscillatory system. *Nucleic Acids Res.* **44**, e80–e80 (2016).
13. N. Wittenbrink *et al.*, High-accuracy determination of internal circadian time from a single blood sample. *J. Clin. Invest.* **128**, 3826–3839 (2018).
14. R. Braun *et al.*, Reply to Laing *et al.*: Accurate prediction of circadian time across platforms. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5206–5208 (2019).
15. C. S. Möller-Levet *et al.*, Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1132–E1141 (2013).
16. S. N. Archer *et al.*, Mistimed sleep disrupts circadian regulation of the human transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E682–E691 (2014).
17. E. S. Arnardottir *et al.*, Blood-gene expression reveals reduced circadian rhythmicity in individuals resistant to sleep deprivation. *Sleep* **37**, 1589–1600 (2014).
18. M. E. Hughes, J. B. Hogenesch, K. Kornacker, Jtk_cycle: An efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. Rhyth.* **25**, 372–380 (2010).
19. G. Wu, R. C. Anafi, M. E. Hughes, K. Kornacker, J. B. Hogenesch, Metacycle: An integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016).
20. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **67**, 301–320 (2005).
21. N. D. Price *et al.*, Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3414–3419 (2007).
22. D. Geman, C. d'Avignon, D. Q. Naiman, R. L. Winslow, Classifying gene expression profiles from pairwise mRNA comparisons. *Stat. Appl. Genet. Mol. Biol.* **3**, 1–22 (2004).
23. L. Xu, A. C. Tan, R. L. Winslow, D. Geman, Merging microarray data from separate breast cancer studies provides a robust prognostic test. *BMC Bioinf.* **9**, 1–14 (2008).
24. J. Duan *et al.*, tauFisher accurately predicts circadian time from a single sample of bulk and single-cell transcriptomic data. *bioRxiv [Preprint]* (2023). <https://www.biorxiv.org/content/10.1101/2023.04.04.535473v1.full>. Accessed 1 September 2023.
25. F. O. James, N. Cermakian, D. B. Boivin, Circadian rhythms of melatonin, cortisol, and clock gene expression during simulated night shift work. *Sleep* **30**, 1427–1436 (2007).
26. D. B. Boivin, P. Boudreau, Impacts of shift work on sleep and circadian rhythms. *Pathol. Biol.* **62**, 292–301 (2014).
27. S. Hood *et al.*, The aging clock: Circadian rhythms and later life. *J. Clin. Invest.* **127**, 437–446 (2017).
28. B. D. Weger *et al.*, The mouse microbiome is required for sex-specific diurnal rhythms of gene expression and metabolism. *Cell Metabol.* **29**, 362–382 (2019).
29. L. Talamanca, C. Gobet, F. Naef, Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans. *Science* **379**, 478–483 (2023).
30. C. A. Wolff *et al.*, Defining the age-dependent and tissue-specific circadian transcriptome in male mice. *Cell Rep.* **42**, 111982 (2023).
31. S. K. Smith *et al.*, Validation of blood-based transcriptomic circadian phenotyping in older adults. *Sleep* **45**, zsc148 (2022).
32. M. B. Maas *et al.*, Circadian gene expression rhythms during critical illness. *Crit. Care Med.* **48**, e1294 (2020).
33. N. Cermakian, D. B. Boivin, The regulation of central and peripheral circadian clocks in humans. *Obesity Rev.* **10**, 25–36 (2009).
34. J. A. Mohawk, C. B. Green, J. S. Takahashi, Central and peripheral circadian clocks in mammals. *Annu. Rev. Neurosci.* **35**, 445–462 (2012).
35. Y. Huang, R. Braun, TimeMachine Source Code. Github. <https://github.com/pepperhuang/TimeMachine>. Deposited 13 September 2023.
36. C. Möller-Levet *et al.*, Effect of sleep restriction on the human transcriptome during extended wakefulness. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39445>. Deposited 20 February 2019.
37. C. Möller-Levet *et al.*, Mistimed sleep disrupts circadian regulation of the human blood transcriptome. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48113>. Deposited 19 June 2013.
38. K. R. Shockley, A. I. Pack, A. Podtelezhinikov, Blood gene expression reveals reduced circadian rhythmicity in individuals resistant to sleep deprivation. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56931>. Deposited 4 March 2015.
39. R. Allada *et al.*, TimeSignature: A Universal Method for Robust Detection of Circadian State from Gene Expression. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113883>. Deposited 26 March 2019.