Research

A computational model for functional mapping of genes that regulate intra-cellular circadian rhythms

Tian Liu[1], Xueli Liu[1], Yunmei Chen[2] and Rongling Wu*[1]

Address: [1]Department of Statistics, University of Florida, Gainesville, FL 32611, USA and [2]Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

Email: Tian Liu - tianliu@stat.ufl.edu; Xueli Liu - xueli@stat.ufl.edu; Yunmei Chen - yun@math.ufl.edu; Rongling Wu* - rwu@stat.ufl.edu

* Corresponding author

Abstract

Background: Genes that control circadian rhythms in organisms have been recognized, but have been difficult to detect because circadian behavior comprises periodically dynamic traits and is sensitive to environmental changes.

Method: We present a statistical model for mapping and characterizing specific genes or quantitative trait loci (QTL) that affect variations in rhythmic responses. This model integrates a system of differential equations into the framework for functional mapping, allowing hypotheses about the interplay between genetic actions and periodic rhythms to be tested. A simulation approach based on sustained circadian oscillations of the clock proteins and their mRNAs has been designed to test the statistical properties of the model.

Conclusion: The model has significant implications for probing the molecular genetic mechanism of rhythmic oscillations through the detection of the clock QTL throughout the genome.

Background

Rhythmic phenomena are considered to involve a mechanism, ubiquitous among organisms populating the earth, for responding to daily and seasonal changes resulting from the planet's rotation and its orbit around the sun [1]. All these periodic responses are recorded in a circadian clock that allows the organism to anticipate rhythmic changes in the environment, thus equipping it with regulatory and adaptive machinery [2]. It is well recognized that circadian rhythms operate at all levels of biological organization, approximating a twenty-four hour period, or more accurately, the alternation between day and night [3]. Although there is a widely accepted view that the normal functions of biological processes are strongly correlated with the genes that control them, the detailed

genetic mechanisms by which circadian behavior is generated and mediated are poorly understood [4].

Several studies have identified various so-called circadian clock genes and clock-controlled transcription factors through mutants in animal models [5,6]. These genes have implications for clinical trials: their identification holds great promise for determining optimal times for drug administration based on an individual patient's genetic makeup. It has been suggested that drug administration at the appropriate body time can improve the outcome of pharmacotherapy by maximizing the potency and minimizing the toxicity of the drug [7], whereas drug administration at an inappropriate body time can induce more severe side effects [8]. In practice, body-time-dependent therapy, termed chronotherapy [9], can be

optimized via the genes that control expression of the patient's physiological variables during the course of a day.

With the completion of the Human Genome Project, it has been possible to draw a comprehensive picture of the genetic control of the functions of the biological clock and, ultimately, to integrate genetic information into routine clinical therapies for disease treatment and prevention. To achieve this goal, there is a pressing need to develop powerful statistical and computational algorithms for detecting genes or quantitative trait loci that determine circadian rhythms as complex dynamic traits. Unlike many other traits, rhythmic oscillations are generated by complex cellular feedback processes comprising a large number of variables. For this reason, mathematical models and numerical simulations are needed to grasp the molecular mechanisms and functions of biological rhythms fully [10]. These mathematical models have proved useful for investigating the dynamic bases of physiological disorders related to perturbations of biological behavior.

In this article, we will develop a statistical model for genetic mapping of QTL that determine patterns of rhythmic responses, using random samples from a natural population. This model is implemented by the principle of functional mapping [11], a statistical framework for mapping dynamic QTL for the pattern of developmental changes, by considering systems of differential equations for biological clocks. Simulation studies have been performed to investigate the statistical properties of the model.

Model
Mathematical Modeling of Circadian Rhythms
In all organisms studied so far, circadian rhythms that allow adaptation to a periodically changing environment originate from negative autoregulation of gene expression. Scheper et al. [10] illustrated and analyzed the generation of a circadian rhythm as a process involving a reaction cascade containing a loop, as depicted in Fig. 1A. The reaction loop consists in the production of the effective protein from its mRNA and negative feedback from the effective protein on mRNA production. The protein production process involves translation and subsequent processing steps such as phosphorylation, dimerization, transport and nuclear entry. It is assumed that the protein production cascade and the negative feedback are nonlinear processes in the reaction loop (Fig. 1B), with a time delay between protein production and subsequent processing. These nonlinearities and the delay critically determine the free-running periodicity in the feedback loop.

Scheper et al. [10] proposed a system of coupled differential equations to describe the circadian behavior of the intracellular oscillator:

$$\frac{dM}{dt} = \frac{r_M}{1+\left(\dfrac{P}{k}\right)^n} - q_M M$$

$$\frac{dP}{dt} = r_P M(t-\tau)^m - q_P P \qquad (1)$$

where M and P are, respectively, the relative concentrations of mRNA and the effective protein measured at a particular time, r_M is the scaled mRNA production rate constant, r_P is the protein production rate constant, q_M and q_P are, respectively, the mRNA and protein degradation rate constants, n is the Hill coefficient, m is the nonlinear exponent in the protein production cascade, τ is the total duration of protein production from mRNA, and k is a scaling constant.

Equation 1 constructs an unperturbed (free-running) system of the intracellular circadian rhythm generator that is defined by seven parameters, $\Theta_u = (n, m, \tau, r_M, r_P, q_M, q_P, k)$. The behavior of this system can be determined and predicted by changes in these parameter combinations. For a given QTL, differences in the parameter combinations among genotypes imply that this QTL is involved in the regulation of circadian rhythms. Statistical models will be developed to infer such genes from observed molecular markers such as single nucleotide polymorphisms (SNPs).

Statistical Modeling of Functional Mapping
Suppose a random sample of size N is drawn from a natural human population at Hardy-Weinberg equilibrium. In this sample, multiple SNP markers are genotyped, with the aim of identifying QTL that affect circadian rhythms. The relative concentrations of mRNA (M) and the effective protein (P) are measured in each subject at a series of time points (1, ..., T), during a daily light-dark cycle. Thus, there are two sets of serial measurements, expressed as [$M(1)$, ..., $M(T)$] and [$P(1)$, ..., $P(T)$]. According to the differential functions (1), these two variables, modeled in terms of their change rates, are expressed as differences between two adjacent times, symbolized by

$\mathbf{y} = [M(2) - M(1),..., M(T) - M(T - 1)]$

$= [\gamma(1),..., \gamma(T - 1)]$

for the protein change and

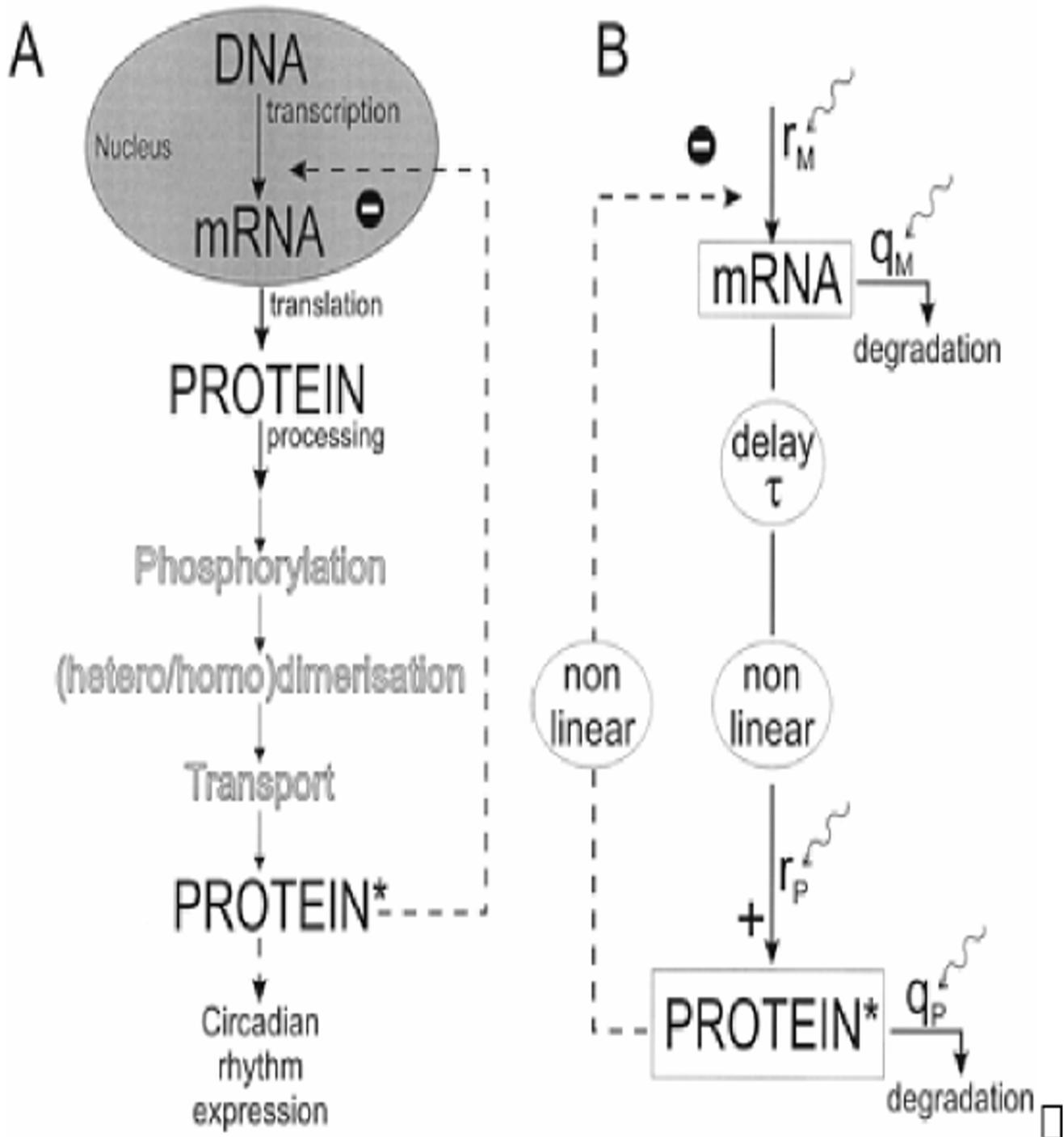$\mathbf{z} = [P(2) - P(1),..., P(T) - P(T - 1)]$

Figure 1
(A) Diagram of the biological elements of the protein synthesis cascade for a circadian rhythm generator. (B) Model interpretation of A showing the delay (τ) and nonlinearity in the protein production cascade, the nonlinear negative feedback, and mRNA and protein production (r_M, r_P) and degradation (q_M, q_P). Adapted from ref. [10].

$$= [z(1),..., z(T - 1)]$$

for the mRNA change.

Assume that a putative QTL with alleles *A* and *a* affecting circadian rhythms is segregated in the population. The frequencies of alleles *A* and *a* are *q* and 1 - *q*, respectively. For a particular genotype *j* of this QTL *(j* = 0 for *aa*, 1 for *Aa* and 2 for *AA*), the parameters describing circadian rhythms are denoted by $\Theta_{uj} = (n_j, m_j, \tau_j, r_{Mj}, r_{Pj}, q_{Mj}, q_{Pj}, k_j)$. Comparisons of these quantitative genetic parameters among the three different genotypes can determine whether and how this putative QTL affects circadian rhythms.

The time-dependent phenotypic changes in mRNA and protein traits for individual *i* measured at time *t* due to the QTL can be expressed by a bivariate linear statistical model

$$y_i(t) = \sum_{j=0}^{2} \xi_{ij} u_{Mj}(t) + e_i^y(t)$$
$$z_i(t) = \sum_{j=0}^{2} \xi_{ij} u_{Pj}(t) + e_i^z(t) \qquad (2)$$

where ξ_{ij} is an indicator variable for the possible genotypes of the QTL for individual *i*, defined as 1 if a particular QTL genotype *j* is indicated and 0 otherwise, $u_{Mj}(t)$ and $u_{Pj}(t)$ are the genotypic values of the QTL for mRNA and protein changes at time *t*, respectively, which can be determined using the differential functions expressed in equation (1), and $e_i^y(t)$ and $e_i^z(t)$ are the residual effects in individual *i* at time *t*, including the aggregate effect of polygenes and error effects.

The dynamic features of the residual errors of these two traits can be described by the antedependence model, originally proposed by Gabriel [12] and now used to model the structure of a covariance matrix [13]. This model states that an observation at a particular time *t* depends on the previous ones, the degree of dependence decaying with time lag. Assuming the 1st-order structured antedependence (SAD(1)) model, the relationship between the residual errors of the two traits *y* and *z* at time *t* for individual *i* can be modeled by

$$e_i^y(t) = \phi_z e_i^y(t - 1) + \psi_z e_i^z(t - 1) + \varepsilon_i^y(t)$$
$$e_i^z(t) = \phi_z e_i^z(t - 1) + \psi_z e_i^y(t - 1) + \varepsilon_i^z(t) \qquad (3)$$

where ϕ_k and ψ_k are, respectively, the antedependence parameters caused by trait *k* itself and by the other trait, and $\varepsilon_i^y(t)$ and $\varepsilon_i^z(t)$ are the time-dependent innovation error terms, assumed to be bivariate normally distributed with mean zero and variance matrix

$$\Sigma_\varepsilon(t) = \begin{pmatrix} \delta_x^2(t) & \delta_x(t)\delta_y(t)\rho(t) \\ \delta_x(t)\delta_y(t)\rho(t) & \delta_y^2(t) \end{pmatrix},$$

where $\delta_x^2(t)$ and $\delta_y^2(t)$ are termed time-dependent innovation variances. These variances can be described by a parametric function such as a polynomial of time [14], but are assumed to be constant in this study. $\rho(t)$ is the correlation between the error terms of the two traits, specified by an exponential function of time *t* [14], but is assumed to be time-invariant for this study. It is reasonable to say that there is no correlation between the error terms of two traits at different time points, i.e. $Corr(\varepsilon_i^y(t_y), \varepsilon_i^z(t_z)) = 0$ $(t_y \neq t_z)$.

Based on the above conditions, the covariance matrix (Σ) of phenotypic values for traits *y* and *z* can be structured in terms of ϕ_y, ϕ_z, ψ_y, ψ_z and $\Sigma_\varepsilon(t)$ by a bivariate SAD(1) model [15,16]. Also, the closed forms for the determinant and inverse of Σ can be derived as given in [15,16]. We use a vector of parameters arrayed in $\Theta_v = (\phi_y, \phi_z, \psi_y, \psi_z, \delta_y, \delta_z, \rho)$ to model the structure of the covariance matrix involved in the function mapping model.

Likelihood
The likelihood of samples with 2(*T* 1)-dimensional measurements, $\mathbf{x} = (\mathbf{x}_i) = \{y_i(t), z_i(t)\}_{t=1}^{T}$, for individual *i* and marker information, M, in the human population affected by the QTL is formulated on the basis of the mixture model, expressed as

$$L(\omega, \Theta \mid \mathbf{x}, M) = \prod_{i=1}^{N} \left[\sum_{i=1}^{2} \omega_{j|i} f_j(\mathbf{x}_i; \Theta_{uj}, \Theta_v) \right] \qquad (4)$$

where the unknown parameters include two parts, $\omega = (\omega_{j|i})$ and $\Theta = (\Theta_{uj}, \Theta_v)$. In the statistics, the parameters $\omega = (\omega_{j|i})$ determine the proportions of different mixture normals, and actually reflect the segregation of the QTL in the population, which can be inferred on the basis of non-random association between the QTL and the markers. For a mapping population, *N* progeny can be classified into different groups on the basis of known marker genotypes. Thus, in each such marker genotype group, the mixture proportions of QTL genotypes ($\omega_{j|i}$) can be expressed

as the conditional probability of QTL genotype j for subject i given its marker genotype.

Suppose that this QTL is genetically associated with a codominant SNP marker that has three genotypes, *MM*, *Mm* and *mm*. Let p and $1 - p$ be the allele frequencies of marker alleles *M* and *m*, respectively, and D be the coefficient of (gametic) linkage disequilibrium between the marker and QTL. According to linkage disequilibrium-based mapping theory [17], the detection of significant linkage disequilibrium between the marker and QTL implies that the QTL may be linked with and, therefore, can be genetically manipulated by the marker. The four haplotypes for the marker and QTL are *MA, Ma, mA* and *ma*, with respective frequencies expressed as $p_{11} = pq + D$, $p_{10} = p(1 - q) - D$, $p_{01} = (1 - p)q - D$ and $p_{00} = (1 - p)(1 - q) + D$. Thus, the population genetic parameters p, q, D can be estimated by solving a group of regular equations if we can estimate the four haplotype frequencies. The conditional probabilities of QTL genotypes given marker genotypes in a natural population can be expressed in terms of the haplotype frequencies (see [18]).

In the mixture model (4), $\Theta = \{(\Theta_{uj}, \Theta_v)\}_{j=0}^2$ is the unknown vector that determines the parametric family f_j, described by a multivariate normal distribution with the genotype-specific mean vector

$$\mathbf{u}_j = \left(\mathbf{u}_{Mj}; \mathbf{u}_{Pj}\right) = \{u_{Mj}(t); u_{Pj}(t)\}_{t=1}^{T-1}$$

$$= \left[\frac{r_{Mj}}{1 + \left(\frac{P(t+1)}{k_j}\right)^{n_j}} - q_{Mj}M(t+1); r_{Pj}M(t+1-\tau_j)^{m_j} - q_{Pj}P(t+1) \right]_{t=1}^{T-1} \quad (5)$$

and the covariance matrix Σ. While the mean vector is determinedby genotype-specific parameters, $\Theta_{uj} = (n_j, m_j, \tau_j, r_{Mj}, r_{Pj}, q_{Mj}, q_{Pj}, k_j)$, $j = (2,1,0)$ the covariance matrix is structured by common parameters, $\Theta_v = (\phi_y, \phi_z, \psi_y, \psi_z, \delta_y, \delta_z, \rho)$.

Algorithm
Wang and Wu [18] proposed a closed form for the EM algorithm to obtain the maximum likelihood estimates (MLEs) of haplotype frequencies p_{11}, p_{10}, p_{01} and p_{00}, and thus the allele frequencies of the marker (p) and QTL (q) and their linkage disequilibrium (D). Genotype-specific mathematical parameters in $\mathbf{u}_j$ (5) for the two differential functions of circadian rhythms, and the parameters that specify the structure of the covariance matrix, Σ, can be theoretically estimated by implementing the EM algorithm. But it would be difficult to derive the log-likelihood equations for these parameters by this approach because they are related in a complicated nonlinear way. The simplex algorithm, which relies only upon a target

function, has proved powerful for estimating the MLEs of these parameters [19] and will be used in this study. As discussed above, closed forms exist for the determinant and inverse and should be incorporated into the estimation process to increase computational efficiency.

Hypothesis Testing
One of the most significant advantages of functional mapping is that it can ask and address biologically meaningful questions about the interplay between gene actions and trait dynamics by formulating a series of hypothesis tests. Wu et al. [20] described several general hypothesis tests for different purposes. Although all these general tests can be used directly in this study, we propose here the most important and specific tests for the existence of QTL that affect mRNA and protein changes pleiotropically or separately, and for the effects of the QTL on the shape of differential functions.

Existence of QTL
Testing whether a specific QTL is associated with the differential functions (1) is a first step toward understanding the genetic architecture of circadian rhythms. The genetic control of the entire rhythmic process can be tested by formulating the following hypotheses:

$$H_0 : D = 0 \; vs. \; H_1 : D \neq 0 \quad (6)$$

H_0 states that there are no QTL affecting circadian rhythms (the reduced model), whereas H_1 proposes that such QTL do exist (the full model). The statistic for testing these hypotheses (6) is calculated as the log-likelihood ratio (LR) of the reduced to the full model:

$$LR_1 = -2[\ln L(\tilde{\Theta}, \tilde{\omega} \,|\mathbf{x}, M) - \ln L(\hat{\Theta}, \hat{\omega} \,|\mathbf{x}, M)], \quad (7)$$

where the tildes and hats denote the MLEs of the unknown parameters under H_0 and H_1, respectively. The LR is asymptotically χ^2-distributed with one degree of freedom.

A similar test for the existence of a QTL can be performed on the basis of these hypotheses, as follows:

$$H_0 : \Theta_{uj} \equiv \Theta_u, \, j = (2,1,0) \quad (8)$$

H_1 : At least one of the equalities above does not hold;

from which the LR is calculated by

$$LR_2 : -2[\ln L(\tilde{\tilde{\Theta}} \,|\mathbf{x}) - \ln(\hat{\Theta}, \hat{\omega} \,|\mathbf{x}, M)], \quad (9)$$

with the doubled tildes denoting the estimates under H_0 of hypothesis (8). It is difficult to determine the distribu-

tion of the LR_2 because the linkage disequilibrium is not identifiable under H_1. An empirical approach to determining the critical threshold is based on permutation tests, as advocated by Churchill and Doerge [21]. By repeatedly shuffling the relationships between marker genotypes and phenotypes, a series of maximum LR_2 values are calculated, from the distribution of which the critical threshold is determined.

Is the QTL for mRNA or protein rhythms?

After the existence of a QTL that affects circadian rhythms is confirmed, we need to test whether it affects the rhythmic responses of mRNA and protein jointly or separately. The hypothesis for testing the effect of the QTL on the mRNA response is formulated as

$$H_0 : (r_{Mj}, q_{Mj}, k_j, n_j) \equiv (r_M, q_M, k, n) \text{ for } j = 0, 1, 2 \quad (10)$$

H_1 : At least one of the equalities above does not hold.

The log-likelihood values under H_0 and H_1 are calculated, and thus the corresponding LR.

A similar test is formulated for detecting the effect of the QTL on the protein rhythm:

$$H_0 : (r_{Pj}, q_{Pj}, \tau_j, m_j) \equiv (r_P, q_P, \tau, m) \text{ for } j = 0, 1, 2 \quad (11)$$

H_1 : At least one of the equalities above does not hold.

For both hypotheses (10) and (11), an empirical approach to determining the critical threshold is based on simulation studies. If the null hypotheses of (10) and (11) are both rejected, this means that the QTL exerts a pleiotropic effect on the circadian rhythms of mRNA and protein.

The QTL responsible for the behavior and shape of circadian rhythms

Two different subspaces of parameters are used to define the features of circadian rhythms: $\{n, m, \tau\}$, determining the nonlinearity and delay in the system, and $\{r_M, r_P, q_M, q_P\}$, determining the phase-response curves. The null hypotheses regarding the genetic control of the system's oscillatory behavior and the shape of the rhythmic responses are:

$$\begin{aligned} H_0 &: (n_j, m_j, \tau_j) \equiv (n, m, \tau) \\ H_0 &: (r_{M_j}, r_{P_j}, q_{M_j}, q_{P_j}) \equiv (r_M, r_P, q_M, q_P) \end{aligned} \text{ for } j = 0, 1, 2 \quad (12)$$

The oscillatory behavior of a circadian rhythm can also be determined by the amplitude of the rhythm, defined as the difference between the peak and trough values; its phase, defined as the timing of a reference point in the cycle (e.g. the peak) relative to a fixed event (e.g. begin-

ning of the night phase); and its period, defined as the time interval between phase reference points (e.g. two peaks). The genetic determination of all thesevariables can be tested.

Simulation

Simulation experiments are performed to examine the statistical properties of the model proposed for genetic mapping of circadian rhythms. We choose 200 individuals at random from a human population at Hardy-Weinberg equilibrium. Consider one of the markers genotyped for all subjects. This marker, with two alleles M and m, is used to infer a QTL with two alleles A and a for circadian rhythms on the basis of non-random association. The allele frequencies are assumed to be $p = 0.6$ for allele M and $q = 0.6$ for allele A. A positive value of linkage disequilibrium $(D = 0.08)$ between M and A is assumed, suggesting that these two more common alleles are in coupled phase [22].

The three QTL genotypes, AA, Aa and aa, are each hypothesized to have different response curves for circadian rhythms of mRNA and protein as described by equation (1). The rhythmic parameters $\Theta_{uj} = (n_j, m_j, \tau_j, r_{Mj}, r_{Pj}, q_{Mj}, q_{Pj}, k_j)$ for the three genotypes, given in Table 1, are determined in the ranges of empirical estimates of these parameters [10]. Note that for computational simplicity the scaling constant k and the total duration of protein production from mRNA are given values 1 and 4.0, respectively. We used the SAD(1) model to structure the covariance matrix based on the antedependence parameters (ϕ_x, ϕ_y, ψ_x, ψ_y) and innovation variances (δ_x^2, δ_y^2) (Table 1). The innovation variances for each of the two rhythmic traits were determined by adjusting the heritability of the curves to $H^2 = 0.1$ and 0.4, respectively, due to the QTL for the rhythmic response at a middle measurement point.

Many factors have been shown to affect the precision of parameter estimation and the power of QTL detection by functional mapping. These factors are related to experimental design (sample size and number and pattern of repeated measures), the genetic properties of the circadian rhythm (heritability of the curves, population genetic parameters of the underlying QTL), and the analytical approach to modeling the structure of the covariance matrix. Previous studies have investigated the properties of functional mapping when different experimental designs are used [15,18]. For this simulation study, we focus on the influence of different heritabilities on parameter estimation using a practically reasonable sample size ($n = 200$). We assumed that the relative concentrations of

Table 1: The MLEs of parameters that define circadian rhythms for three different QTL genotypes, the structure of the covariance matrix and the association between the marker and QTL in a natural population, taking the heritability of the assumed QTL as $H^2 = 0.1$. The numbers in parentheses are the square roots of the mean square errors of the MLEs.

	Rhythmic Parameters					
	AA		*Aa*		*aa*	
	Given	MLE	Given	MLE	Given	MLE
n	1.50	1.57(0.067)	1.40	1.54(0.139)	1.70	1.74(0.045)
m	3.00	3.02(0.018)	2.90	2.91(0.010)	2.80	2.79(0.050)
r_M	1.10	1.16(0.060)	1.30	1.35(0.054)	0.90	0.98(0.083)
r_P	0.30	0.30(0.020)	0.35	0.36(0.007)	0.40	0.39(0.001)
q_M	0.16	0.15(0.008)	0.17	0.16(0.011)	0.18	0.17(0.012)
q_P	0.16	0.16(0.001)	0.17	0.16(0.005)	0.18	0.18(0.001)

	Matrix Structuring Parameters	
	Given	MLE
ϕ_x	0.010	0.011(0.001)
ϕ_y	-0.100	0.098(0.001)
ψ_x	0.100	0.105(0.005)
ψ_y	-0.200	-0.206(0.006)
δ_x^2	0.223	0.223(0.001)
δ_y^2	1.842	1.742(0.100)
ρ	0.200	0.216(0.016)

	Genetic Parameters	
	Given	MLE
p	0.6	0.601(0.003)
q	0.6	0.501(0.094)
D	0.08	0.068(0.012)

mRNA and protein are measured at eight equally-spaced time points in each subject, although these measurements can be made differently in terms of the number and pattern of repeated measures.

The phenotypic values of circadian rhythms for the mRNA and protein traits are simulated by summing the genotypic values predicted by the rhythmic curves and residual errors following a multivariate normal distribution, with $MVN(0, \Sigma)$. The simulated phenotypic and marker data were analyzed by the proposed model. The population genetic parameters of the QTL can be estimated with reasonably high precision using a closed-form solution approach [18]. We compare the estimation of the marker allele frequencies, QTL allele frequencies and marker-QTL linkage disequilibria under different heritability levels.

The precision of estimation of marker allele frequency is not affected by differences in heritability, but estimates of QTL allele frequency and marker-QTL linkage disequilibrium are more precise for a higher (Table 1) than a lower (Table 2) heritability.

Figure 2A illustrates different forms of circadian rhythms for three QTL genotypes, *AA, Aa* and *aa*, with the rhythmic values for the protein and mRNA responses given in Tables 1 and 2. Pronounced differences among the genotypes imply that the QTL may affect the joint rhythmic response of the protein and mRNA concentrations. The rhythmic values can be estimated reasonably from the model. Using the estimates of the rhythmic parameters from one random simulation, we draw the oscillations of the two traits. The shapes of these curves seem to be

Table 2: The MLEs of parameters that define circadian rhythms for three different QTL genotypes, the structure of the covariance matrix and the association between the marker and QTL in a natural population, taking the heritability of the assumed QTL as $H^2 = 0.4$. The numbers in parentheses are the square roots of the mean square errors of the MLEs.

	Rhythmic Parameters					
	AA		*Aa*		*aa*	
	Given	MLE	Given	MLE	Given	MLE
n	1.50	1.52(0.059)	1.40	1.42(0.006)	1.70	1.71(0.060)
m	3.00	3.02(0.015)	2.90	2.90(0.009)	2.80	2.80(0.020)
r_M	1.10	1.14(0.052)	1.30	1.34(0.028)	0.90	0.93(0.066)
r_P	0.30	0.30(0.009)	0.35	0.35(0.002)	0.40	0.40(0.001)
q_M	0.16	0.16(0.003)	0.17	0.17(0.011)	0.18	0.18(0.005)
q_P	0.16	0.16(0.004)	0.17	0.17(0.003)	0.18	0.18(0.001)

	Matrix Structuring Parameters					
	Given	MLE				
ϕ_x	0.010	0.010(0.001)				
ϕ_y	-0.100	0.095(0.002)				
ψ_x	0.100	0.102(0.005)				
ψ_y	-0.200	0.201(0.006)				
δ_x^2	0.307	0.309(0.011)				
δ_y^2	0.200	0.204(0.011)				
ρ	0.037	0.038(0.002)				

	Genetic Parameters					
	Given	MLE				
p	0.6	0.601(0.002)				
q	0.6	0.67(0.091)				
D	0.08	0.067(0.022)				

broadly consistent with those of the hypothesized curves, although the curve estimates are more accurate under higher (Fig. 2C) than lower (Fig. 2B) heritability.

The estimates of the rhythmic parameters for each response curve also display reasonable precision, as assessed by the square roots of the mean square errors over 100 repeated simulations. As expected, the estimate is more precise when the heritability increases from 0.1 (Table 1) to 0.4 (Table 2). The model displays great power in detecting a QTL responsible for circadian rhythms using the marker associated with it. Given the above simulation conditions, a significant QTL can be detected with about 75% power for a heritability of 0.1. The power increases to over 90% as the heritability increases to 0.4.

The model can be used to test whether the QTL detected for overall protein and mRNA rhythm responses also affects key features of circadian rhythms, such as period, amplitude or phase shift, by formulating the corresponding hypotheses. For a real data set, it is exciting to test these hypotheses because they may enable the mechanistic basis of the genetic regulation of circadian rhythms to be identified. In the current simulation, these hypothesis tests were not performed.

Discussion
One of the most important aspects of life is the rhythmic behavior that is rooted in the many regulatory mechanisms that control the dynamics of living systems. The most common biological rhythms are circadian rhythms,
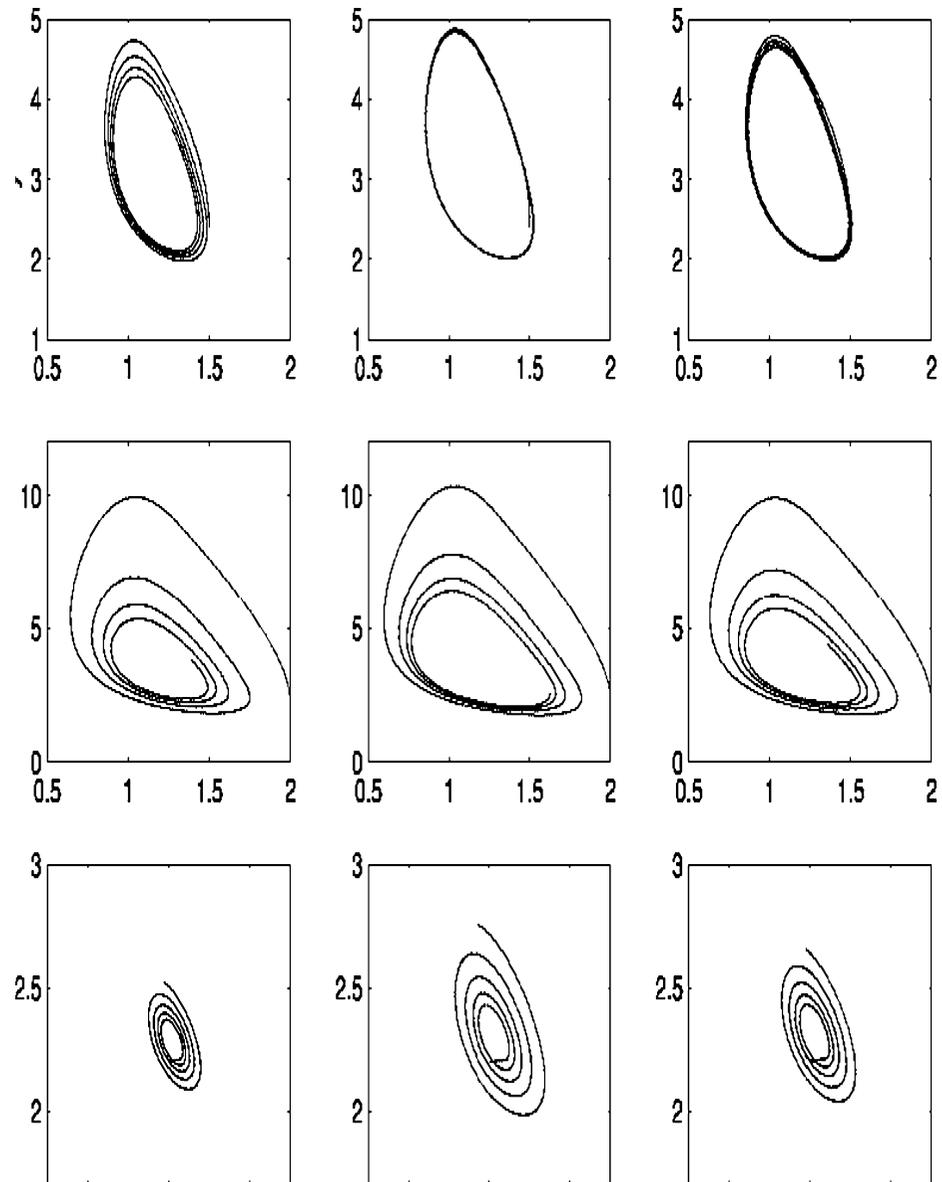
Figure 2
Free-running oscillation of mRNA abundance (*x*) and protein abundance (*y*) in a rhythmic system, expressed as limit cycle contour, annotated with the time points within the 24.6 h circadian cycle, for three assumed QTL genotypes using given rhythmic parameter values (A), estimated values under $H^2 = 0.1$ (B), and estimated values under $H^2 = 0.4$ (C). The three plots within each column correspond to QTL genotypes *AA*, *Aa* and *aa*, respectively.

which occur with a period close to 24 h, allowing organisms to adapt to periodic changes in the terrestrial environment [1]. With the rapid accumulation of new data on gene, protein and cellular networks, it is becoming increasingly clear that genes are heavily involved in the cellular regulatory interactions underpinning circadian rhythms [4,23]. However, a detailed picture of the genetic architecture of circadian rhythms has not been obtained, although ongoing projects such as the Human Genome Project will assist in the characterization of circadian genetics.

Traditional strategies for identifying circadian clock genes in mammals have been based on the analysis of single gene mutations and the characterization of genes identified by cross-species homology, and have laid an essential groundwork for circadian genetics [6,23]. However, these strategies do not include a more thorough examination of the breadth and complexity of influences on circadian behavior throughout the entire genome. Genetic mapping relying upon genetic linkage maps has provided a powerful tool for identifying the quantitative trait loci (QTL) responsible for circadian rhythms. In a mapping study of 196 F_2 hybrid mice, Shimomura et al. [24] detected 14 interacting QTL that contribute to the variation of rhythmic behavior in mice by analyzing different discrete aspects of circadian behavior: free-running circadian period, phase angle of entrainment, amplitude of the circadian rhythm, circadian activity level, and dissociation of rhythmicity.

The data of Shimomura et al. [24] point to promising approaches for genome-wide analysis of rhythmic phenotypes in mammals including humans. Their most significant drawback is the lack of robust statistical inferences about the dynamic genetic control of circadian rhythms. Typically, biological rhythms are dynamic traits, and the pattern of their genetic determination can change dramatically with time. In this article, we have incorporated mathematical models and concepts regarding the molecular and cellular mechanisms of circadian rhythms into a general framework for mapping dynamic traits, called functional mapping [11]. Based firmly on experiments, robust differential equations have been established to provide an essential tool for studying and comprehending the cellular networks for circadian rhythms [1,25-27]. As an attempt to integrate differential equations into functional mapping, the statistical model shows favorable properties in estimating the effects of a putative QTL and its association with polymorphic markers. The simulation study results suggest that the parameters determining the behavior and shape of circadian rhythmic curves can be estimated reasonably even if the QTL effect is small to moderate. As seen in general functional mapping [11], the model implemented with a system of differential equa-

tions also allows us to make a number of biologically meaningful hypothesis tests for understanding the genetic control of rhythmic responses in organisms.

As a first attempt of its kind, the model proposed in this article has only considered one QTL associated with circadian rhythms. A one-QTL model is definitely not sufficient to explain the complexity of the genetic control of this trait. A model incorporating multiple QTL and their interactive networks should be derived; this is technically straightforward. In addition, the system of circadian rhythms is characterized by two variables, and this may also be too simple to reflect the complexity of rhythmic behavior. A number of more sophisticated models, governed by systems of five [28], ten [29] or 16 kinetic equations [4,30,31], have been constructed to describe the detailed features of a rhythmic system in regard to responses to various internal and environmental factors. While the identification of circadian clock genes can elucidate the molecular mechanism of the clock, our model will certainly prove its value in elucidating the genetic architecture of circadian rhythms and will probably lead to the detection of the driving forces behind circadian genetics and its relationship to the organism as a whole.

Competing interests
The author(s) declare that they have no competing interests.

Authors' contributions
The entire theoretical concept of the work was envisaged by RLW. The mathematical and statistical modeling was carried out by TL with feedback from XLL and YMC.

Acknowledgements

References
1. Goldbeter A: **Computational approaches to cellular rhythms.** *Nature* 2002, **420**:238-245.
2. Webb AAR: **The physiology of circadian rhythms in plants.** *New Phytologist* 2003, **160**:281-303.
3. Goldbeter A: *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour* Cambridge: Cambridge University Press; 1996.
4. Ueda HR, Hagiwara M, Kitano H: **Robust oscillations within the interlocked feedback model of Drosophila circadian rhythm.** *J Theor Biol* 2001, **210**:401-406.
5. Takahashi JS: **Gene regulation. Circadian clocks a la CREM.** *Nature* 1993, **365**:299-300.
6. Reppert SM, Weaver DR: **Coordination of circadian timing in mammals.** *Nature* 2002, **418**:935-941.
7. Levi F, Zidani R, Misset JL: **Randomised multicentre trial of chronotherapy with oxaliplatin, uorouracil, and folinic acid in metastatic colorectal cancer.** *International Organization for Cancer Chronotherapy. Lancet* 1997, **350**:681-686.
8. Ohdo S, Koyanagi S, Suyama H, Higuchi S, Aramaki H: **Changing the dosing schedule minimizes the disruptive effects of interferon on clock function.** *Nature Med* 2001, **7**:356-360.

9.　Labrecque G, Belanger PM: **Biological rhythms in the absorption, distribution, metabolism and excretion of drugs.** *Pharmacol Ther* 1991, **52**:95-107.
10.　Scheper TO, Klinkenberg D, Pennartz C, van Pelt J: **A mathematical model for the intracellular circadian rhythm generator.** *J Neuroscience* 1999, **19**:40-47.
11.　Wu RL, Lin M: **Functional mapping How to study the genetic architecture of dynamic complex traits.** *Nature Rev Genet* 2006, **7**:229-237.
12.　Gabriel KR: **Ante-dependence analysis of an ordered set of variables.** *Ann Math Stat* 1962, **33**:201-212.
13.　Nunez-Anton V, Zimmerman DL: **Modeling nonstationary longitudinal data.** *Biometrics* 2000, **56**:699-705.
14.　Pourahmadi M: **Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation.** *Biometrika* 1999, **86**:677-690.
15.　Lin M, Wu RL: **Theoretical basis for the identification of allelic variants that encode drug ecacy and toxicity.** *Genetics* 2005, **170**:919-928.
16.　Zhao W, Hou W, Littell RC, Wu RL: **Structured antedependence models for functional mapping of multivariate longitudinal quantitative traits.** *Statistical Methods in Molecular Genetics and Biology 4* 2005 [http://www.bepress.com/sagmb/vol4/iss1/art33/].
17.　Wu RL, Ma C-X, Casella G: **Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations.** *Genetics* 2002, **160**:779-792.
18.　Wang ZH, Wu RL: **A statistical model for high-resolution mapping of quantitative trait loci determining human HIV-1 dynamics.** *Stat Med* 2004, **23**:3033-3051.
19.　Zhao W, Wu RL, Ma C-X, Casella G: **A fast algorithm for functional mapping of complex traits.** *Genetics* 2004, **167**:2133-2137.
20.　Wu RL, Ma C-X, Lin M, Casella G: **A general framework for analyzing the genetic architecture of developmental characteristics.** *Genetics* 2004, **166**:1541-1551.
21.　Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** *Genetics* 1994, **138**:963-971.
22.　Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits* Sunderland, MA: Sinauer; 1998.
23.　Tafti M, Petit B, Chollet D, Neidhart E, de Bilbao F, Kiss JZ, Wood PA, Franken P: **Deficiency in short-chain fatty acid beta-oxidation affect theta oscillations during sleep.** *Nature Genet* 2003, **34**:320-325.
24.　Shimomura K, Low-Zeddies SS, King DP, Steeves TD, Whiteley A, Kushla J, Zemenides PD, Lin A, Vitaterna MH, Churchill GA, Takahashi JS: **Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice.** *Genome Res* 2001, **11**:959-980.
25.　Leloup JC, Gonze D, Goldbeter A: **Limit cycle models for circadian rhythms based on transcriptional regulation in *Drosophila* and *Neurospora*.** *J Biol Rhythms* 1999, **14**:433-448.
26.　Gonze D, Halloy J, Goldbeter A: **Stochastic versus deterministic models for circadian rhythms.** *J Biol Physics* 2002, **28**:637-653.
27.　Gonze D, Halloy J, Leloup JC, Goldbeter A: **Stochastic models for circadian rhythms: inuence of molecular noise on periodic and chaotic behavior.** *C R Biologies* 2003, **326**:189-203.
28.　Goldbeter A: **A model for circadian oscillations in the *Drosophila* period protein (PER).** *Proc R Soc Lond B* 1995, **261**:319-324.
29.　Leloup JC, Goldbeter A: **A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins.** *J Biol Rhythms* 1998, **13**:70-87.
30.　Smolen P, Baxter DA, Byrne JH: **Modeling circadian oscillations with interlocking positive and negative feedback loops.** *J Neurosci* 2001, **21**:6644-6656.
31.　Leloup JC, Goldbeter A: **Toward a detailed computational model for the mammalian circadian clock.** *Proc Natl Acad Sci USA* 2003, **100**:7051-7056.