Submitted to the Annals of Applied Statistics

BAYESIAN RHYTHMIC MODEL FOR JOINTLY DETECTING CIRCADIAN BIOMARKERS AND PREDICTING MOLECULAR CIRCADIAN TIME IN HUMAN POST-MORTEM BRAIN TRANSCRIPTOME

BY XIANGNING XUE[a], HAO WANG[a], ZIHAN ZHU[b], YU GUAN[a], MARIANNE L. SENEY[c], LIANGLIANG ZHANG[b], GEORGE C. TSENG[*,a]

Department of Biostatistics and Health Data Science
University of Pittsburgh, Pittsburgh, PA 15261 [a]*ctseng@pitt.edu*

Department of Population and Quantitative Health Sciences
Case Western Reserve University, Cleveland, OH 44106 [b]

Department of Psychiatry
University of Pittsburgh, Pittsburgh, PA 15213 [c]

Transcriptomic circadian analysis of human post-mortem brain provides a unique opportunity to characterize in vivo molecular circadian rhythms across brain regions implicated in aging and psychiatric disorders. A primary goal in such analyses is the detection of circadian biomarkers. However, this task is complicated by the frequent mismatch between a subject's recorded circadian clock time and their true molecular circadian time — arising from observational or recording errors, as well as intrinsic biological variability. Existing methods typically address either biomarker detection or circadian time prediction in isolation. Because errors in one task can degrade performance in the other, the lack of a unified approach remains a key limitation. We propose BayCT — a Bayesian model for simultaneous circadian marker detection and molecular circadian time estimation. The model extends naturally to repeated measurements from multiple brain regions or organs. For circular data, we employ a von Mises prior distribution, with slice sampling and reversible-jump Markov chain Monte Carlo (MCMC) for Bayesian inference. Through extensive simulations and applications to transcriptomic data from three human brain regions and from 12 mouse organs, BayCT demonstrates superior performance in both biomarker detection and circadian time estimation. Furthermore, we highlight the advantages of integrating data across brain regions, achieving substantial improvements in both tasks.

1. Introduction. Circadian rhythms are endogenous ∼24-hour oscillations that synchronize physiological activities to the daily light–dark cycle. This intrinsic timekeeping system optimizes the timing of diverse biological processes, thereby enhancing organismal survival. At the molecular level, circadian rhythms are driven by a transcriptional–translational feedback loop that regulates thousands of circadian-controlled genes in a coordinated manner (Pittendrigh, 1960). In mammals, the suprachiasmatic nucleus (SCN) serves as the master clock, maintaining self-sustained oscillations independent of environmental cues (Aschoff, 1960). Although the free-running period is close to — but not exactly — 24 hours, it varies with age, sex, and ancestry (Crowley and Eastman, 2018; Eastman et al., 2015), necessitating entrainment to environmental signals such as light to maintain precise phase alignment. However, modern lifestyle factors — including artificial lighting, altered work schedules, and late-night blue light exposure — challenge this entrainment process (Finger and Kramer, 2021; Tähkämö, Partonen and Pesonen, 2019). Chronic circadian misalignment, such as that

[*]George C. Tseng is corresponding author
Keywords and phrases: Molecular circadian time, Circadian biomarker, Bayesian hierarchical models.

induced by rotating night shifts, has been linked to elevated risks for cancer and other chronic diseases (Schernhammer et al., 2001; Wegrzyn et al., 2017).

Detection of circadian biomarkers from high-throughput omics data is fundamental for studying rhythmic regulation in health and disease. A variety of computational methods have been proposed, including parametric cosinor regression (Cornelissen, 2014), non-parametric approaches such as JTK_CYCLE (Hughes, Hogenesch and Kornacker, 2010), spectral methods such as the Lomb–Scargle periodogram (Glynn, Chen and Mushegian, 2006), and deep learning methods such as BIO_CYCLE (Agostinelli et al., 2016). Flexible models like JTK_CYCLE and BIO_CYCLE can capture diverse waveform shapes beyond the sinusoidal assumption of cosinor regression. However, they may suffer from inflated or deflated type I error rates and present challenges in accurately estimating rhythmic parameters such as phase and amplitude (Ding et al., 2021).

Our motivating example is from a human post-mortem brain transcriptomic study for circadian research, in which bulk RNA-seq data are profiled in three brain regions - nucleus accumbens (NAc), caudate and putamen - of the same $N = 59$ subjects with time of death (TOD) information for each subject (Section 4.2). In post-mortem human studies, a subject's recorded time of death (TOD) often deviates from their true molecular circadian time (MCT), due to observational or recording errors and intrinsic inter-individual variation in circadian phase. Such variability arises from environmental influences (e.g., light exposure, temperature), behavioral patterns (e.g., sleep–wake schedules, meal timing), genetic differences (Jones et al., 2019), and lifestyle-induced chronodisruption. Accurate estimation of MCT has substantial implications for chronotherapy, which aims to align treatment timing with an individual's circadian phase to maximize efficacy and minimize toxicity (Dallmann, Okyar and Lévi, 2016; Cederroth et al., 2019). Meta-analyses and clinical trials have shown that dosing time can significantly influence drug toxicity and therapeutic response (Printezi et al., 2022; Ruben et al., 2019).

Several computational approaches have been developed to predict MCT from transcriptomic data. Early work such as the Molecular Timetable (Ueda et al., 2004) correlated samples to reference gene-phase profiles, while more recent methods employ machine learning (e.g., ZeitZeiger (Butte and Hastie, 2016), TimeSignature (Braun et al., 2018)) or Bayesian inference (e.g., Tempo (Auerbach et al., 2022) and CHIRAL (Talamanca, Gobet and Naef, 2023)). Deep learning models such as CYCLOPS (Anafi et al., 2017) and Cyclum (Liang et al., 2020) leverage nonlinear representation learning to improve prediction accuracy. However, these methods address MCT prediction and circadian biomarker detection as separate tasks, despite the fact that errors in one can propagate and degrade performance in the other. Figure 1A visually demonstrates the need for joint modeling of both tasks by scatter plots for time of death (TOD; x-axis) versus gene expression level (y-axis) in four selected core clock genes (CIART, CRY1, PER2 and PER3) in 59 human caudate samples. Subject 103 shows clear deviance between recorded TOD (solid circle) and the predicted MCT (empty circle). Correcting recorded TOD=-5.61 to predicted MCT=0.43 yields improved sinusoidal fits for circadian biomarker detection, and these gains in biomarker detection subsequently enhance the accuracy of MCT prediction.

To address this gap, we propose BayCT — a unified Bayesian model for simultaneous Circadian biomarker detection and molecular circadian Time estimation in Section 2.1. BayCT employs a von Mises prior to accommodate the circular nature of time and uses slice sampling and reversible-jump Markov chain Monte Carlo (MCMC) for efficient Bayesian inference (Section 2.2). In Section 2.3, we further extend the model to BayCT-M to integrate data from multiple brain regions or organs. In Section 3 and Section 4, we evaluate BayCT and BayCT-M through extensive simulations and real-data applications involving transcriptomic measurements from 12 mouse organs and three human brain regions. Our results demonstrate substantial improvements in both biomarker detection and MCT estimation, and highlight the benefits of integrating data across multiple brain regions or organs.

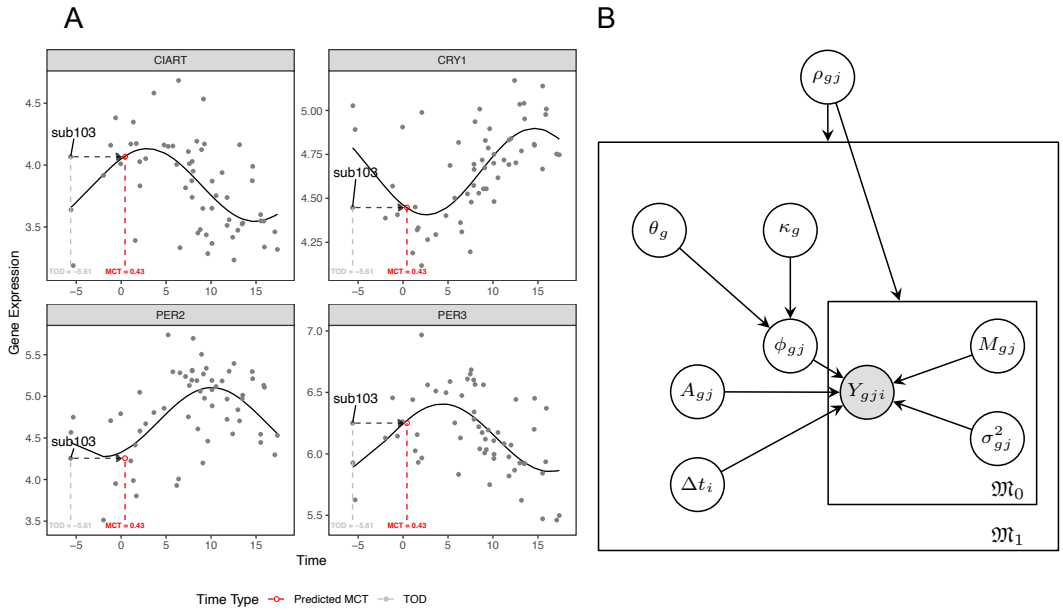

FIG 1. *A. Circadian gene expression patterns of four selected core clock genes in the human caudate. x axis is time on a ZT scale (-6 to 18 hour) and y axis is gene expression level. The black line is the fitted sinusoidal curve using TOD. Correcting TOD=-5.61 to predicted MCT=0.43 yields improved sinusoidal fits for circadian biomarker detection, and the improvement in biomarker detection will subsequently enhance MCT prediction accuracy. B. Graphical representation of BayCT-M. The graphical representation of BayCT is a simplified version without subscript for brain region j and parameters θ_g and κ_g.*

2. Methods: BayCT and BayCT-M. In this section, we present the BayCT and BayCT-M model. In Section 2.1, we begin with introducing a basic Bayesian circadian model (BayC) without aligning molecular circadian time, followed by a full BayCT model including MCT estimation. The BayC model will also be used as a baseline method in numerical studies when MCT correction is ignored in biomarker detection. Section 2.2 shows the corresponding Monte-Carlo Markov Chain (MCMC) sampling procedures for Bayesian inference. Finally, we extend the BayCT model to BayCT-M (Section 2.3) allowing multi-brain-region integrative analysis.

2.1. *BayCT model.* Consider Y_{ig} the expression values of G genes ($1 \leq g \leq G$) and N subjects ($1 \leq i \leq N$) with the time of death (TOD) of subject i standardized to zeitgeber time (ZT) as t_i. Here, zeitgeber time is adjusted by factors such as time zone, daylight savings time, and sunrise time (calculated by date of death and longitude and latitude of the place of death) and and ZT=0 is standardized to 6:00AM as previously described (Chen et al., 2016). For simplicity, we begin with a basic Bayesian model for circadian marker detection without TOD correction by MCT, which is denoted as BayC model. The cosinor model of BayC for a given gene g is:

$$Y_{ig} = M_g + A_g \cos(\omega(t_i - \phi_g)) + \epsilon_{ig}, \tag{1}$$

where A_g is the oscillation amplitude, ϕ_g is the peak time of gene expression, M_g represents the MESOR (midline-estimating statistic of rhythm; mean expression level), ϵ_{ig} is Gaussian noise with $var(\epsilon_{ig}) = \sigma_g^2$, and $\omega = \frac{2\pi}{P}$ is the oscillation frequency ($P = 24$ hours is pre-defined to reflect circadian rhythm). The range of the amplitude and the peak time are constrained with $\phi \in [0, 24]$ and $A \in (0, +\infty)$.

The hyperparameters and priors are set as:

$$
\begin{aligned}
&M_g \sim N(\mu_M, \sigma_M^2),\\
&\sigma_g^2 \sim IG(V_0/2, V_0\sigma_0^2/2),\\
&A_g \sim (1-\rho_g)\delta_0 + \rho_g N_{I(A_{min},+\infty)}(\mu_A, \sigma_A^2),\\
&\phi_g \sim VM(\omega\theta_0, A_g\kappa_0) * (1/\omega).
\end{aligned} \tag{2}
$$

Here, N is normal distribution, IG is inverse gamma, A_g follows a spike-and-slab prior with spike at 0 and slab as truncated normal distribution $N_{I(a,b)}$ constrained within (a,b), VM is von Mises distribution for circular data. The indicator variable $\rho_g = 1$ indicates gene g is rhythmic and $\rho_g = 0$ arryhythmic. We set $\mu_M = 0$, $\mu_A = 1$, $\sigma_M^2 = \sigma_A^2 = 10^2$, $\sigma_0^2 = 0$ and $V_0 = 2$ to give minimally-informative prior for M and σ. For prior of A, A_{min} can be user-defined, representing the smallest biologically meaningful amplitude but we use $A_{min} = 0$ throughout the paper for simplicity. The von Mises prior for ϕ is used for its periodic nature. We set $\theta_0 = 0$ and $\kappa_0 = 0$ to give non-informative prior to ϕ. In fact, $VM(\theta_0, 0)$ is equivalent to Uniform$(0, P)$.

We note that the value of the circadian indicator ρ generates two possible model states for the gene expression below with different parameter space: $\Theta_0 = \left\{M', \sigma^{2\prime}\right\}$ for $\rho = 0$ and $\Theta_1 = \left\{M, A, \phi, \sigma^2\right\}$ for $\rho = 1$, which calls for reversible jump sampling to be introduced in Section 2.2.

$$
\begin{aligned}
&\mathfrak{M}_0 : Y_i \mid M', \sigma^{2\prime}, \rho = 0 \sim N\left(M', \sigma^{2\prime}\right)\\
&\mathfrak{M}_1 : Y_i \mid M, \sigma^2, A, \phi, \rho = 1 \sim N\left(M + A\cos\left(\omega\left(t_i - \phi\right)\right), \sigma^2\right)
\end{aligned} \tag{3}
$$

To achieve joint biomarker detection and time correction in BayCT model, we introduce t_i^+ to denote the underlying true molecular circadian time (MCT). The circadian time deviation between MCT and TOD is denoted as Δt_i such that $t_i^+ = t_i + \Delta t_i$. The cosinor model in BayCT simply replaces t_i in the BayC model (Equation 1) with t_i^+:

$$
Y_i = A\cos(\omega(t_i^+ - \phi)) + M + \epsilon_i, \text{where } t_i^+ = t_i + \Delta t_i. \tag{4}
$$

Since Δt_i is also periodic, we give a reasonably informative prior

$$
\Delta t_i \sim VM(0, \kappa_t) * (1/\omega).
$$

The mean of the prior is set to 0, and κ_t is set to 6, which is equivalent to a standard deviation of about 1.5 hours. Figure 1B shows the graphical representation of BayCT under $\mathfrak{M}_0$ and $\mathfrak{M}_1$.

2.2. *Sampling procedure.* The sampling procedure for the BayCT model consists of two alternating stages: updating within models and transitioning between the two models in Equation 3 using reversible jump MCMC. The complete procedure is divided into four steps. Step I occurs only when updating within model $\mathfrak{M}_0$, while Steps I, II, and IV are applied to model $\mathfrak{M}_1$. Step III is performed after each update within either model.

Step I updates the two parameters M and σ^2 shared by $\mathfrak{M}_0$ and $\mathfrak{M}_1$. In Step II, the application of slice sampling for jointly updating A and ϕ in $\mathfrak{M}_1$ significantly improves the computing efficiency. Step III applies reversible jump sampling to transit between $\mathfrak{M}_0$ and $\mathfrak{M}_1$. Finally, Step IV updates the time deviance Δt. We use the notation $P^*(.)$ to represent the full conditional posterior probability for a given parameter, and define $z_i = A\cos(\omega(t_i^+ - \phi))$ (with $z_i = 0$ for $\mathfrak{M}_0$).

Step I: Gibbs sampling for M and σ^2

The full conditional posterior distributions of M and σ^2 can be derived and updated using Gibbs sampling:

$$
\begin{aligned}
&P^*(M) \sim N\left(\frac{\frac{\sum_{i=1}^n (y_i - z_i)}{\sigma^2} + \frac{\mu_M}{\sigma_M^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_M^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_M^2}}\right), \\
&P^*(\sigma^2) \sim \text{Inverse Gamma}\left(\frac{V_0 + n}{2}, \frac{V_0\sigma_0^2 + \sum_{i=1}^n (y_i - z_i - M)^2}{2}\right).
\end{aligned} \tag{5}
$$

Step II: Update within $\mathfrak{M}_1$: slice sampling for joint update of A and ϕ

The joint posterior distribution of A and ϕ is

$$
\begin{aligned}
P^*(A, \phi) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left[Y_i - M - A\cos(\omega(t_i^+ - \phi))\right]^2 + \kappa_0 \cos(\omega\phi - \omega\theta_0)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left[Y_i - M - A\cos(\omega(t_i^+ - \phi))\right]^2\right\} \quad (\kappa_0 = 0) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left[Y_i - M - \cos\omega t_i^+ A\cos\omega\phi - \sin\omega t_i^+ A\sin\omega\phi\right]^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{Y}' - \boldsymbol{X}\boldsymbol{\beta}\right]^T\left[\boldsymbol{Y}' - \boldsymbol{X}\boldsymbol{\beta}\right]\right\},
\end{aligned} \tag{6}
$$

where

$$
\boldsymbol{X} = \begin{pmatrix} \cos(\omega t_1^+) & \sin(\omega t_1^+) \\ \cos(\omega t_2^+) & \sin(\omega t_2^+) \\ \vdots & \vdots \\ \cos(\omega t_N^+) & \sin(\omega t_N^+) \end{pmatrix},
$$

$$
\boldsymbol{\beta} = \begin{pmatrix} A\cos(\omega\phi) \\ A\sin(\omega\phi) \end{pmatrix},
$$

and

$$
\boldsymbol{Y}' = \begin{pmatrix} Y_1 - M \\ Y_2 - M \\ \vdots \\ Y_n - M \end{pmatrix}.
$$

If we re-parameterize the above density as a posterior density for $\boldsymbol{\beta}$ with

$$
\begin{aligned}
\beta_1 &= A\cos(\omega\phi), \\
\beta_2 &= A\sin(\omega\phi),
\end{aligned} \tag{7}
$$

then the inverse function can be obtained as

$$
\begin{aligned}
A &= \sqrt{\beta_1^2 + \beta_2^2} \\
\phi &= \frac{1}{\omega}\arctan 2(\beta_1, \beta_2).
\end{aligned} \tag{8}
$$

and the corresponding Jacobian determinant can be calculated as

$$
\begin{aligned}
J &= \begin{vmatrix} \frac{\partial A}{\partial \beta_1} & \frac{\partial A}{\partial \beta_2} \\ \frac{\partial \phi}{\partial \beta_1} & \frac{\partial \phi}{\partial \beta_2} \end{vmatrix} \\
&= \begin{vmatrix} \frac{\beta_1}{\sqrt{\beta_1^2+\beta_2^2}} & \frac{\beta_2}{\sqrt{\beta_1^2+\beta_2^2}} \\ -\frac{\beta_2/\omega}{\beta_1^2+\beta_2^2} & \frac{\beta_1/\omega}{\beta_1^2+\beta_2^2} \end{vmatrix} \\
&= \frac{1}{\omega\sqrt{\beta_1^2+\beta_2^2}}.
\end{aligned} \tag{9}
$$

Thus the transformed conditional posterior density of $\boldsymbol{\beta}$ is

$$
\begin{aligned}
P^*(\beta_1,\beta_2) \propto & \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{Y}-\boldsymbol{X\beta})\right\} \\
& \times\left(\frac{1}{\sqrt{\beta_1^2+\beta_2^2}}\exp\left\{-\frac{1}{\sigma_A^2}\left(\sqrt{\beta_1^2+\beta_2^2}-\mu_A\right)^2\right\}\right)
\end{aligned} \tag{10}
$$

Now we can perform slice sampling as follows:

(1) Sample u from $u \mid \beta_1,\beta_2 \sim U\left(0, \frac{1}{\sqrt{\beta_1^2+\beta_2^2}}\exp\left\{-\frac{1}{\sigma_A^2}\left(\sqrt{\beta_1^2+\beta_2^2}-\mu_A\right)^2\right\}\right)$

(2) Sample (β_1,β_2) from

$$
\begin{aligned}
\beta_1,\beta_2 \mid u \sim & \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{Y}-\boldsymbol{X\beta})\right\} \\
& \times I\left(\frac{1}{\sqrt{\beta_1^2+\beta_2^2}}\exp\left\{-\frac{1}{\sigma_A^2}\left(\sqrt{\beta_1^2+\beta_2^2}-\mu_A\right)^2\right\} > u\right)
\end{aligned}
$$

The function

$$
f(z=\sqrt{\beta_1^2+\beta_2^2}) = \frac{1}{z}\exp\left\{-\frac{1}{\sigma_A^2}(z-\mu_A)^2\right\}
$$

is a decreasing function on $z>0$, so if we solve for $\hat{z}$ s.t. $f(z)=u$, any z with $z<\hat{z}$ will satisfy the constraint.

Finally, Step II can be achieved by a two-step Gibbs sampling:

$$
\beta_1|\beta_2,u \sim \mathcal{N}\left(\mu_1+\sigma_{12}\frac{\sigma_{11}}{\sigma_{22}}(\beta_2-\mu_2), \sigma_{11}^2(1-\sigma_{12}^2)\right) I\left((|\beta_1| < \sqrt{\hat{z}-\beta_2^2}\right),
$$

$$
\beta_2|\beta_1,u \sim \mathcal{N}\left(\mu_2+\sigma_{12}\frac{\sigma_{22}}{\sigma_{11}}(\beta_1-\mu_1), \sigma_{22}^2(1-\sigma_{12}^2)\right) I\left(|\beta_2| < \sqrt{\hat{z}-\beta_1^2}\right).
$$

Step III: Update between $\mathfrak{M}_1$ and $\mathfrak{M}_0$ using reversible jump sampling

Next, we show the between model updating procedure using reversible jump. For reversible jump, when selecting between models with different dimensions, we use augmenting variables to match the parameter space dimension across models. The parameter space of the two models that we want to select from are $\mathfrak{M}_0 : \boldsymbol{\beta_0} = \{M', \sigma^{2\prime}\}$ and $\mathfrak{M}_1 : \boldsymbol{\beta_1} = \{M, \sigma^2, A, \phi\}$.

Suppose the current state is $(\mathfrak{M}_0, \boldsymbol{\beta_0})$, and we propose augmenting variables $\mathbf{u} = \{A, \phi\}$. The acceptance ratio of reversible jump is calculated as below:

$$r = \frac{P(\mathbf{y} \mid \boldsymbol{\beta_1})\, P(\phi)\, P(A)}{P(\mathbf{y} \mid \boldsymbol{\beta_0})} \frac{p}{1-p} \frac{J(\mathfrak{M}_0 \mid \mathfrak{M}_1)}{J(\mathfrak{M}_1 \mid \mathfrak{M}_0)} \frac{1}{g(\phi)\, g(A)} \left|J_{1,2}\right|.$$

In this formula, p, as the prior of $P(\rho = 1)$, is set to $p = 0.2$ to reflect the prior knowledge of proportion of circadian biomarkers. The probability of transition proposal J between models is set to 0.5 in both direction, resulting in $\frac{J(\mathfrak{M}_0|\mathfrak{M}_1)}{J(\mathfrak{M}_1|\mathfrak{M}_0)} = 1$. This indicates that the probability of transitioning from model $\mathfrak{M}_0$ to $\mathfrak{M}_1$ is equal to the probability of transitioning from model $\mathfrak{M}_1$ to $\mathfrak{M}_0$. Furthermore, $|J_{1,2}|$ represents the Jacobian of the one-to-one transformation from the parameter space $\{\boldsymbol{\beta_0}, \mathbf{u}\}$ under model $\mathfrak{M}_0$ to the parameter space $\{\boldsymbol{\beta_1}\}$ under model $\mathfrak{M}_1$. Since no parameter transformation is performed between the two models, the Jacobian determinant equals 1. Finally, the $g(A)$ and $g(\phi)$ are the proposal probability of the two parameters, which we set to be their prior distritions.

After simplification by plugging the above values, we get

$$r = \min\left(\frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - X_i - M)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - M)^2\right]} \frac{P(A)P(\phi)}{g(A)g(\phi)}, 1\right). \tag{11}$$

The sampling procedure is as follows, with probability 0.5, the chain stays at the current state and performs corresponding samplings, otherwise, the chain will jump to the other state with the acceptance ratio r if we are jumping from $\mathfrak{M}_0$ to $\mathfrak{M}_1$ and $1/r$ otherwise.

Step IV: Update Δt_i in BayCT

Finally, the posterior distribution of Δt_i in the BayCT model is updated as:

$$P^*(\Delta t_i) \propto \prod_{g:c_g=1} P\left(Y_g \mid A_g, \phi_g, t, \Delta t, M_g, \sigma_g^2\right) \cdot P\left(\Delta t_i \mid \theta', k'\right)$$

$$\propto \exp\left\{\sum_{g:c_g=1} -\frac{1}{2\sigma_g^2}\left[Y_{gi} - M_g - A_g \cos\left(\omega\left(t_i + \Delta t_i - \phi_g\right)\right)\right]^2 + \kappa' \cos\left(\omega(\Delta t_i - \theta')\right)\right\}.$$

Here, the set of genes with index $c_g = 1$ is chosen to update Δt. Without prior knowledge, we update $g : c_g = 1$ as the rhythmic genes identified in the last MCMC iteration, i.e., $\{g : c_g = 1\} = \{g : \rho_g = 1\}$. In real application, we pre-select a small set of core clock genes with known strong circadian rhythmicity, denoted by $s_g = 1$, and Δt is updated with genes that are both pre-selected and estimated as rhythmic, i.e., $\{g : c_g = 1\} = \{g : s_g \cdot \rho_g = 1\}$.

If we denote $P^*(\Delta t_i) \propto \exp\{f(\Delta t_i)\}$, the sampling procedure is as following:

1. Draw $u \sim U(0, \exp\{f(\Delta t_i)\})$, which is equivalent to firstly drawing $z \sim \exp(1)$ and then calculate $\log u = f(\Delta t_i) - z$.
2. Calculate the range of $\left\{a \le \Delta t_i^{(\text{new})} \le b : f\left(\Delta t_i^{(\text{new})}\right) > \log u\right\}$.
3. Draw $\Delta t_i^{(\text{new})} \sim U(a, b)$.

2.3. *BayCT-M model.* The BayCT-M extends BayCT model from single brain region to multiple brain regions. We denote the expression level of gene g $(g = 1, 2, \ldots, G)$ in brain region j $(j = 1, 2, \ldots, J)$ and subject i $(i = 1, 2, \ldots, n_j)$ at time point t_{ji} as Y_{gji}. In the particular setting of our motivating example, data from the three brain regions are from $N = 59$ identical subjects. As such, $n_1 = n_2 = n_3 = N = 59$ and $t_{1i} = t_{2i} = t_{3i} = t_i$. Notably, our

implementation in R also allows for partially overlapped subjects across brain regions, i.e., $n_{j_1} \neq n_{j_2}$ for some j_1, j_2. The cosinor model is parameterized as:

$$Y_{gji} = A_{gj} \cos\left(\omega\left(t_i + \Delta t_i - \phi_{gj}\right)\right) + M_{gj} + \epsilon_{gji}, \tag{12}$$

where $t_i^+ = t_i + \Delta t_i$ is the MCT. The priors are the same as in BayCT, except that ϕ_{gj} shares a common mean θ_g in the von Mises prior across brain regions. The A_{gj}, σ_{gj}^2, and M_{gj} have more considerable variance across groups and thus are assigned independent priors for each tissue (Zhang et al., 2014; Mure et al., 2018; Talamanca, Gobet and Naef, 2023):

$$\begin{aligned} M_{gj} &\sim N(\mu_M, \sigma_M^2), \\ \sigma_{gj}^2 &\sim IG(V_0/2, V_0\sigma_0^2/2), \\ A_{gj} &\sim (1-\rho_{gj})\delta_0 + \rho_{gj} N(\mu_A, \sigma_A^2) I(A_{\min}, +\infty), \\ \phi_{gj} &\sim VM(\omega\theta_g, \kappa_g A_{gj}) * (1/\omega), \\ \Delta t_i &\sim VM(0, \kappa_t) * (1/\omega). \end{aligned} \tag{13}$$

Similar to BayCT, the indicator variable $\rho_{gj} = 1$ indicates gene g in brain region j is rhythmic and $\rho_{gj} = 0$ otherwise. We set $\mu_M = 0, \mu_A = 1, \sigma_M^2 = \sigma_A^2 = 10^2, \sigma_0^2 = 0, V_0 = 2$ and $\kappa_t = 6$. The dependence of ϕ_{gj} prior on A_{gj} is grounded in the rationale that a smaller A_{gj} implies a weaker signal and thus a greater dispersion in ϕ_{gj}. We further define the conjugate joint hyperprior for the hyperparameters θ_g and κ_g:

$$P(\theta_g, \kappa_g) \sim \{I_0(\kappa_g)\}^{-c_0} \exp\left\{\kappa_g R_0 \cos\left(\omega\theta_g - \omega\theta_0\right)\right\}. \tag{14}$$

In this prior distribution, $I_0(\kappa_g)$ is the Bessel function of the first kind with order 0. The hyperprior indicates c_0 number of prior observations in the direction θ_0, and R_0 corresponds to the component projected along θ_0 (i.e., in the known direction) of the resultant of c_0 observations (Damien and Walker, 1999). We set $c_0 = 1, R_0 = 1$, and $\theta_0 = 0$, which constitutes a weakly informative prior, providing a baseline orientation while allowing the data to largely influence the posterior estimates. The corresponding Gibbs sampling procedure proposed in (Damien and Walker, 1999) is used for updating θ_g and κ_g. Figure 1B shows the graphical representation of BayCT-M under $\mathfrak{M}_0(\rho_{gj} = 0)$ and $\mathfrak{M}_1(\rho_{gj} = 0)$. The whole MCMC sampling procedures for the BayCT-M are detailed in Supplement Section 1 Posterior Inference. Inference outputs (posterior mean and standard deviation) of four parameters from BayCT-M are: (1) ρ_{gj}: circadian indicator for gene g in tissue j, (2) θ_g: canonical phase of gene g, (3) ϕ_{gj}: tissue-specific phase for a circadian gene g, (4) Δt_i: deviation of molecular circadian time of subject i.

3. Simulations.

3.1. *Simulation to evaluate BayCT model.* We generate data for 1000 genes, consisting of 200 rhythmic and 800 arrhythmic genes, to reflect the higher prevalence of arrhythmic genes in real applications. For the rhythmic genes, we set amplitude $A = 0.6, 0.8, 1$, with a constant variance $\sigma^2 = 1$. The time deviation is varied along two dimensions: the magnitude of deviation $|\Delta t|$, set at 0, 1, 2, and 4 hours, and the proportion of the total 30 samples affected by this deviation, set at $r = 0, 20\%, 40\%$, with equal numbers of samples having positive and negative $|\Delta t|$ values. Each setting is repeated 30 times.

We first evaluate MCT estimation accuracy, comparing BayCT with various methods, including TimeSignature (TS), ZeitZeiger (ZZ), CHIRAL, CYCLOPS, and Cyclum. Note that CHIRAL, CYCLOPS and Cyclum are not standard supervised machine learning methods. These tools are unsupervised without utilizing the input TOD. They, however, can generate

pseudo-time, which is prediction of the MCT of each subject with certain unknown shift. Consequently, these three methods are not applicable to perform cross-validation in the MCT prediction evaluation. In Figure 2, the prediction accuracy is calculated by using all subjects to construct the model and then applied to the same subjects. This may cause slight overfitting in the accuracy evaluation but it gives a fair method comparison. To avoid overfitting, Figure S1 conducts five-fold cross validation and compare methods excluding CHIRAL, CYCLOPS and Cyclum.

Next, we compare biomarker detection accuracy among BayCT, BayC and the Cosinor model, using the area under the Receiver Operating Characteristic curve (AUC) as a threshold-free measure. For Bayesian methods, sensitivity and specificity are calculated with varying cutoffs for $P^*(\rho = 1)$, while the frequentist Cosinor method uses varying p-value cutoffs. Additionally, we assess the parameter estimation consistency through the bias and standard deviation of the point estimators from each method.

MCT estimation accuracy: The simulations in Figure 2A shows that BayCT and TS give the most accurate MCT prediction, regardless of Δt and r. In Figure 2B, when there is no time deviation (i.e., $\Delta t_i = 0$ for all i), BayCT has slightly larger standard deviation (SD) than TS due to slight loss of efficiency by estimating Δt. But when time deviation indeed exist, BayCT outperforms TS as well as other methods. Larger and more prevalent time deviations (e.g., $|\Delta t| = 4$) lead to more significant superiority in BayCT.

An extra advantage of BayCT over other machine learning methods is its inference of a Bayesian credible interval for the MCT prediction uncertainty. Supplement Figure S2 shows the actual coverage probability when the 95% credible interval of MCT contains the truth. The result shows general accurate estimation of the 95% credible intervals.

Circadian biomarker detection: Figure 3A presents the area under the curve (AUC) evaluation of circadian gene detection using BayC, BayCT and frequentist Cosinor model. In the baseline scenario where $\Delta t_i = 0$ for all i (i.e., when MCT aligns perfectly with TOD for all samples), BayC and Cosinor achieve comparable high accuracy. The result is consistent with the general observation that Bayesian approaches and their frequentist counterparts often achieve similar high performance when both modelled adequately. BayCT has a slightly lower AUC performance, due to the slight loss of efficiency from estimating Δt_i when they are all zero. However, as the degree of deviance grows, BayCT becomes increasingly effective, surpassing the other two methods. For example, when $A = 0.8$, and $r = 40\%$ of samples have $|\Delta t| = 4$, BayeCT achieves mean AUC=0.909, compared to mean AUC of 0.869 and 0.852 for BayC and Cosinor. Notably, AUC of BayCT remains almost at the same level for fixed A, showing effective time deviation correction and the robustness provided by BayCT.

We further assess estimation accuracy in terms of bias and standard deviation of amplitude A, phase ϕ, and expression error term σ^2 for BayC, BayCT and Cosinor. The result at simulating $A = 0.8$ is presented in Figure 3B. In terms of estimating A and σ^2, BayCT maintains an unbiased estimation performance, whereas both BayC and Cosinor tend to underestimate A and overestimate σ^2 with increasing proportion of samples with time deviation, r. Since A/σ stands for signal-to-noise ratio, this result can be explained by that time deviation introduces uncertainty in signal detection. In terms of ϕ, all three methods are unbiased when r increases. BayC and Cosinor, however, have much larger SD compared to BayCT when r increases.

3.2. *Simulations to evaluate BayCT-M model.* We simulate 2000 genes in 3 brain regions from 30 samples, that is, $G = 2000, J = 3$, and $n_j = 30$. Then, we simulate data for 1300 arrhythmic genes in all 3 regions, 100 genes rhythmic only in each tissue (300 genes in totla), 100 genes rhythmic in any two tissues (300 genes in total), and 100 genes rhythmic in all 3 regions. We set amplitude $A_{gj} = 0.6, 0.8, 1$ and a constant variance $\sigma^2_{gj} = 1$ to generate

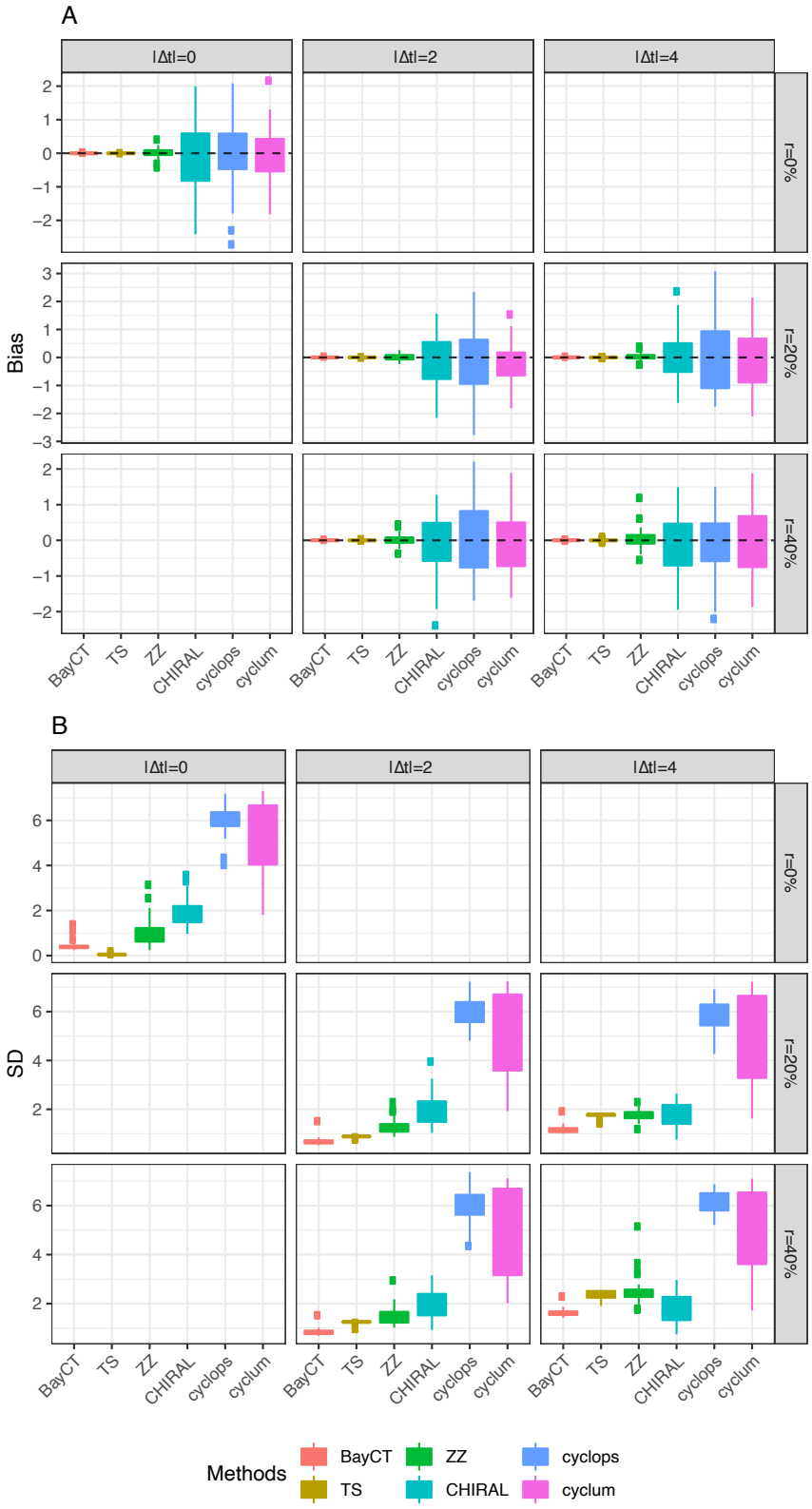

FIG 2. *Bias and standard deviation (SD) of the MCT estimations.*

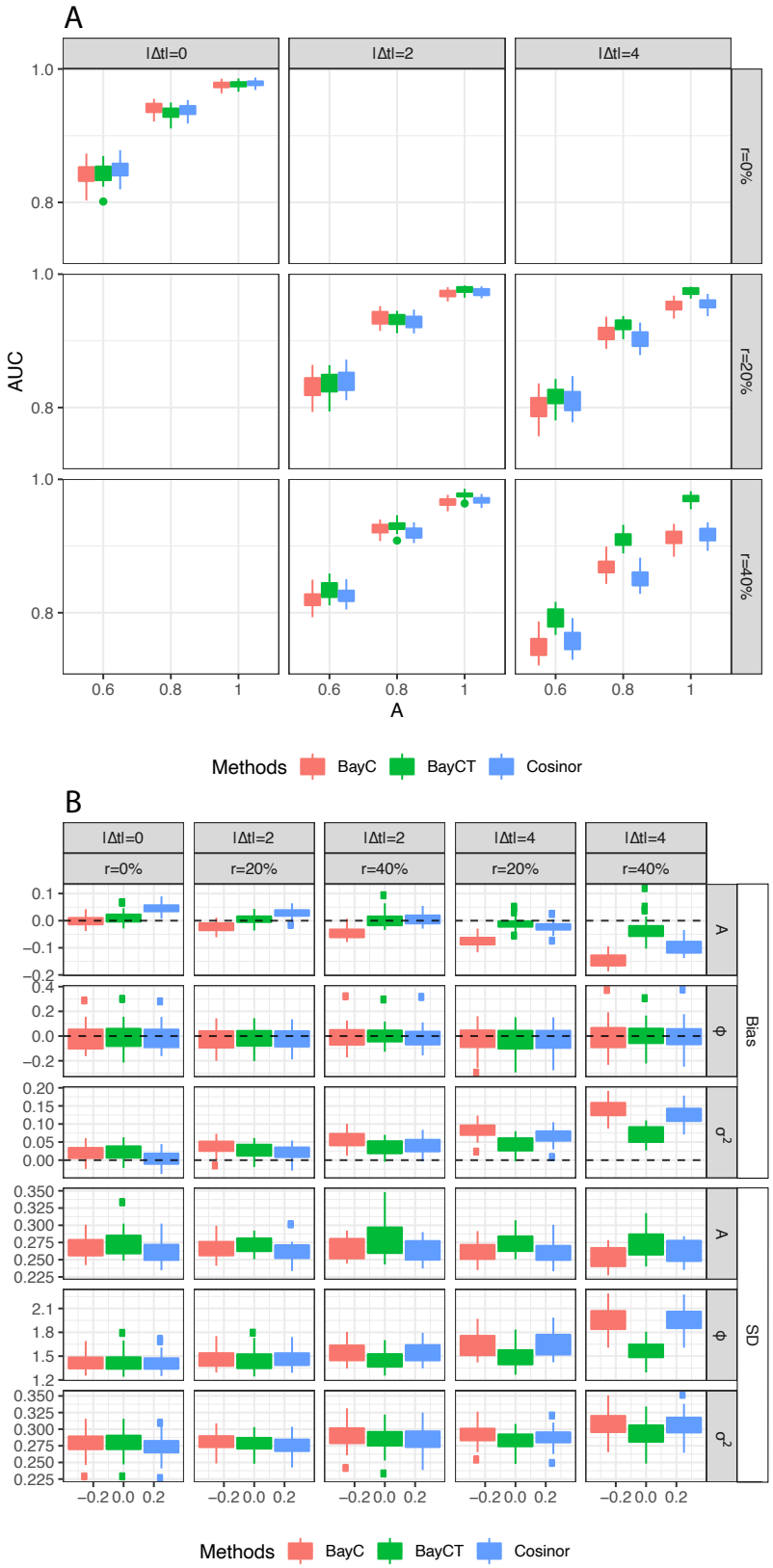

FIG 3. *A. Comparison of rhythmicity accuracy of the three methods with AUC. The columns, moving from the left to right, represent an increasing magnitude of* Δt*, while the rows, from top to bottom, present an increasing percentage of samples affected by* Δt*. B. Bias and SD evaluation for the three methods when simulating* $A = 0.8$*.*

adequate intrinsic effect size (i.e., A/σ) for sufficient statistical power (Zong et al., 2023). Similar to the previous subsection, the circadian time deviation varies in two dimensions: 1) the magnitude of deviation, $|\Delta t| \in \{0, 2, 4\}$. 2) the proportion of total samples affected by this deviation was set at $r = 0\%, 20\%, 40\%$, with equal samples with positive and negative $|\Delta t|$ values. We categorize simulations into two settings: 1) SamePhase: ϕ_{gj} are same across tissues for rhythmic genes. 2) DiffPhase: ϕ_{gj} are different across tissues for rhythmic genes with magnitude 2 and 4, i.e., $\phi_{g2} - \phi_{g1} = 2$ and $\phi_{g3} - \phi_{g1} = 4$. Each combination of these variations is simulated 30 times. The performance of BayCT-M in the two settings are highly similar, demonstrating robust capacity of BayCT-M under heterogeneity.

MCT estimation accuracy: Supplement Figure S3 and Figure S5 show bias and SD of MCT estimates under SamePhase and DiffPhase simulations, respectively. BayCT-M consistently gives unbiased estimates regardless of $|\Delta t|$ and r. As amplitude A increases, BayCT-M gives more accurate estimates with smaller SD, indicating that the model is efficient if we have strong rhythmicity signal.

To assess the efficiency of BayCT-M relative to BayCT in estimating Δt, we apply BayCT separately on three tissues from the simulated data and compute the SD ratio of BayCT-M to BayCT, as shown in Figure 4. Notably, the efficiency gain is most pronounced at lower amplitude ($A = 0.6$), where the SD ratio drops below $1/\sqrt{3}$ (i.e., the expected asymptotic optimal efficiency). As the amplitude increases, these ratios converge to $1/\sqrt{3}$. The result shows that BayCT-M is asymptotically efficient in integrating information across brain regions for predicting MCT.

Circadian biomarker detection: The evaluation on circadian gene detection accuracy in Figure S4 and Figure S6 show that BayCT-M is capable to identify circadian gene with high performance. The AUC approaches 1 when the signal is strong enough (i.e., $A = 1$). Comparisons in Figure S7 confirm that BayCT-M performs comparably to the single-tissue BayCT method. BayCT-M shows a slight advantage at lower signal amplitudes, indicating that integrating multi-tissue information preserves robust biomarker detection.

4. Real applications. In this section, we evaluate performance of BayCT and BayCT-M with two real datasets. In the first application, since biological/genetic and environmental variations among mice are minimal, we assume TOD=MCT (i.e., $\Delta t = 0$) in the mouse study. We simulate Δt to generate semi-synthetic data from real data to evaluate MCT prediction performance of BayCT and BayCT-M. In the second application, we apply BayCT to a human postmortem transcriptomic dataset with three brain regions for both biomarker detection and MCT estimation.

4.1. *Evaluation in mouse data with synthetic* Δt. Circadian gene markers have been found to be tissue-specific. As such, a transcriptomic study in mouse by Zhang et al. (2014) provided an circadian atlas composing 12 tissues, each sampled in every 2 hours for 2 days (24 samples per tissue). Since all mice were entrained with the same schedule before the sample collection, the MCT is considered to be aligned perfectly to the experiment time (i.e., $\Delta t = 0$).

After filtering out probes with negative values and duplication, 35,556 genes (transcripts) are left and we further normalize the data with quantile normalization. To assess the MCT prediction methods on real data, we generate pseudo datasets by artificially introducing noise to MCT. Specifically, for a randomly selected half of the samples, we add Δt to the observed time, while for the other half, we subtract Δt from the observed time to create the new observed TOD. This process simulates scenarios where the recorded TOD might deviate from the underlying MCT due to various factors.

Then we apply the BayCT and other MCT prediction methods to the data using the pseudo time as the input. For BayCT, we update $|\Delta t|$ for the 12 tissues separately with the 11 present

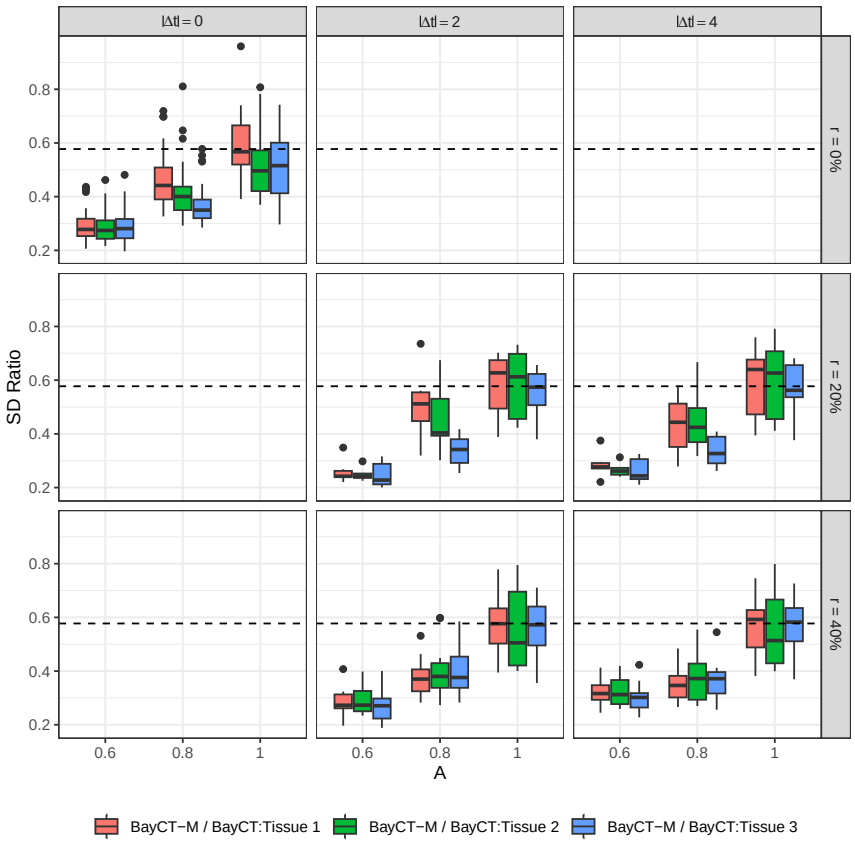

FIG 4. *The SD ratio of BayCT-M to BayCT on each tissue with increasing amplitude A when estimating Δt. The horizontal dashed line is $1/\sqrt{3}$. The columns, moving from the left to right, represent an increasing magnitude of Δt, while the rows, from top to bottom, present an increasing percentage of samples affected by Δt.*

core rhythmic genes: NPAS2, PER2, CRY1, CIART, PER1, NR1D1, NR1D2, TEF, CRY2, PER3 and DBP, while simultaneously estimating the rhythmicity parameters for all the genes. The results show that the 95% credible intervals (CIs) covered 94.4% of the artificial Δt, indicating the model's ability to accurately capture the introduced noise. Additionally, we apply other MCT prediction methods with the same set of 11 core clock genes, using procedures described in Section 3. As shown in Figure 5A, the comparison yielded results similar to those from our simulation, with BayCT model exhibiting consistent unbiasedness and small variance overall. These findings demonstrate the BayCT's robust performance on real data.

Subsequently, we evaluate the accuracy of rhythmicity identification of BayC, BayCT, and Cosinor with pseudo $|\Delta t| = 1, 2, 4$. To obtain a reference rank of gene rhythmicity which we can compare to, we applied the three models on the original data (when $|\Delta t = 0|$) and obtain the gene rhythmicity rank for each method. For BayC and BayCT, genes with larger $P^*(\rho = 1)$ have smaller ranks, while for Cosinor, genes with smaller p-value have smaller ranks. The "ensemble reference" top genes were ordered by the smallest average rank across the three methods. We then examin the percentage of the n-top-ranked rhythmic genes from the "ensemble reference list" that are overlapping with the n-top-ranked rhythmic genes from each method, with $n = 100, 200, 300$, and 500. Figure 5B shows that when $|\Delta t| = 0$, BayC has the highest overlapping percentage. However, as $|\Delta t|$ increases, both BayC and Cosinor lost accuracy in identifying the top rhythmic genes. In contrast, BayCT displays robust performance across varying $|\Delta t|$, with the results being comparable to results at $|\Delta t| = 0$. This

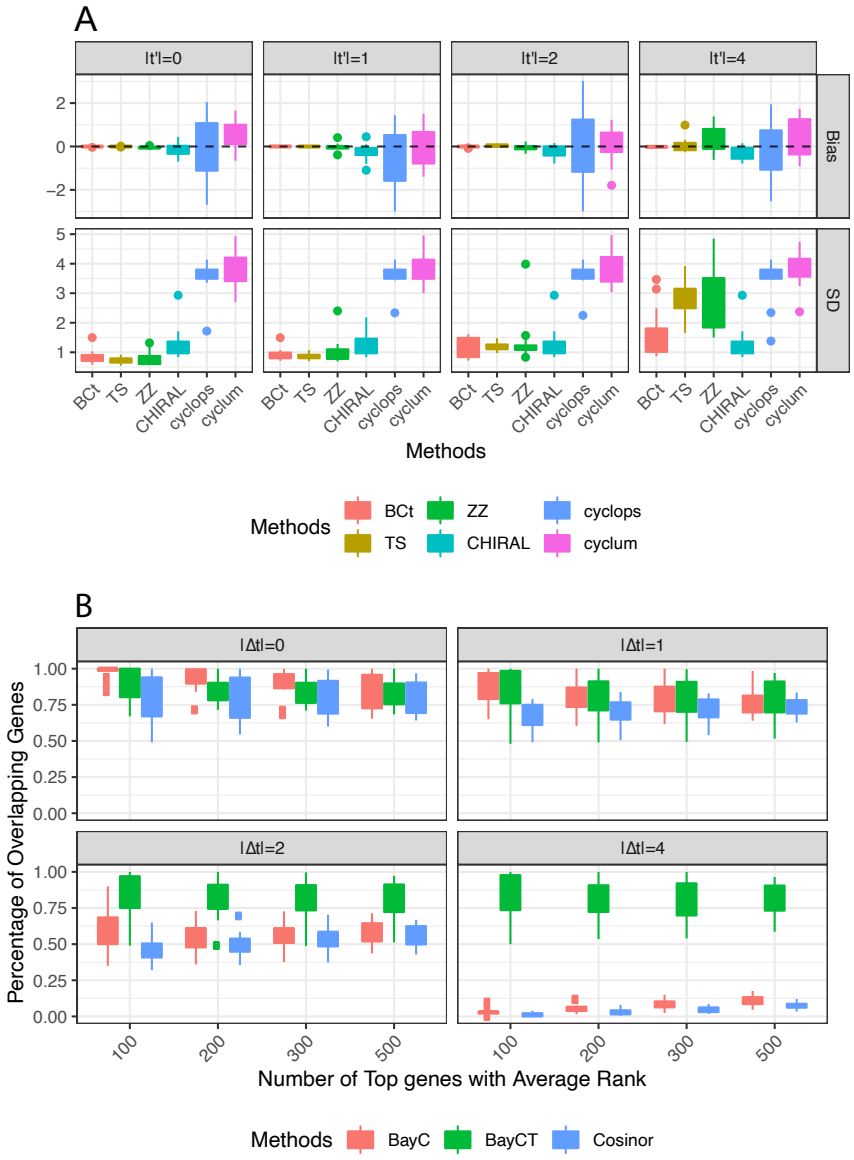

FIG 5. *A. Bias and SD evaluation for MCT prediction, pooled over 12 tissues. B. The overlapping of top rhythmic genes for BayC, BayCT, and Cosinor with increasing artificial* Δt.

demonstrates the BayCT's overall accurate and robust identification of circadian genes under the impact of varying $|\Delta t|$.

4.2. *Human postmorten data with three brain regions.* In the second case study, we apply BayCT and BayCT-M to RNA-seq data from Ketchesin et al. (2021), which involves samples from three distinct brain regions - nucleus accumbens (NAc), caudate and putamen - of the same 59 subjects.

We independently apply BayCT to each of the three brain regions and use BayCT-M on all three brain regions. Figure 6A shows the BayCT-M Δt ordered by the mean with the corresponding 90% credible interval. All of the BayCT-M Δt have a 90% credible interval width of less than 2.4 hours, showing high concordance of Δt estimates across three brain regions.

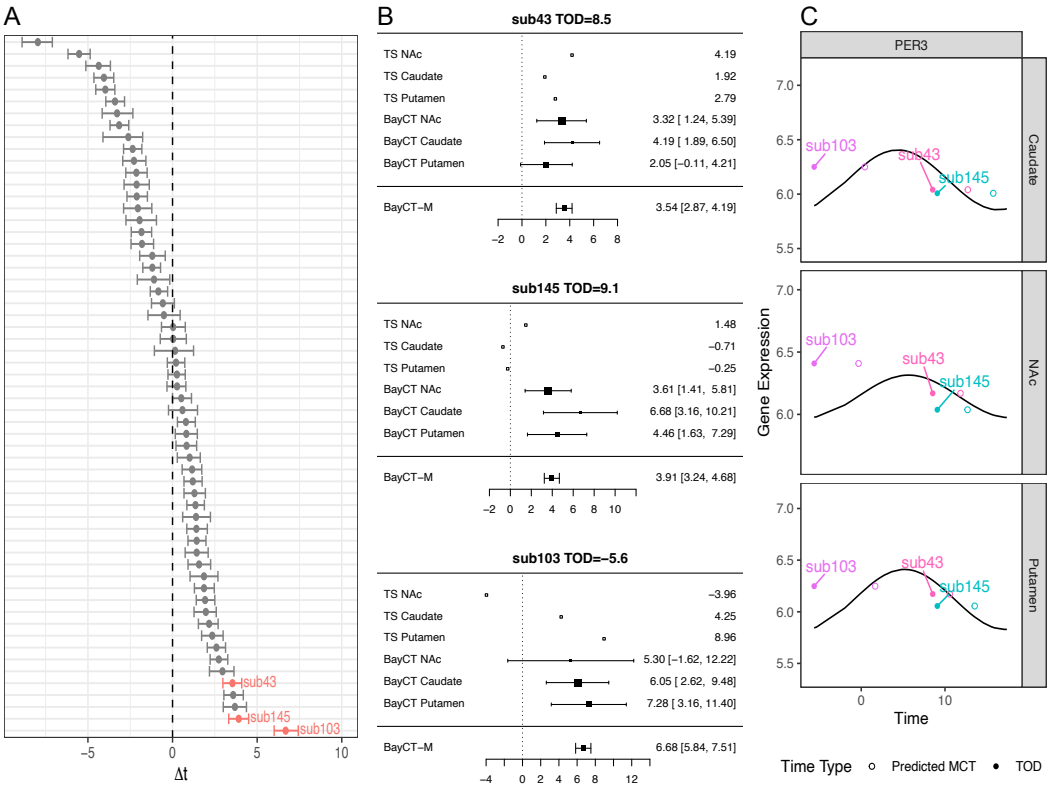

FIG 6. *The BayCT, BayCT-M, and TS Δt estimates for three brain regions. A. BayCT-M Δt estimates and 90% credible interval for all subjects. B. Forest plots of subject 43, 145, and 103. C. Circadian gene expression pattern of PER3 in Caudate, Nac, and Putamen. x axis is TOD or predicted MCT and y axis is gene expression level. The black line is the fitted sinusoidal curve using TOD*

For further comparison, we compare these BayCT-M Δt estimates with those obtained from TimeSignature (TS), which is identified as the second-best performer in the extensive simulation. Remarkably, the comparison reveals a high degree of similarity, with 54 out of 59 (91.5%) samples exhibiting a difference of less than 44 hours (Figure S8). For demonstration, Figure 6B provides forest plots of BayCT Δt in three subjects, BayCT-M Δt, and the Δt estimations from TS. Since TS is a machine learning method, only the Δt point prediction is available. In contrast, BayCT and BayCT-M provides uncertainty quantification by 95% credible interval, in addition to the mean point estimation. In Subject 43, the mean $\Delta t = 3.54$ and narrow Δt credible interval $= [2.87, 4.19]$ from BayCT-M shows high inference confidence of the difference. TS predictions in three brain regions gives consistent conclusions of $\Delta t = 4.19, 1.92$ and 2.79, respectively. In subject 145, BayCT-M also concludes highly confident Δt with mean $= 3.91$ and credible interval $= [3.24, 4.68]$. In contrast, TS shows inconclusive Δt prediction at $1.48, -0.71$ and -0.25. In Subject 103, BayCT concludes consistent signals for Δt across three brain regions although NAc has larger variance. As such, the BayCT-M result gives significant conclusion of mean $\Delta t = 6.68$ and credible interval $= [5.84, 7.51]$. Prediction result from TS is relatively consistent with BayCT in Caudate and Putamen ($\Delta t = 4.25$ and 8.96) while very different in NAc ($\Delta t = -3.96$). Again TS can only provide predication without uncertainty quantification in the prediction compared to BayCT. We also notice that BayCT-M gives us much narrower credible interval in this real application. This is possibly due to finite sample or data noise and our simulation results in Figure 4

guarantees that BayCT-M Δt estimate has a reasonable efficiency when rhythmicity signal is strong. In Figure 6C, we use a core clock gene PER3 to illustrate how MCT adjusts from TOD to provide better circadian curve fitness. PER3 appears to be a circadian genes in all three brain regions (posterior probability of $\rho_{gj} = 1$ is 1, for $j = 1, 2, 3$). Expression levels plotted with MCT (circles) fit the circadian curves much better than TOD (solid dots), except for poor fitness in both MCT and TOD for NAc in Subject 103. This raw data visualization confirms the likely time deviations Δt in these three subjects.

5. Discussion. The inherent biological variability of MCT presents significant challenges to not only the chronotherapy but also the broader researches in human circadian biology. Traditional studies often equates the TOD with MCT, introducing measurement noise into the analyses. On the other hand, existing time prediction methods focus on estimating MCT, but not on inferences of circadian biomarkers and their associated rhythmic parameters.

In response to these challenges, we developed BayCT, a novel Bayesian circadian detection model that precisely adjusts the discrepancies between the observed time and MCT. The model employs full Bayesian inference, enabling the inference of both the rhythmic parameters in detected biomarkers and the uncertainty in MCT estimation. BayCT has shown superior performance in circadian biomarker detection compared to the traditional frequentis method Cosionor and BayC without MCT adjustment. It also provides more accurate MCT estimation than other machine learning approaches. Moreover, BayCT's extension to BayCT-M enables us to deal with multi-tissue data within a full Bayesian framwork.

The application of BayCT to real datasets further validates its effectiveness. By introducing pseudo Δt into the mouse data, we are encouraged by its robust performance in estimating the MCT and rhythmicity identification across various levels of noise strength. Additionally, the concordant Δt estimations across multiple human brain tissues demonstrates the reliability of its results.

In our motivating example, data from the three brain regions are from $N = 59$ identical subjects. However, in our R package, BayCT, the BayCT-M method can implement the general situation when data across brain regions or organs are from partially overlapped subjects.

BayCT is designed assuming Gaussian distributed residual errors, making it applicable for a variety of omics data types, such as transcriptomics, proteomics, and log-transformed methylation data. For the human postmorten RNA-seq data from Ketchesin et al. (2021) with $G = 14,540$ transcripts, $J = 3$ brain regions, and $n_j = 59$ subjects, BayCT-M takes 4.45 hours to complete 3,000 MCMC simulations on a personal laptop with 18GB RAM. The computing is expected to significantly improve if calling Rcpp from R.

In summary, BayCT enhances circadian analysis by adjusting for discrepancies between the MCT and observed TOD via a full Bayesian inference framework. It presents robust circadian detection and inference with varying levels of biological variation of MCT, which enhance the power and accuracy of circadian analysis.

Acknowledgments. The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding. This work was supported by NIH Award Number R01LM014142 and R01CA285337. This research was supported in part by the University of Pittsburgh Center for Research Computing and Data, RRID:SCR_022735, through the resources provided. Specifically, this work used the HTC cluster, which is supported by NIH award number S10OD028483.

SUPPLEMENTARY MATERIAL

Supplement to "Bayesian rhythmic model for jointly detecting circadian biomarkers and predicting molecular circadian time in human post-mortem brain transcriptome" The supplement contains detailed posterior inference procedures for the BayCT and BayCT-M models, along with additional simulation results and supplementary figures.

REFERENCES

AGOSTINELLI, F., CEGLIA, N., SHAHBABA, B., SASSONE-CORSI, P. and BALDI, P. (2016). What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics* **32** i8–i17.

ANAFI, R. C., FRANCEY, L. J., HOGENESCH, J. B. and KIM, J. (2017). CYCLOPS reveals human transcriptional rhythms in health and disease. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/PNAS.1619320114

ASCHOFF, J. (1960). Exogenous and endogenous components in circadian rhythms. In *Cold Spring Harbor symposia on quantitative biology* **25** 11–28. Cold Spring Harbor Laboratory Press.

AUERBACH, B. J., FITZGERALD, G. A., LI, M., AUERBACH, B. J., FITZGERALD, G. A. and LI, M. (2022). Tempo: an unsupervised Bayesian algorithm for circadian phase inference in single-cell transcriptomics. https://doi.org/10.1038/S41467-022-34185-W

BRAUN, R., KATH, W. L., IWANASZKO, M., KULA-EVERSOLE, E., ABBOTT, S. M., REID, K. J., ZEE, P. C. and ALLADA, R. (2018). Universal method for robust detection of circadian state from gene expression. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/PNAS.1800314115

BUTTE, A. J. and HASTIE, T. (2016). ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. https://doi.org/10.1093/NAR/GKW030

CEDERROTH, C. R., ALBRECHT, U., BASS, J., BROWN, S. A., DYHRFJELD-JOHNSEN, J., GACHON, F., GREEN, C. B., HASTINGS, M. H., HELFRICH-FÖRSTER, C., HOGENESCH, J. B. et al. (2019). Medicine in the fourth dimension. *Cell metabolism* **30** 238–250.

CHEN, C.-Y., LOGAN, R. W., MA, T., LEWIS, D. A., TSENG, G. C., SIBILLE, E. and MCCLUNG, C. A. (2016). Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proceedings of the National Academy of Sciences* **113** 206–211.

CORNELISSEN, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling* **11** 1–24.

CROWLEY, S. J. and EASTMAN, C. I. (2018). Free-running circadian period in adolescents and adults. *Journal of sleep research* **27** e12678.

DALLMANN, R., OKYAR, A. and LÉVI, F. (2016). Dosing-time makes the poison: circadian regulation and pharmacotherapy. *Trends in molecular medicine* **22** 430–445.

DAMIEN, P. and WALKER, S. (1999). A full Bayesian analysis of circular data using the von Mises distribution. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* 291–298.

DING, H., MENG, L., LIU, A. C., GUMZ, M. L., BRYANT, A. J., MCCLUNG, C. A., TSENG, G. C., ESSER, K. A. and HUO, Z. (2021). Likelihood-based tests for detecting circadian rhythmicity and differential circadian patterns in transcriptomic applications. *Briefings in Bioinformatics* **22** bbab224.

EASTMAN, C. I., SUH, C., TOMAKA, V. A. and CROWLEY, S. J. (2015). Circadian rhythm phase shifts and endogenous free-running circadian period differ between African-Americans and European-Americans. *Scientific reports* **5** 8381.

FINGER, A.-M. and KRAMER, A. (2021). Mammalian circadian systems: Organization and modern life challenges. *Acta Physiologica* **231** e13548.

GLYNN, E. F., CHEN, J. and MUSHEGIAN, A. R. (2006). Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms. *Bioinformatics* **22** 310–316.

HUGHES, M. E., HOGENESCH, J. B. and KORNACKER, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of biological rhythms* **25** 372–380.

JONES, S. E., LANE, J. M., WOOD, A. R., VAN HEES, V. T., TYRRELL, J., BEAUMONT, R. N., JEFFRIES, A. R., DASHTI, H. S., HILLSDON, M., RUTH, K. S. et al. (2019). Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nature communications* **10** 343.

KETCHESIN, K. D., ZONG, W., HILDEBRAND, M. A., SENEY, M. L., CAHILL, K. M., SCOTT, M. R., SHANKAR, V. G., GLAUSIER, J. R., LEWIS, D. A., TSENG, G. C. et al. (2021). Diurnal rhythms across the human dorsal and ventral striatum. *Proceedings of the National Academy of Sciences* **118** e2016150118.

LIANG, S., WANG, F., HAN, J. and CHEN, K. (2020). Latent periodic process inference from single-cell RNA-seq data. *Nature communications* **11** 1441.

MURE, L. S., LE, H. D., BENEGIAMO, G., CHANG, M. W., RIOS, L., JILLANI, N., NGOTHO, M., KARIUKI, T., DKHISSI-BENYAHYA, O., COOPER, H. M. et al. (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* **359** eaao0318.

PITTENDRIGH, C. S. (1960). Circadian rhythms and the circadian organization of living systems. In *Cold Spring Harbor symposia on quantitative biology* **25** 159–184. Cold Spring Harbor Laboratory Press.

PRINTEZI, M. I., KILGALLEN, A. B., BOND, M. J., ŠTIBLER, U., PUTKER, M., TESKE, A. J., CRAMER, M. J., PUNT, C. J., SLUIJTER, J. P., HUITEMA, A. D. et al. (2022). Toxicity and efficacy of chronomodulated chemotherapy: a systematic review. *The Lancet Oncology* **23** e129–e143.

RUBEN, M. D., SMITH, D. F., FITZGERALD, G. A. and HOGENESCH, J. B. (2019). Dosing time matters. *Science* **365** 547–549.

SCHERNHAMMER, E. S., LADEN, F., SPEIZER, F. E., WILLETT, W. C., HUNTER, D. J., KAWACHI, I. and COLDITZ, G. A. (2001). Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. *Journal of the national cancer institute* **93** 1563–1568.

TÄHKÄMÖ, L., PARTONEN, T. and PESONEN, A.-K. (2019). Systematic review of light exposure impact on human circadian rhythm. *Chronobiology international* **36** 151–170.

TALAMANCA, L., GOBET, C. and NAEF, F. (2023). Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans. *Science* **379** 478–483.

UEDA, H., CHEN, W., MINAMI, Y., HONMA, S., HONMA, K., IINO, M. and HASHIMOTO, S. (2004). Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/PNAS.0401882101

WEGRZYN, L. R., TAMIMI, R. M., ROSNER, B. A., BROWN, S. B., STEVENS, R. G., ELIASSEN, A. H., LADEN, F., WILLETT, W. C., HANKINSON, S. E. and SCHERNHAMMER, E. S. (2017). Rotating night-shift work and the risk of breast cancer in the nurses' health studies. *American journal of epidemiology* **186** 532–540.

ZHANG, R., LAHENS, N. F., BALLANCE, H. I., HUGHES, M. E. and HOGENESCH, J. B. (2014). A circadian gene expression atlas in mammals: implications for biology and medicine. *Proceedings of the National Academy of Sciences* **111** 16219–16224.

ZONG, W., SENEY, M. L., KETCHESIN, K. D., GORCZYCA, M. T., LIU, A. C., ESSER, K. A., TSENG, G. C., MCCLUNG, C. A. and HUO, Z. (2023). Experimental design and power calculation in omics circadian rhythmicity detection using the cosinor model. *Statistics in Medicine* **42** 3236–3258.