Statistical mechanics of clock gene networks underlying circadian rhythms

Lidan Sun[1,2], Ang Dong[1,2], Christopher Griffin[3], and Rongling Wu[4]

[1]Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China

[2]Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

[3]Applied Research Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

[4]Center for Statistical Genetics, Departments of Public Health Sciences and Statistics, The Pennsylvania State University, Hershey, PA 17033, USA

Corresponding author: Rongling Wu, rwu@phs.psu.edu

Abstract

All multicellular organisms embed endogenous circadian oscillators or clocks that rhythmically regulate a wide variety of processes in response to daily environmental cycles. Previous molecular studies using rhythmic mutants for several model systems have identified a set of genes responsible for rhythmic activities and illustrated the molecular mechanisms underlying how disruptions in circadian rhythms are associated with the sort of aberrant cell cycling. However, the wide use of these forward genetic studies is impaired by a limited number of mutations that can be identified or induced only in a single genome, limiting the identification of many other conserved or non-conserved clock genes. Genetic linkage or association mapping provides an unprecedented glimpse into the genome-wide scanning and characterization of genes underlying circadian rhythms. The implementation of sophisticated statistical models into mapping studies can not only identify key clock genes or clock quantitative trait loci (cQTL) but also, more importantly, reveal a complete picture of the genetic control mechanisms constituted by gene interactomes. Here, we introduce and review an advanced statistical mechanics framework for coalescing all possible clock genes into intricate but well-organizing interaction networks that regulate rhythmic cycles. The application of this framework to widely available mapping populations will reshape and further our understanding of the genetic signatures behind circadian rhythms for an enlarged range of species including microbes, plants, and humans.

Keywords: circadian clock, quantitative trait loci, genetic mapping, genetic network, statistical modeling

1. Introduction

The 24-h rotation of the Earth causes predictable changes in light and temperature in our natural environment. Accordingly, all living organisms from microorganisms to insects, plants, and mammals exhibit circadian rhythms, i.e., sustained oscillations with a period close to 24 h [1]. Circadian rhythms are mediated by an internal body clock, which appears nearly ubiquitous in life, and regulates a wide array of metabolic and physiological functions, such as hormone

production, cell regeneration, brain wave activity and organism behavior [2-5]. Disruptions in biological rhythms can be associated with aberrant cell cycling, ultimately leading to disease such as tumorigenesis, cardiovascular disease, and neurodegenerative disorders [6-8] and reduced productivity in plants [9-12].

The biological process of circadian rhythms involves three fundamental components: Input pathways that transmit environmental cues to the circadian clock, the clock gene itself, which generates the biological rhythm, and output pathways that entrain the clock's information regarding phase and periodicity to the rest of the organism [13]. Extensive molecular studies have successfully identified specific clock genes that regulate an organism's cyclic response to its surrounding environment [14-16]. The first clock gene, *per*, was characterized and cloned in *Drosophila* [17-19], which was subsequently found to regulate circadian rhythms through its protein product PER [20-23]. However, the question of how the PER protein enters the nucleus to act as a transcriptional factor was not answered until the second clock gene *tim* was discovered [24,25]. Takahashi and his group found that the transcription of *tim* and *per* were crucial for sustaining an autonomous oscillation that is activated by a positive input, *clock*, the first clock gene detected in mammals [26-28], that functions through the CLOCK-BMAL1 heterodimeric transcription factor [29]. Tremendous efforts have been made to understand the molecular basis of how the clock genes receive input signals, drive their entrainment, and regulate cellular aspects of circadian rhythms [1,8].

With the continuous improvement of molecular and cloning techniques, an increasing number of clock genes have been detected and characterized [1,30-34]. These genes were revealed to encode proteins through multiple interconnected transcriptional and translational feedback loops, having various impacts on physiological and behavioral rhythmicity [1,16,34-38]. Despite this progress, a complete characterization of clock genes and their rhythmic functions is still far from clear. First, the current identification of molecular components for circadian clocks is mostly based on forward genetics approaches that utilize mutants with abnormal behavioral cycles to map genes [39]. In practice, only a limited number of rhythm mutations can be detected or induced in a single genome for several well-studied model systems, such as cyanobacteria, *Neurospora crassa*, *Arabidopsis thaliana*, *Drosophila melanogaster* and *Mus musculus* (mice).

[1,40,41]. It is largely unknown how clock genes occur and function in those organisms for which mutants are hardly available. Second, although circadian rhythms are omnipresent, their underlying molecular mechanisms are not conserved among evolutionarily divergent organisms [42-45], which makes it difficult or even impossible to chart a complete picture of clock machineries from only model systems.

Linkage or association mapping is a forward genetic tool that can serve as an alternative approach for clock gene detection [46]. This approach, not relying on rhythmic mutants, can take advantage of increasingly available genotypic and sequencing data collected at unprecedented resolution for almost all species and make a full use of considerable allelic variation in clock function that has been accumulated during evolution in natural populations [47,48]. It displays a formidable ability to map a complete set of quantitative trait loci (QTLs) throughout the entire genome that control a rhythmic trait. For example, using linkage mapping or association studies, important clock QTLs (cQTLs) that mediate rhythmic activities have been mapped in several species [49-55]. These studies perform association analysis between marker genotypes and chronophenotypes to identify and map significant genetic loci. To leverage the biological relevance of cQTL detection, several dynamic mapping approaches have been proposed to characterize how QTLs globally regulate the periodic pattern and form of circadian rhythms expressed in various stages from gene expression to protein turnover to metabolic rhythm and ultimately to cell cycles [56-63]. These dynamic approaches, referred to collectively as functional mapping or systems mapping (reviewed in [64-66]), integrate mathematical aspects of circadian rhythms into a mapping setting, and provide a capacity to test the temporal trajectories of genetic effects, exerted by cQTLs, on rhythmic patterns.

Existing mapping models were developed to detect individual significant QTLs from a large pool of genome-wide molecular markers. These models work well in specific situations, but may not work for rhythmic mapping because circadian clocks involve a number of heterogeneous genes that act singly and work together via local or non-local interactions. The past three decades have seen the tremendous development of statistical models for reconstructing interaction networks from gene expression data (see a number of excellent reviews [67-70] from different perspectives). Reconstruction of genetic networks at the QTL level is much more challenging,

although a genome-wide QTL-QTL interaction network can provide direct insight into the genetic control of complex biological processes. More recently, Jiang et al. [71] have proposed a statistical model for mapping QTL networks underlying developmental trajectories by integrating functional mapping and evolutionary game theory. We argue that this model can be modified to map new cQTLs for rhythmic processes and unveil how these cQTLs interact with each other through an intrinsic but well-orchestrated network. Here, by reviewing the fundamental utility of this model to study clock genetics, we augment it into a generic paradigm in which genome-wide interactome networks can be inferred at any dimension. We also integrate QTL control networks and gene regulatory networks to establish an intertwined bidirectional, signed, and weight circuit that can better reveal key organizing principles of how biological rhythms are regulated through a web of interacting cQTLs and transcriptomic networks.

2. A general framework for systems evolutionary game networks

Notation: Different from traditional forward genetics, reverse genetics based on mapping or association studies can simultaneously detect and map a wide array of new genes regulating rhythmic pattern and function for any species, without relying on the characterization of rhythmic mutants. Also, unlike the notion of clock genes limited to transcriptional genes and their products in forward genetics, clock mapping studies based on reverse genetic thinking are concerned with gene detection that covers an entire domain of the central dogma of biology ranging from DNA to RNA to proteins to cellular physiology and behavior. To clarify some issues, we provide several relevant notations. In a mapping population, we genotype DNA markers, e.g., single nucleotide polymorphisms (SNPs), measure the profiles of genes and their products, and phenotype complex traits, aimed at illustrating a DNA to RNA to phenotype pathway. We call the networks composed of DNA markers, transcriptional genes, and phenotypic traits genetic or SNP-SNP interaction networks, gene (regulatory) networks, and phenotypic networks, respectively. We use positive or negative epistasis to define interactions between different markers and synergism and antagonism to define the pattern of how transcriptional genes (or phenotypic traits) are co-regulated. We assume that phenotypic traits are causally regulated by transcriptional genes, which are controlled by DNA markers. We refer to the DNA markers that are significantly associated with gene expression or phenotypic variation

QTLs.

Circadian oscillators function through highly interconnected, autoregulatory gene networks that contain transcription-translation feedback loop motifs [1,16,34,72]. To accommodate this complexity of rhythmic processes, we describe a general framework for reconstructing SNP-SNP interaction and gene regulatory networks that cover all genome-wide genes in order to chart a complete atlas of the molecular mechanisms underlying circadian rhythms. Suppose there is a circadian clock constituted by a number of genes that are rhythmically expressed to regulate behavior and physiological traits in a way that conform to the daily environmental cycle of light/dark. All these processes are encoded by an unknown number of DNA variants distributed throughout the genome. An oscillating clock is large in dimension, complex in structure, heterogeneous in organization, and diverse in function. Despite these recalcitrant characteristics, we can view it as a multiplayer game. Originating in economic research [73], game theory studies and models the payoff of one player based on the strategy implemented by the other player. The application of game theory has been largely popularized by the concept of the Nash equilibrium, a proposed solution of a non-cooperative game, at which each rational agent tends to choose an optimal strategy to maximize its payoff, conditional on the strategies of its opponents, as long as the latter remains unchanged [74]. By combining game theory and evolutionary biology, Smith and Price [75] formulated evolutionary game theory to interpret how frequency dependent fitness drives strategies to evolution [76]. This theory's core is the concept of an evolutionarily stable strategy regarded as an equilibrium refinement of the Nash equilibrium and its extension to population evolution. However, Smith and Price's evolutionary game theory serves as the *static* analysis tool of evolutionary stability because it does not attempt to model how strategies change in a population. By adding the time dimension, we expand evolutionary game theory to its *dynamic* domain, making it possible to explicitly model the change of strategy frequencies in the population. Such a dynamic evolutionary game theory (dEGT) does not need to specifically define a notion of evolutionary stability because, by specifying a population dynamic model, all of the standard stability concepts may be used to characterize dynamical systems. As such, if a dynamic model is developed, we can implement dEGT to characterize how a player achieves its payoff differently over time through its own strategy and the strategies implemented by other players.

We interpret a circadian clock through the lens of dEGT. We view entities comprising the clock as interactive players. The magnitude at which an entity regulates rhythmicity depends on the intrinsic capacity of this entity and the extrinsic influences of other entities on it. Let $g_j(t)$ denote the effect of entity j on a rhythmic trait at time t ($t = t_1, \ldots, t_T$), which can be characterize by a non-linear quasispecies (or non-linear Lotka-Volterra) equation. A whole rhythmic network is composed of q such quasispecies equations, specified by a system of ordinary differential equations (ODEs), i.e.,

$$g_j'(t) = Q_j\left(g_j(t): \Theta_j\right) + \sum_{j'=1, j' \neq j}^{q} Q_{jj'}\left(g_{j'}(t): \Theta_{jj'}\right), j = 1, \ldots, q \quad (1)$$

where the net effect of entity j includes two components: *independent* effect $Q_j(g_j(t): \Theta_j)$ that is expected to occur when this entity is assumed to be socially isolated and accumulated *dependent* effect $\Sigma Q_{jj'l}(g_{j'l}(t): \Theta_{jj'l})$ that results from the influence of other entities j' ($j' = 1, \ldots, j-1, j+1, \ldots, m$) on the focal entity. The pattern and strength of how an entity acts independently are determined by its own innate strategy, whereas how and how much the action of this entity is affected depend on the strategies of other entities. Thus, we express the independent effect of entity j as a function of $g_j(t)$ and its dependent effect as a function of $g_{j'}(t)$. Equation (1) represents a mathematically formulated dEGT framework for systematically characterizing inter-entity interactions and their impacts on circadian changes. Given the uniqueness of the above derivation procedure, we call networks reconstructed under this procedure systems evolutionary game networks (SEGNs).

3. Statistical reconstruction of genetic SEGNs

3.1. Building up SEGN equations by functional mapping

To reconstruct the SEGN at the DNA level, we need to obtain genetic effects of individual SNPs on biological rhythmicity, which can be used to formulate the nonlinear quasi-species equations, as described by equation (1). At this time, j denotes a SNP, q denotes the number of SNPs, and $g_j(t)$ denotes the genetic effect of SNP j on circadian rhythms at time t. Functional mapping is a dynamic mapping approach that can estimate the temporal pattern of genetic effect or genetic variance due to single significant SNPs or single SNP pairs chosen from a genome-wide pool of markers for any mapping or association populations [64-66]. This approach has proven itself to be powerful for QTL mapping in a wide variety of species [77-84]. We initiate a mapping or association study composed of n individuals that are genotyped at q genome-wide SNPs and phenotyped for a rhythmic trait repeatedly at a series of T discrete time points. Let $\mathbf{y}_i = (y_i(t_1), \ldots, y_i(t_T))$ denote a vector of measured values of a rhythmic trait for individual i at T discrete time points. Consider a SNP with K genotypes whose observations are denoted by n_k ($k = 1, \ldots, K$), respectively. Functional mapping formulates a likelihood for n trait vectors at this SNP, expressed as:

$$L(\boldsymbol{\mu}; \boldsymbol{\Sigma}) = \sum_{k=1}^{J} \prod_{i=1}^{n_k} f_k(\mathbf{y}_i | \boldsymbol{\mu}_k; \boldsymbol{\Sigma}), \quad (2)$$

where $f_k(\mathbf{y}_i | \boldsymbol{\mu}_k; \boldsymbol{\Sigma})$ is a T-dimensional normal distribution for individual i with mean vector for genotype j ($\boldsymbol{\mu}_k$) and covariance matrix $\boldsymbol{\Sigma}$. Functional mapping implements biologically meaningful mathematical equations of trait formation to model genotype-typical mean vectors [77]. Many parametric approaches, such as the first-order autoregressive (AR(1)) model [77,78], the first-order structured antedependence (SAD(1)) model [85,86], the autoregressive moving average (ARMA) model [87], and Brownian motion process, have been used to fit the structure of the covariance matrix. Compared with parametric approaches, nonparametric approaches based on B-splines or Legendre Orthogonal Polynomials (LOP) may better model the covariance structure [88-90]. The best approach that structures the longitudinal covariance matrix of real

data can be chosen based on information criteria. Joint mean-covariance modeling in functional mapping can enhance the biological relevance of QTL detection and its statistical power.

In rhythmic biology, the cyclic change of trait values can be approximated by mathematical functions [91,92]. Fourier series are considered one of the universal approximators for rhythmic models [58,59,93-96]. We model time-varying genotypic values in $\boldsymbol{\mu}_k$ by the Fourier signal, expressed as

$$\mu_k(t) = a_{k0} + \sum_{r=1}^{R}\left(a_{kr}\cos\left(\frac{2\pi rt}{\mathcal{T}_k}\right)\right) + \sum_{r=1}^{R}\left(b_{kr}\sin\left(\frac{2\pi rt}{\mathcal{T}_r}\right)\right), \quad (3)$$

where a_{k0} is the trait mean of genotype k over time, $\mathcal{T}_k$ is the period of rhythmic cycle for genotype k, a_{kr} and b_{kr} are the coefficients of cosines and sines at the rth harmonic, from which the amplitude and phase of rhythmic change for genotype k are calculated as $A_{kr} = \sqrt{a_{kr}^2 + b_{kr}^2}$ and $\phi_{kr} = \tan^{-1}(-b_{kr}/a_{kr})$, and R is the number of harmonics that best fits the observed data by statistical reasoning. A number of parametric, nonparametric, or semiparametric approaches have been developed to model the covariance matrix [59,88].

Statistical algorithms are implemented to solve the likelihood of equation (2) and obtain the maximum likelihood estimates (MLEs) of the Fourier series parameters (a_{k0}, $\mathcal{T}_k$, a_{kr}, b_{kr}) for each genotype j. We plug in these estimates into equation (3) to calculate time-varying genotypic values of each genotype from which we calculate time-varying genetic effects or variances explained by the SNP under consideration. In a hypothetical example of rhythmic mapping (**Fig. 1**), we demonstrate how functional mapping can be used to detect genotypic differences in rhythmic curve. At SNP 1, three genotypes AA, Aa, and aa were detected to differ in amplitude but not in phase and period (Fig. 1**A**). Genotypic differences at SNP 2 follow a different pattern; genotypes AA and Aa differ in amplitude but are similar in phase and period, both differ from genotype aa in the three rhythmic parameters (Fig. 1**B**). For a backcross or recombinant inbred line mapping population, there are only two genotypic curves at a SNP, whose difference is used to describe the genetic effect of this SNP. For an F_2 mapping population or human association study population, three genotypes at a SNP may produce two effects, additive and dominant.

These estimated effects at different SNPs are implemented into nonlinear quasi-species equations (1) for network reconstruction. In quantitative genetics, the contribution of a gene to trait variation can be described by the genetic variance that is explained by this gene. Here, we estimate and use the genetic variance of each SNP for the subsequent modeling. It is interesting to find that genetic variance at a SNP also changes rhythmically, although the pattern of change differs between SNPs (Fig. 1**C**). Let $g_j(t)$'s that build up equation (1) denote time-varying genetic variances accounted for by a SNP j ($j = 1, \ldots, p$), which are marginal or net genetic variances used for SEGN inference.

3.2. Network sparsity and variable selection

In a mapping study, Jiang et al. [71] showed that the estimated net genetic variances by functional mapping implemented in equation (1) can reconstruct a gene network for growth trajectories. This implementation can characterize the detailed interaction pattern of genes, but it only illustrates the partial architecture of epistasis among a limited set of significant loci that are first chosen from single marker analysis. It is possible that virtually all genes in the genome participate in mediating a complex trait or disease [97] such that it is essential to reconstruct a genome-wide SEGN that covers the interactome of all genes.

Although we attempt to encapsulate all genes into circadian networks, this does not mean that we would reconstruct a completely connected network in which each gene is linked with every other gene. Instead, we will need to reconstruct sparsely connected networks, in which most genes have a low number of links. Ample evidence from a variety of data analyses suggests that biological networks, ranging from metabolic gene-regulatory to species interaction networks, are sparse; i.e., the percentage of active interactions scales inversely with the network size [98-102]. This is different from inanimate networks and telecommunication networks that are connected by a complete number of links. Several studies have begun to explore why interaction networks in living systems universally possess a non-random architecture and sparsity [103]. It has been generally suggested that sparsity is an emergent property that enables the interactive

system to better adapt to newly intervening changes and remain stable after perturbations of the underlying dynamics [104,105].

Reconstructing a sparse circadian network based on the ODEs of equation (1) is equivalent to identifying and choosing a small number of SNPs that regulate a focal SNP from a huge pool of genome-wide SNPs. To do so, we formulate a regression model to describe the genetic variance value of each SNP as a linear combination of the genetic variance of all other SNPs across time points (equivalent to samples for a traditional regression model). However, because the number of SNPs (q) is significantly larger than the number of time points (T), we encounter the "curse of dimensionality" for model overfitting. We implement LASSO-based variable selection [106] to choose a subset of the most significant SNPs (predictors) that interact with a focal SNP. LASSO can only select at most T SNPs before it saturates, but several versions of its modifications, such as elastic net [107], group LASSO [108], and adaptive group LASSO [109], can address the technical issue of $q >> T$. We can also address this issue by augmenting "sample size" through the curve fitting of genetic variance over time.

The Fourier approximations of genotype values by equation (3) allows us to calculate and interpolate an infinite number of genetic variance values of each SNP expressed during the rhythmic cycle, which provide an infinite number of "samples" to perform LASSO-based variable selection with any dimension of SNP data. Using the interpolated genetic variance values denoted as $z_j(t)$ ($t = 1, \ldots, T$), we formulate a regression model to characterize how a SNP (say j) is affected by all other SNPs (j'), expressed as

$$z_j(t) = \alpha_j + \sum_{j'=1, j' \neq j}^{q} \beta_{j'} z_{j'}(t) + \varepsilon_j(t), j = 1, \ldots, q \quad (4)$$

where α_j and $\beta_{j'}$ are the constant and the regression coefficient of SNP j'as a predictor, respectively, and $\varepsilon_j(t)$ is the residual error. The basic principle of LASSO to disentangle the $q >> T$ problem in order to minimize squared error loss (L^2 loss) under a penalty on the sum of the absolute values of the coefficients (an L^1 penalty). Under this principle, the solution of equation

(4) tends to find estimates of $\beta_{j'}$ ($j' = 1, \ldots, j-1, j+1, \ldots, m$) that are mostly zero. By plotting the genetic variance of SNP 2 against that of SNP 1 (used in Fig. 1's example), we found that their relationship can be better fitted by a nonparametric approach (Fig. 1**C**). Thus, we implement a general nonparametric approach for variable selection on regression model (4). Thus, through variable selection procedure, we will find a small subset of the most significant SNPs (say d_j) that link with each focal SNP j ($j = 1, \ldots, m$), which allows us to rewrite the ODEs of equation (1) as

$$g_j'(t) = Q_j\left(g_j(t); \Theta_j\right) + \sum_{j'=1, j' \neq j}^{d_j} Q_{jj'}\left(g_{j'}(t); \Theta_{jj'}\right), j = 1, \ldots, m \quad (5)$$

where $Q_j(g_j(t); \Theta_j)$ and $Q_{jj'}(g_{j'}(t); \Theta_{jj'})$ are defined as above. Although a majority of regression coefficients in equation (4) are shrunk to be zero, we pose no constraints on the number of ODEs such that the dimension of networks remains unchanged. Equation (5) affords an interdisciplinary platform on which evolutionary game theory and network theory are cross-pollinated through statistical variable selection to reconstruct a series of gene networks. These networks are high-dimensional (q SNPs), highly sparse ($d_j << q$), and mobile (as a function of t).

3.3. Statistical algorithms for ODE solving

The ODEs in equation (5) can be solved by implementing mathematical and statistical algorithms. Genetic variance $g_j(t)$ is calculated from genotypic values $\mu_k(t)$ that have an explicit periodic form, fitted by a Fourier series approximation (2), but we do not know about the form of $Q_j(g_j(t); \Theta_j)$ and $Q_{jj'}(g_{j'}(t); \Theta_{jj'})$. Thus, these two functions can be better fitted by a nonparametric approach. Because of its favorable property as an infinitely differentiable function, we implement the LOP-based approach for smoothing time-varying independent and dependent genetic variances through the parameters Θ_j and $\Theta_{jj'}$, respectively. Many mature mathematical techniques have been available for studying numerical or theoretical properties of ODE models (e.g., sensitivity and bifurcation analysis) [110-112]. In the past decade, many statistical algorithms have emerged for estimating ODE parameters from the noisy data. These

methods includes Ramsay et al.'s generalized profiling approach [113], Liang and Wu's two-step derivative-based local polynomial regression approach [114], Cao et al.'s penalized least square method [115], Brunel et al.'s gradient matching approach [116], Li et al.'s regularization estimation approach [117], and Chen et al.'s derivative-free approach [118]. Each of these methods has its own advantages and disadvantages in parameter estimation, power, and computational efficiency. For example, the derivative-free approach is very flexible to handle noisy data. Because gene expression data often contain unknown noises, we hybridize the derivative-free approach with the simplex (Nelder-Mead) algorithm under a likelihood setting to obtain the maximum likelihood estimates (MLEs) of ODE parameters in equation (5).

Let $P_j(t)$ and $P_{jj'}(t)$ denote the integrals of the MLEs of $Q_j(g_j(t):\Theta_j)$ and $Q_{jj'}(g_{j'}(t):\Theta_{jj'})$, respectively. Then, we code $P_j(t)$ as a node and $P_{jj'}(t)$ as an edge into q-node networks. Such networks, i.e., SEGNs, cover all possible genes that take a part in circadian rhythms along direct or indirect pathways. SEGNs can contextualize bidirectional, signed, and weighted gene interactions into fully informative graphs and, thereby, own many favorable features that are unavailable to commonly used correlation- and Bayesian-based networks.

3.4. Mechanistic characterization of epistasis constituting rhythmic clock networks

The phenomenon by which the impact of one gene on a phenotype is determined by other genes is called epistasis [119]. Classic quantitative genetic theory can estimate the size of epistasis, but fails to characterize its causality and the direction of its causality [120]. The SEGN is the interdisciplinary integration that combines separate perspectives through the development of mechanistic connections among them to establish a more cognitive and empirical approach toward the epistatic identification of clock genes. The pattern of how and how strongly SNP *j* is affected by SNP *j'* can be assessed by $P_{jj'}(t)$. If this value is positive, zero, or negative, then this suggests that SNP *j'* activates, is neutral to, or inhibits SNP *j*, respectively. By comparing $P_{jj'}(t)$ and $P_{j'j}(t)$, we can classify SNP-SNP interactions into five qualitatively different types:

- Positive epistasis by which two interactive SNPs activate each other. This can be seen if both $P_{jj'}(t)$ and $P_{j'j}(t)$ are positive;

- Negative epistasis by which two interactive SNPs inhibit each other. This can be seen if both $P_{jj'}(t)$ and $P_{j'j}(t)$ are negative;
- Directional positive epistasis by which SNP j' activates SNP j but the latter is neutral to the former. This can be seen if $P_{jj'}(t)$ is positive but $P_{j'j}(t)$ is zero;
- Directional negative epistasis by which SNP j' inhibits SNP j but the latter is neutral to the former. This can be seen if $P_{jj'}(t)$ is negative but $P_{j'j}(t)$ is zero;
- Altruistic/exploitation epistasis in which one SNP activates the other but the latter inhibits the former. If $P_{jj'}(t)$ is positive whereas $P_{j'j}(t)$ is negative, this suggests that SNP j' offers altruism to SNP j, or say, SNP j exploits SNP j'.

It is possible that the two SNPs may peacefully coexist when they do not affect each other. This can be seen if both $P_{jj'}(t)$ and $P_{j'j}(t)$ are zero. The SEGN is also a quantitative network, because each activation or inhibition is quantified by a value. If $P_{jj'}(t)$ and $P_{j'j}(t)$ are positive and their values are equal, the positive epistasis of two SNPs j and j' is regarded as symmetrical positive epistasis. If $P_{jj'}(t)$ and $P_{j'j}(t)$ are positive but their values are not equal, then positive epistasis becomes asymmetrical positive epistasis. Similarly, we can distinguish between symmetrical negative epistasis and asymmetrical negative epistasis. Table 1 condenses the salient features of a SEGN. Taken together, the definitions and interpretations of various patterns of gene interactions can facilitate the exploration of the mass, energetic, or signal basis for each interaction, surpassing the traditional notation of epistasis in terms of its biological relevance.

The central themes of network reconstruction include sparsity, stability and causality [121-123]. As described above, the implementation of ODEs meets the causality property of a network by determining the direction of gene interaction. As shown in Box 1, the statistical procedure for learning the SEGN is formulated under the maximum likelihood and convex optimality setting. Thus, various strategies each SNP chooses to interact with different SNPs can be thought to achieve the maximum stability of the network [122]. As predicted by network theory, there is a limit to the number of links owned by each node in a network [124]. We can implement variable

selection to detect the number of the most significant SNPs that affect a focal SNP. Taken together, we can reconstruct sparse, stable, and casual gene networks for circadian rhythms.

Networks are regarded as snapshots of biological systems at different times. Uncovering the dynamic nature of genetic networks can shed light on the genomic mechanisms that drive circadian rhythms. As a function of time t, $P_{jj'}(t)$ can be calculated at any time point from $t = 0$ to T. In an example illustrated by **Figure 2**, we simulated 10 SNPs each with genetic variance spinning cyclically with time in a different manner. We can reconstruct mobile SEGNs using these SNPs along the time axis. We show such SEGNs at three representative time points, 15, 30, and 60 h, in a rhythmic cycle. Although network topologies do not change from time 15 to 60, the quantitative organization of the network dramatically varies from time to time. For example, SNP3 and SNP 8 establish a relationship of weak symmetric negative epistasis at time 15, but this relationship is changed to sizeable altruistic/repressive epistasis with SNP 8 promoting SNP 3 but with SNP 3 inhibiting SNP 8 at time 30, which becomes even stronger at time 60. In general, the strength of SNP-SNP interactions increases with time. Indeed, SEGNs can be reconstructed instantaneously, which are equipped with a capacity to establish a real-time visualization of genetic networks during biological processes. Such momentary monitoring facilitates the detection of genetic disruption in circadian rhythms, thereby providing a quantitative approach for rhythmically-related disorder prediction.

3.5. Network modularity and functional clustering

It has been widely recognized that biological networks across nearly an entire range of scales from molecules all the way up to the whole organism can be divided into smaller communities or modules that have strong internal interactions but are relatively autonomous with respect to each other [125,126]. This phenomenon, called network modularity, has received considerable attention in biological and biomedical research [127-129]. Genes within modules function similarly and vary together, but they are independent from the function of other genes. Such structural and functional diversity of gene networks enhance the robustness of biological systems to environmental perturbations, showing a widespread implication for mediating developmental and evolutionary processes. A number of computational algorithms have been developed to

detect and characterize modular structure in networks by revealing the occurrence of densely connected groups of vertices, with only sparser connections between groups [130-134].

Genes that display a similar pattern of time-varying gene expression profiles are attributed to the same group. These similarly differentiated genes form the same modules, which are less similar in expression pattern to those from different modules. We implement functional clustering [59,60,135] to classify SNPs into an optimal number of distinct groups, each representing a different module within clock networks. Let $\mathbf{g}_j = (g_j(t_1), \ldots, g_j(t_T))$ denote a vector of genetic variances due to SNP j at time points (t_1, …, t_T). To group q SNPs into R modules according to how they act with time, functional clustering formulates a mixture-based likelihood as

$$L(\mathbf{y}) = \prod_{j=1}^{q} \left[\pi_1 f_1(\mathbf{g}_j) + \cdots + \pi_R f_R(\mathbf{g}_j)\right] \tag{6}$$

where π_r is the proportion of SNPs within module r (r = 1, …, R) to all SNPs, $f_r(\mathbf{g}_j)$ is the T-variate normal distribution of SNP j over time with mean vector $\boldsymbol{\mu}_r = \left(\mu_r(t_1), \ldots, \mu_r(t_T)\right)$ and covariance matrix $\boldsymbol{\Sigma}$. The form of rhythmic genetic variance curves may be unknown so that a nonparametric smoothing approach can be used to model mean vector $\boldsymbol{\mu}_r$. In a simulated example derived from the emulated real data, we found that genetic variance explained by a SNP obeys pattern of periodic change with time (Fig. 1**C**), in which case Fourier series approximation can be used to model $\boldsymbol{\mu}_r$ for more efficient fitting. Many approaches can be used to model the matrix $\boldsymbol{\Sigma}$, including AR(1), SAD(1), and AR(u,v)MA models [57] and nonparametric models [88,89]. A specific optimal procedure must be formulated to select a model that structures the covariance matrix for a given dataset.

Module proportions and parameters that model mean vectors and the covariance matrix can be estimated by implementing an EM-(Nelder-Mead) simplex hybrid algorithm. In the E step, we

calculate the posterior probabilities of each SNP *j* that belongs to a particular module *r*, by equation

$$\Pi_{rj} = \frac{\pi_r f_r(\mathbf{g}_j)}{\pi_1 f_1(\mathbf{g}_j) + \cdots + \pi_L f_L(\mathbf{g}_j)}, \quad (7)$$

and in the M step, we estimate the proportion of module *l* among all genes by equation

$$\pi_r = \frac{1}{q}\sum_{j=1}^{q} \Pi_{rj}. \quad (8)$$

In the M step, the vector-covariance modeling parameters are estimated by the simplex algorithm. The E and M steps are repeated until the estimates are stable. The optimal number of modules, L_o, can be determined by information criteria, such as AIC or BIC. Based on the posterior probabilities of each SNP estimated by equation (7), we can assign SNPs into these *R* distinct modules. We used a hypothetical example to show how functional clustering can be used to classify 35,000 SNPs based on their rhythmic patterns of genetic variation. Under three different orders of the Fourier series approximation, we calculated AIC values when different number of modules are assumed. An optimal number of modules occurs at 24 for the third order where the AIC value is minimized (**Fig. 3**). We chose 10 of the 24 modules to show that the time-varying trends of genetic variances are dramatically different among rather than within modules. For example, genetic variances of SNPs within modules 1, 6, 7, and 10 spin slightly rhythmically with time, although their values are highly module-dependent. There are remarkable periodic changes in genetic variance for other modules, but the phase, period and amplitude of rhythmic cycles are largely different among the modules.

Let q_r denote the number of SNPs that belong to module *r* (*r* = 1, …, *R*). We take the time-varying means of genetic variance over all SNPs within the same modules and use these means to reconstruct an *R*-node interaction network among modules. We can further reconstruct *R* interaction networks among SNPs within each of the *R* modules. Thus, a large SNP network is

decomposed into multiple functionally different but interconnected network communities based on the theory of biological modularity. If the number of SNPs within a module is still large, a further analysis using functional clustering can be conducted to identify more fine-grained network communities. In the end, we will reconstruct a multilayer SNP interactome network, i.e., SEGN, that encapsulates all types of interactions for a complete set of genome-wide SNPs from a circadian study. Figure 4 illustrates the hierarchical structure of such a multiplayer SEGN, at the top tier of which a network was reconstructed with five modules. Each module contains SNPs that display a similar rhythmic pattern of genetic variance according to a certain criterion, but some modules can be further classified if a more stringent criterion is used. For example, module 1 is dissected into four submodules that build a network at the second tier, among which submodule 3 is further split into 8 sub-submodules to form the third-tier network. Similarly, module 3 and 5 can be decomposed into more fine-grained networks. Multiplayer networks facilitate the fine detection of gene-gene interactions, but at the same time, cover all SNPs from an association study, thus allowing a platform to test a variety of hypotheses regarding the molecular mechanisms of biological clocks at unprecedented resolution and coverage.

4. QTL control of gene regulatory SEGNs

4.1. Clocks as a high-dimensional, dynamic, and mechanistic network

Circadian clocks can be regarded as a molecular apparatus composed of clock genes at the transcriptional level. Some of these transcriptional genes regulate rhythmic patterns, whereas others cyclically change their expression in response to the regulation of biological cycles [136-138]. Network alteration of these regulating or regulated processes may interpret causes of circadian clock dysregulation for diseased patients [139]. Reconstructing such gene regulatory networks (GRNs) from expression data has not only attracted the attention of computational biologists [67-70], but also captivated the interest of engineers [121-123]. However, although a plethora of computational approaches have been developed, fully informative networks that capture all fundamental properties of interactions, including causality, the direction of causality, feedback cycle, strength, and mobility, are quite lacking. More recently, Chen et al. [140] proposed a new computational model for reconstructing mechanistic and dynamic interaction

networks from gene expression data at any dimension, showing its potential application to disentangle biological complexities [141]. By accommodating the cyclic property of clock genes, this model can be modified and implemented to identify oscillating gene networks.

There has been ample evidence for the genetic control of gene co-expression related to various biological processes [142-144]. To our knowledge, there have been no studies thus far that can characterize whether and how specific QTLs control large-scale GRNs for circadian rhythms, but we argue that such QTLs exist, which may affect biological systems at various scales. Our tasks are not only to reconstruct GRNs from gene expression data, as reported in previous studies [67-70,121-123], but also to develop a methodology for mapping QTLs that modulate these GRNs. The identification of QTLs involved in gene networks allow us to gain new insight into the molecular mechanisms of rhythmic processes.

4.2. Mapping oscillating gene networks

Likelihood model: In a mapping or association study population, n individuals are genotyped, monitored for m transcriptional genes, and phenotyped for p rhythmic traits. Transcriptional profiles and phenotypic traits are measured repeatedly at a series of T discrete time points. Let $\mathbf{y}_{li} = (y_{li}(t_1), \ldots, y_{li}(t_T))$ denote the expression vector of gene l measured for individual i at T time points. Consider a SNP of K genotypes each with a size denoted as n_k ($k = 1, \ldots, K$). The likelihood of gene expression data at this SNP is formulated as

$$L_1(\boldsymbol{\mu}; \boldsymbol{\Sigma}) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} f_k(\boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{mi} | \boldsymbol{\mu}_{1k}, \ldots, \boldsymbol{\mu}_{mk}; \boldsymbol{\Sigma}) \quad (9)$$

where $f_k(\cdot)$ is the n-dimensional m-variate normal distribution for m genes across T time points with the mean vector $\boldsymbol{\mu}_k$ for SNP genotype k and covariance matrix $\boldsymbol{\Sigma}$. Specifically, we have

$$\boldsymbol{\mu}_k = (\boldsymbol{\mu}_{1k}; \ldots; \boldsymbol{\mu}_{mk}) = \big(\mu_{1k}(t_1), \ldots, \mu_{1k}(t_T); \ldots; \mu_{mk}(t_1), \ldots, \mu_{mk}(t_T)\big), \quad (10)$$

where $\mu_{lk}(t)$ is the mean expression level of genotype k for gene l (l = 1, …, m) at time t. The expression amount of each gene includes independent and dependent expression components. The independent component is one that occurs when this gene is assumed to be in isolation, whereas the dependent component is formed due to the regulation of other genes for the focal gene. This argument can be mathematically formulated by a system of genotype-specific ODEs, expressed as

$$\mu'_{lk}(t) = W_{lk}(\mu_{lk}(t): \Theta_{lk}) + \sum_{l'=1, l' \neq l}^{d_{lk}} W_{ll'k}(\mu_{l'k}(t): \Theta_{ll'k}) \quad (11)$$

where $W_{lk}(\mu_{lk}(t): \Theta_{lk})$ is the independent expression component of gene l, determined by this gene's innate capacity expressed as a function of $\mu_{lk}(t)$; $W_{ll'k}(\mu_{l'k}(t): \Theta_{ll'k})$ is the dependent expression component of gene l that results from the strategy implemented by gene l', expressed as a function of $\mu_{l'k}(t)$; and d_{lk} (<< m) is the number of genes that significantly influence gene l for genotype k. The determination of d_{lk} is made by a LASSO-based variable selection built on a regression model of $y_{li}(t)$ as a response on $y_{l'i}(t)$ (l' = 1, …, l-1, l+1, …, m) as predictors over all individuals carrying the same genotype across time points. We allow d_{lk} to vary across SNP genotypes for the same gene, because a focal gene may be regulated by different sets of genes for different genotypes.

We implement a nonparametric approach to smooth $W_{lk}(\mu_{lk}(t): \Theta_{lk})$ and $W_{ll'k}(\mu_{l'k}(t): \Theta_{ll'k})$, specified by the unknown parameters Θ_{lk} and $\Theta_{ll'k}$, respectively. These independent and dependent expression components constitute the time-varying mean vector of gene l for SNP genotype k. Since the basis function is built on $\mu_{lk}(t)$ or $\mu_{l'k}(t)$ that contains time information, we can still implement autoregressive models, such as AR(1) or SAD(1), to fit the structure of the covariance matrix $\boldsymbol{\Sigma}$ in the likelihood (9). Under the mean-covariance modeling as described above, we implement mathematical and statistical algorithms to solve the likelihood (9) and obtain the MLEs of ODE parameters for each SNP genotype. By plugging these MLEs into the

independent and dependent expression components in equation (11), we reconstruct a series of oscillating m-node sparse gene networks, denoted as $\mathbb{Q}_k(t)$, for each genotype k.

To determine whether the SNP determines oscillating gene networks, we need to formulate a procedure for hypothesis testing. In case of no significant QTL, we formulate the null hypothesis that gene networks characterized by ODE parameters in equation (11) are invariant among genotypes, i.e., H_0: $\mathbb{Q}_k(t) \equiv \mathbb{Q}(t)$, under which the likelihood is written as

$$L_0(\boldsymbol{\mu}; \boldsymbol{\Sigma}) = \prod_{i=1}^{n} f(\boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{mi} | \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m; \boldsymbol{\Sigma}) \quad (12)$$

where $f(\cdot)$ is parameterized by the mean vector of time-varying expression of m genes over all individuals and covariance matrix modelled by an autoregressive process. The likelihood under the alternative hypothesis, i.e., all individuals with the same SNP genotype share the same network, but individuals with different genotypes have different networks, is formulated by equation (9). The log-likelihood ratio (LR) calculated under the null and alternative hypotheses is used as a test statistic. The genome-wide critical threshold is determined by permutation tests. A SNP is significant for gene networks if the LR value is beyond a threshold value.

Network dissection: Beyond GRNs reconstructed by commonly used correlation-based and Bayesian network approaches, $\mathbb{Q}_k(t)$ can capture all network features by characterizing bidirectional, signed, and weighted gene co-regulation. As shown by equation (11), the pattern of gene co-regulation can be assessed by $W_{ll'}(t)$. The sign and size of this value reflect the promotion, inhibition, or lack of regulation of gene l' on gene l. Based on $W_{ll'}(t)$ and $W_{l'l}(t)$, We can classify gene co-regulation into different types as follows:

- Synergism by which two genes activate each other if both $W_{ll'}(t)$ and $W_{l'l}(t)$ are positive;
- Antagonism by which two genes inhibit each other if both $W_{ll'}(t)$ and $W_{l'l}(t)$ are negative;
- Directional synergism by which gene l' activates gene l but the latter is neutral to the former if $W_{ll'}(t)$ is positive but $W_{l'l}(t)$ is zero or vice versa;

- Directional antagonism by which gene l' inhibits gene l but the latter is neutral to the former if $W_{ll'}(t)$ is negative but $W_{l'l}(t)$ is zero or vice versa;
- Altruism by which gene l activates gene l' but the latter inhibits the former if $W_{l'l}(t)$ is positive but $W_{ll'}(t)$ is negative. Here, the altruism of gene l is the egoism of gene l'.

If both $W_{ll'}(t)$ and $W_{l'l}(t)$ are zero, then this indicates that the two genes l and l' peacefully coexist. For synergism and antagonism, we can further define symmetrical synergism and symmetrical antagonism if $W_{ll'}(t)$ and $W_{l'l}(t)$ of the same sign are equal in size and asymmetrical synergism and asymmetrical antagonism if $W_{ll'}(t)$ and $W_{l'l}(t)$ of the same sign are not equal in size.

Result interpretation: To explain how a QTL determines gene networks, we consider a simulated example with results illustrated in **Fig. 5**. Suppose there is a genotyped mapping population designed to characterize the genetic architecture of rhythmic co-regulation of genes in response to the environmental cycle. As an example, we assume that 10 genes are monitored, but the model can handle any number of genes by incorporating network modularity theory. Based on the above LR test, we identify a set of significant QTLs for rhythmic GRNs from a pool of genome-wide SNPs. By reconstructing 10-node gene networks for three different genotypes at a QTL, we found remarkable genotypic differences in network features. First, three genotypes display pronounced differences in the rhythmic pattern of gene expression profiles (**Fig. 5A**). Notably, 10 genes are expressed more rhythmically in genotype *AA* than genotype *Aa*, both genotypes having a stronger rhythmic pulse than genotype *aa*. Second, GRNs reconstructed from 10 genes are sparser in genotype *AA* than genotypes *Aa* and *aa*. The three gene networks considerably differ in both structure and organization among genotypes. For example, in $\mathbb{Q}_{AA}(t)$, genes 9 and 6 form an asymmetric antagonistic relationship, but this relationship does not exist in $\mathbb{Q}_{Aa}(t)$ and $\mathbb{Q}_{aa}(t)$. Gene 5 exerts a directional synergistic effect on gene 9 in $\mathbb{Q}_{Aa}(t)$, whereas these genes co-exist peacefully in $\mathbb{Q}_{AA}(t)$ and $\mathbb{Q}_{aa}(t)$. By closely investigating how gene-gene relationships differ among genotypes, one can decipher a global and detailed view into the genetic control mechanisms of specific QTLs on rhythmic GRNs.

4.3. How a QTL modulates the emergent properties of gene networks

The procedure described in the preceding section can identify and map network QTLs that mediate the overall structure and organization of oscillating networks. Here, we develop a framework to test how a QTL affects the emergent properties of gene networks. The properties of a network can be described by many parameters, including **connectivity**, describing the number of nodes with which a node links within a network [145]; **closeness**, describing the degree of linkage of one node to other genes [146]; **betweenness**, reflecting the importance of a node as a bridge across the network [147]; **eccentricity**, defined as the longest distance of one node to other nodes [148]; **eigenvector centrality**, describing the importance of node to neighboring nodes [149]; and **PageRank**, evaluating the quality and quantity of links to a network [150]. As defined, these parameters determine the property of a network from different topological aspects.

For each genotype-specific oscillating gene network, we calculate connectivity, closeness, betweenness, eccentricity, eigenvector, and PageRank at different time points and plot these network parameters against time. By comparing the genotype-specific curves of network properties, we can address the following questions: (1) How does a QTL affect network structure in terms of network parameters? (2) How does a QTL pleiotropically affect different network parameters? Answers to these questions can help understand how the human genome encodes instructions of gene regulation for circadian rhythms and find genetic variants that drive genomic differences distinguishing a healthy status from a diseased status.

4.4. Genetic determination of casual gene networks

Clocks contain cyclic genes that drive behavior and physiology to change rhythmically in response to daily cycles. This process operates through high-dimensional, complex casual networks, and, more likely, is controlled by QTLs. The identification of such QTLs can facilitate our mechanistic understanding of genotype-phenotype relationships. Let $(\mathbf{y}_{li}; \mathbf{z}_{li}) = (y_{li}(t_1), \ldots, y_{li}(t_T); z_{si}(t_1), \ldots, z_{si}(t_T))$ denote the expression vector of gene l ($l = 1, \ldots, m$) and the phenotypic vector of trait s ($s = 1, \ldots, p$) measured for individual i at T time points. We extend the likelihood function given in equation (9) for a given SNP to include information about both gene expression and rhythmic traits. Under the expanded likelihood, we model the expanded mean vector for

SNP genotype k by

$$\begin{cases} \mu'_{lk}(t) = W_{lk}(\mu_{lk}(t): \Theta_{lk}) + \sum_{l'=1,l'\neq l}^{d_{lk}} W_{ll'k}(\mu_{l'k}(t): \Theta_{ll'k}) & (13A) \\ \mu'_{sk}(t) = R_{sk}(\mu_{sk}(t): \Theta_{sk}) + \sum_{s'=1,s'\neq s}^{b_{sk}} R_{ss'k}(\mu_{s'k}(t): \Theta_{ss'k}) + \sum_{l=1}^{d_{sk}} R_{slk}(\mu_{lk}(t): \Theta_{slk}) & (13B) \end{cases}$$

where equation (13A) is described as equation (11), which models m-node gene networks, and equation (13B) characterizes how p rhythmic traits interact with each other to form phenotypic networks and how gene networks causally regulate rhythmic processes. In Equation (13B), $R_{sk}(\cdot)$ is the independent phenotypic value of trait s (under the assumption that this trait is isolated), $R_{ss'k}(\cdot)$ is the dependent phenotypic value of traits s that forms due to the influence of other trait s', and $R_{slk}(\cdot)$ is the dependent phenotypic value of trait s resulting from the regulation of gene l. The first two terms can reconstruct phenotypic networks that reflect trait-trait interactions and the third term is used to reconstruct casual networks from genes to phenotypes. These three terms are a function of $\mu_{sk}(t)$, the time-varying genotypic mean of trait s, $\mu_{s'k}(t)$, the time-varying genotypic mean of trait s', and $\mu_{lk}(t)$, the time-varying genotypic mean of gene l, specified by parameters Θ_{sk}, $\Theta_{ss'k}$, and Θ_{slk}, respectively, in a nonparametric way.

The expanded likelihood contains a bivariate covariance matrix, composed of m-dimensional covariance submatrix (genes), p-dimensional covariance submatrix (traits), and ($m\times p$)-dimensional covariance submatrix (genes vs. traits). A bivariate autoregressive model, such as bivariate SAD(1), has proven to be powerful for modeling the longitudinal structure of large covariance matrices [86]. Moreover, the existence of closed forms for the determinant and inverse of the bivariate SAD(1) matrix can increase the computational efficiency, despite the high dimensionality of the covariance matrix.

A similar log-likelihood procedure can be implemented to test whether a SNP is significant in affecting causal networks from genes to rhythmic phenotypes. The advantage of network analysis is that one can identify the roadmap of each node through direct and indirect paths

towards a targeted phenotype. A further testing procedure can be developed to identify which genes and which paths play a critical role in mediating casual relationships.

5. QTL networks of oscillating gene networks

Traditional approaches for mapping genotype-phenotype relationships are mostly based on a marginal single-marker analysis. These approaches are not particularly powerful for studying rhythmic behaviors that contain a large number of genes and their complex genetic interaction networks. In the preceding section, we describe a statistical model for reconstructing QTL networks that can systematically characterize the genetic control of circadian rhythms. To reveal the genomic internal workings behind biological processes from genotype to rhythmic phenotype, we describe a second statistical model for mapping individual QTLs that mediate gene regulatory networks of circadian rhythms. Here, we formulate a framework that unifies the above two models to identify and reconstruct QTL networks of gene networks. Let $g_{lj}(t)$ denote the genetic variance of gene l (l = 1, …, m) due to SNP j (j = 1, …, q) at time t (t = 1, …, T). We argue that the effect of SNP j on gene l is different in terms of three scenarios as follows. In scenario 1, SNP j only affects gene l but not any other genes, in a way that is independent from other SNPs. In scenario 2, SNP j receives epistatic interactions in a way that its effect on gene l is influenced by other SNPs. In scenario 3, SNP j pleiotropically affects multiple genes so that its effect on gene l is regulated by other genes. The value of $g_{lj}(t)$ that contributes to the genetic architecture is the net consequence of these three scenarios. Using the dEGT model, we decompose $g_{lj}(t)$ using the following equations,

$$g'_{lj}(t) = Q_{lj}\left(g_{lj}(t):\Theta_{lj}\right) + \sum_{j'=1,j'\neq j}^{d_{lj}} Q_{ljj'}\left(g_{lj'}(t):\Theta_{ljj'}\right) + \sum_{l'=1,l'\neq l}^{b_{lj}} W_{ll'j}\left(g_{l'j}(t):\Theta_{ll'j}\right) \quad (14)$$

where $Q_{lj}\left(g_{lj}(t):\Theta_{lj}\right)$ is the independent genetic variance of gene l due to SNP j that is expected to occur when both SNP j and gene l are assumed to be in isolation, $\sum Q_{ljj'}\left(g_{lj'}(t):\Theta_{ljj'}\right)$ is the *epistatic dependent genetic variance* of gene l due to interactions of other d_{lj} SNPs with SNP j, and $\sum W_{ll'j}\left(g_{l'j}(t):\Theta_{ll'j}\right)$ the *pleiotropic dependent variance* of gene l due to influences of

other b_{lj} genes under the control of SNP j. Note that the number of the most significant SNPs that interact with SNP j ($d_{lj} << p$) and the number of the most significant genes that regulate gene l ($b_{lj} << m$) are determined by LASSO-based variable selection. A likelihood is formulated to solve the ODEs of equation (14) by implementing LOP-based nonparametric models to smooth the independent genetic variance, the epistatic dependent genetic variance, and pleiotropic dependent variance, with ODE parameters Θ_{lj}, $\Theta_{ljj'}$, and $\Theta_{ll'j}$, respectively, and autoregressive models to fit the covariance structure. The MLEs of ODE parameters allow us to reconstruct a tridimensional QTL interaction network of gene networks.

The nonlinear quasispecies equations (14) can be expanded to include rhythmic phenotypes. Let $g_{sj}(t)$ denote the genetic variance of rhythmic trait s ($s = 1, \ldots, p$) explained by SNP j ($j = 1, \ldots, q$) at time t ($t = 1, \ldots, T$). By adding $g_{sj}(t)$ to the replicator equations, we can reconstruct a tridimensional QTL interaction network of causal gene-phenotype networks for circadian rhythms. To explain how such a tridimensional network works, we hypothesize a clock mediated by three QTLs that regulate three genes and two rhythmic traits (**Fig. 6**). This tridimensional network includes intertwined pleiotropic networks and epistatic networks across DNA sequences. In the pleiotropic networks, we can characterize how a QTL pleiotropically affects the expression of multiples genes and multiple rhythmic traits to form dynamic causal regulatory networks from genes to phenotypes. Under the control of SNP 1, gene 1 inhibits gene 3 and also promotes gene 2 that promotes phenotype 1, but these causal roadmaps change under the control of the other SNPs. For example, SNP 2 changes the relationship between gene 1 and 3; under its control, gene 1 is inhibited by gene 3 that promotes phenotype 1.

In the epistatic networks (**Fig. 6**), we can characterize how different SNPs interact with each other to determine the expression of a gene or a phenotypic trait. Gene 1 is affected by an epistatic network in which SNP 1 promotes SNP 3 and SNP 2, whereas SNP 2 inhibits SNP 3. Phenotype 1 is affected by an epistatic network that is structurally the same as but quantitatively different from the epistatic network of gene 1. For other genes and phenotypes, we found distinct differences in epistatic network topology. Taken together, the tridimensional network charts the change of the pleiotropic landscape of genes and phenotypes from SNP to SNP and the change of the epistatic landscape of SNPs from genes to phenotypes.

6. Conclusions and future directions

A clock contains numerous cyclically expressed genes that mediate biological rhythms, a process encoded by the genome. Over the past decades, mutagenesis-based molecular genetic analysis has considerably contributed to the identification of clock genes that are required for rhythmic oscillations in response to the light/dark cycle, establishing the fundamental understanding of the genetic mechanisms underlying circadian rhythms. The 2017 Nobel Prize in Physiology or Medicine was awarded to Jeffrey C. Hall, Michael Rosbash and Michael W. Young for their pioneering work in this establishment [20]. With the advent of advanced sequencing techniques, current molecular studies have shifted from the identification of individual rhythmic genes to the genome-wide landscaping of transcriptomic genes [137,138,151]. This paradigm shift has led to the discoveries of a number of new genes that rhythmically synchronize cellular metabolism and organismal behavior through the internal oscillators, or clocks.

The statistical models reviewed in this article can facilitate the promotion of this shift to a generic and wide domain without relying on the use of rhythmic mutants. As a routine genetic approach, linkage or association mapping populations have been produced worldwide for a wide array of species during the past three decades. These populations provide a rich biobank of genetic variants that may be responsible for rhythmic variation [49,152,153]. More importantly, mapping approaches can stimulate the discoveries of new or non-conserved clock genes that are involved in circadian rhythms at the transcriptional level and beyond.

A number of statistical methods have been widely developed and applied to map circadian rhythms. One model, called functional mapping, integrates the mathematical aspects of circadian rhythms to map how a cQTL regulates molecular and physiological profiles rhythmically and test by which parameter, period, phase, or amplitude the cQTL determines the temporal pattern of circadian rhythms [56-62]. Because of a full use of longitudinal measures across multiple points, functional mapping can increase the power of QTL detection. However, most existing models aim to find individual clock QTLs (cQTLs), failing to characterize the genetic complexity of rhythmic physiology and behavior. The omnigenic theory even suggests that a complex trait is determined by all genome-wide distributed genes carried by an organism [97].

Thus, it is highly crucial for reconstructing a systematic network of how each gene acts and interacts with every other gene to contributes to phenotypic variation. The inference of such an omnigenic network is statistically challenging, but once reconstructed, it can provide a powerful tool to extract and excavate the new organizing principles of circadian rhythms.

In this article, we assemble and integrate advanced approaches for QTL functional mapping and gene network reconstruction through high-dimensional statistical modeling into a unified framework for inferring large-scale genetic networks that encompass circadian clocks. Different from GRNs widely reviewed in a range of biological, physical, and engineering literature, this review represents the first among its kinds regarding SNP interactome networks. The reviewed statistical methodology overcomes several technical issues, typical of SNP-SNP network reconstruction. First, GRNs are reconstructed from continuous or semi-continuous abundance data that directly reflect the expression levels of different genes, whereas SNP data describe discrete genotypes of different individuals, which become meaningful for network reconstruction only after genotypes are associated with phenotypes of interest. We translate genetic information carried by individual SNPs into their continuous genetic effects by functional mapping, with which a series of nonlinear quasispecies equations are derived on the basis of evolutionary game theory. These replicator equations establish a basis for a graph theory that codes SNP-SNP interactions into mathematical networks, i.e., SEGNs.

Second, since a complex trait may be controlled by all genes an organism carries, there is a necessity to reconstruct interaction networks that cover a complete set of genes. However, reconstructing a completely linked network from high-dimensional gene data is highly challenging in practice and also is not meaningful from an evolutionary perspective. We implement two advanced statistical approaches, functional clustering and variable selection, to classify all SNPs into distinct modules according to their similarity of temporal genetic effects and select a small set of the most significant SNPs that influences a focal SNP. This type of implementation facilitates the reconstruction of multilayer, sparse, and large-scale genetic networks filled by all SNPs from a mapping or association study. Third, a SNP is determined to be insignificant by commonly used marginal statistical approaches, such as functional mapping, but this detection is the net consequence of the intrinsic action of this SNP and its interactions

with other SNPs. The statistical model reviewed in this article can decompose the net effect of a SNP into its independent effect, expected to occur when this SNP is assumed to be in isolation, and dependent effect resulting from the interactions of other SNPs with the focal SNP. Thus, the insignificance of a SNP by marginal mapping does not necessarily indicates that this SNP is not important in mediating circadian rhythms since it may be confounded by its negative epistasis triggered by other SNPs. Thus, by knocking out these epistatic SNPs, we may clearly understand and better use the genetic effect of this SNP on rhythmic activities.

Fourth, to reveal the causal links from genotype to high-order phenotype, increasing studies have begun to integrate transcriptional, proteomic, and metabolomic profiles into mapping paradigms. However, most of these studies model the individual roles of different genes, proteins, or metabolites, rather the synthetic role of all these entities as a cohesive whole, in mediating genotype-phenotype relationships. By reconstructing a series of networks of networks, the statistical methods reviewed in this article leverage networks as a backbone of linking genotype to phenotype. Networks inferred at a single level of biological organization have been used in the past, but charting information flows of horizontal networks from lower microscopic organization levels (upstream) of molecules to higher macroscopic levels (downstream) of the whole organism via vertical networks has not been explored. Reconstructing intertwined networks is founded on the fundamental premise that networks at different scales share similar global statistical features and structural design principles. The reviewed methods will potentially fill the gap in the systematic, mechanistic characterization of holistic genotype-phenotype relationships for network biology and network medicine.

From the networks reconstructed at different scales, we can not only characterize key significant QTLs responsible for circadian rhythms, but also chart a global picture of how each (significant or insignificant) QTL interacts with every other possible QTL to regulate rhythmic phase, amplitude, and period. A detailed analysis of rhythmic networks enables the discovery of intricate but well-orchestrated structural design principles underlying circadian rhythms. We describe and review the advanced statistical methods for genetic network reconstruction, but there remains much work to be done. One prominent research direction is the incorporation of environmental factors, spatial scales (such as different tissues or cells), and physiological states

into network reconstruction at different levels, allowing the fundamental questions of how environmental and developmental agents affect network topology or how alteration of gene networks mediates rhythmic changes in physiology and behavior to be addressed. We deploy a general statistical procedure to establish a network framework, but mathematical, statistical, computational solutions of framework are likely to remain data-dependent. Optimal techniques should be developed to suit the given data sets especially when the data are heterogeneous. Lastly, and not least, we need to closely collaborate with experimental biologists or clinicians to justify the networks reconstructed by the statistical models reviewed by performing in-vitro or in-vivo experiments. Gene discoveries made by these justified methods will greatly advance the study of chronobiology and the development of chronotherapies.

Acknowledgements

This work is partially supported by NICHD 5R01HD086911-02 from the National Institute of Health (C.G.). C.G.'s work was supported in part by the National Science Foundation under grant DMS-1814876.

References

[1] Young MW, Kay SA. Time zones: a comparative genetics of circadian clocks. Nat Rev Genet 2001;2(9):702-15.

[2] Albrecht U. Timing to perfection: the biology of central and peripheral circadian clocks. Neuron 2012;74:246–60.

[3] Sahar S, Sassone-Corsi P. Regulation of metabolism: the circadian clock dictates the time. Trends Endocrinol Metab 2012;23(1):1-8.

[4] Paschos GK, FitzGerald GA. Circadian clocks and metabolism: implications for microbiome and aging. Trends Genet 2017;33(10):760-69.

[5] Serin Y, Acar Tek N. Effect of circadian rhythm on metabolic processes and the regulation of energy balance. Ann Nutr Metab 2019;74(4):322-30.

[6] Bass J, Lazar MA. Circadian time signatures of fitness and disease. Science 2016;354(6315):994-9.

[7] Bass JT. The circadian clock system's influence in health and disease. Genome Med 2017;9(1):94.

[8] Rijo-Ferreira F, Takahashi JS. Genomics of circadian rhythms in health and disease. Genome Med 2019;11(1):82.

[9] Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH. Resonating circadian clocks enhance fitness in cyanobacteria. Proc Natl Acad Sci U S A 1998;95:8660–8664.

[10] Dodd AN, Salathia N, Hall A, et al. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science 2005;309:630–633.

[11] Kim JA, Kim HS, Choi SH, et al. The importance of the circadian clock in regulating plant metabolism. Intl J Mol Sci 2017;18(12):2680.

[12] Srivastava D, Shamim M, Kumar MA, et al. Role of circadian rhythm in plant system: An update from development to stress response. Environ Exp Bot 2019;162:256-71.

[13] Buhr ED, Takahashi JS. Molecular components of the Mammalian circadian clock. Handb Exp Pharmacol 2013;(217):3-27.

[14] Lowrey PL, Takahashi JS. Genetics of circadian rhythms in Mammalian model organisms. Adv Genet 2011;74:175–230.

[15] Huang RC. The discoveries of molecular mechanisms for the circadian rhythm: The 2017 Nobel Prize in Physiology or Medicine. Biomed J 2018;41(1):5-8.

[16] Cox KH, Takahashi JS. Circadian clock genes and the transcriptional architecture of the clock mechanism. J Mol Endocrinol 2019;63(4):R93-R102.

[17] Bargiello TA, Young MW. Molecular genetics of a biological clock in *Drosophila*. Proc Natl Acad Sci USA. 1984;81:2142–6.

[18] Bargiello TA, Jackson FR, Young MW. Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. Nature 1984;312:752–4.

[19] Reddy P, Zehring WA, Wheeler DA, Pirrotta V, Hadfield C, Hall JC. Molecular analysis of the *period* locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. Cell 1984;38:701–10.

[20] Ibäñez C. The 2017 Nobel prize in physiology or medicine – advanced information: discoveries of molecular mechanisms controlling the circadian rhythm. Nobelprize.org. http://www.nobelprize.org/nobel_prizes/medicine/laureates/2017/advanced.html.

[21] Rosbash MA. 50-year personal journey: location, gene expression, and circadian rhythms. Cold Spring Harb Perspect Biol 2017;9:a032516.

[22] Crews ST, Thomas JB, Goodman CS. The *Drosophila single-minded* gene encodes a

nuclear protein with sequence similarity to the *per* gene product. Cell 1988;52:143–52.

[23] Siwicki KK, Eastman C, Petersen G, Rosbash M, Hall JC. Antibodies to the *period* gene product of *Drosophila* reveal diverse tissue distribution and rhythmic changes in the visual system. Neuron 1988;1:141–50.

[24] Myers MP, Wager-Smith K, Wesley CS, Young MW, Sehgal A. Positional cloning and sequence analysis of the *Drosophila* clock gene *timeless*. Science 1995;270:805–8.

[25] Sehgal A, Rothenfluh-Hilfiker A, Hunter-Ensor M, Chen Y, Myers M, Young MW. Rhythmic expression of *timeless*: a basis for promoting circadian cycles in *period* gene autoregulation. Science 1995;270:808–10.

[26] Vitaterna MH, King DP, Chang AM, et al. Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. Science 1994;264:719–25.

[27] King DP, Zhao Y, Sangoram AM, et al. Positional cloning of the mouse circadian clock gene. Cell 1997;89(4):641-53.

[28] Antoch MP, Song EJ, Chang AM, et al. Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. Cell 1997;89(4):655-67.

[29] Gekakis N, Staknis D, Nguyen HB, et al. Role of the CLOCK protein in the mammalian circadian mechanism. Science 1998;280:1564–1569.

[30] Eriksson ME, Millar AJ. The circadian clock: a plant's best friend in a spinning world. Plant Physiol 2003;132:732-8.

[31] Loza-Correa M, Gomez-Valero L, Buchrieser C. Circadian clock proteins in prokaryotes: hidden rhythms? Front Microbiol 2010;1:130.

[32] Kamioka M, Takao S, Suzuki T, et al. Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the *Arabidopsis* circadian clock. Plant Cell 2016;28(3):696-711.

[33] Michael AK, Fribourgh JL, Chelliah Y, et al. Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. Proc Natl Acad Sci U S A 2017;114(7):1560-5.

[34] Takahashi J. Transcriptional architecture of the mammalian circadian clock. Nat Rev Genet 2017;18:164-79.

[35] Yan J, Wang H, Liu Y, Shao C. Analysis of gene regulatory networks in the mammalian circadian rhythm. PLoS Comput Biol 2008;4(10):e1000193.

[36] Koike N, Yoo SH, Huang HC, et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science 2012;338(6105):349-54.

[37] Menet JS, Rodriguez J, Abruzzi KC, Rosbash M. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. eLife 2012;1:e00011.

[38] Podkolodnaya OA, Podkolodnaya NN, Podkolodnyy NL. The mammalian circadian clock: Gene regulatory network and computer analysis. Russ J Genet Appl Res 2015;5:354-62.

[39] Salathia N, Edwards K, Millar AJ. QTL for timing: a natural diversity of clock genes. Trends Genet 2002;18(3):115-8.

[40] Konopka RJ, Benzer S. Clock mutants of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 1971;68(9):2112-2116.

[41] Vitaterna MH, King DP, Chang AM, et al. Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. Science 1994;264(5159):719-25.

[42] Bell-Pedersen D, Cassone V, Earnest D, et al. Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet 2005;6:544–56.

[43] Niwa Y, Matsuo T, Onai K, et al. Phase-resetting mechanism of the circadian clock in Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A 2013;110(33):13666-71.

[44] Oike H. Modulation of circadian clocks by nutrients and food factors. Biosci Biotechnol Biochem 2017;81(5):863-70.

[45] Poliner E, Cummings C, Newton L, Farré EM. Identification of circadian rhythms in Nannochloropsis species using bioluminescence reporter lines. Plant J 2019;99(1):112-27.

[46] Miyoshi C, Kim SJ, Ezaki T, et al. Methodology and theoretical basis of forward genetic screening for sleep/wakefulness in mice. Proc Natl Acad Sci U S A 2019;116(32):16062-7.

[47] Swarup K, Alonso-Blanco C, Lynn JR, et al. Natural allelic variation identifies new genes in the Arabidopsis circadian system. Plant J 1999;20(1):67-77.

[48] Anwer MU, Davis SJ. An overview of natural variation studies in the *Arabidopsis thaliana* circadian clock. Semin Cell Dev Biol 2013;24(5):422-9.

[49] Shimomura K, Low-Zeddies SS, King DP, et al. Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice. Genome Res 2001;11(6):959-80.

[50] Kerwin RE, Jimenez-Gomez JM, Fulop D, et al. Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in Arabidopsis. Plant Cell 2011;23(2):471-85.

[51] Jones SE, Lane JM, Wood AR, et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nat Commun 2019;10:343.

[52] Jones SE, Tyrrell J, Wood AR, et al. Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. PLoS Genet 2016;12(8):e1006125.

[53] Jagannath A, Taylor L, Wakaf Z, et al. The genetics of circadian rhythms, sleep and health. Hum Mol Genet 2017;26(R2):R128-R138.

[54] Ferguson A, Lyall LM, Ward J, et al. Genome-wide association study of circadian rhythmicity in 71,500 UK biobank participants and polygenic association with mood instability. EBioMedicine 2018;35:279-87.

[55] Rubin MJ, Brock MT, Davis SJ, Weinig C. QTL underlying circadian clock parameters under seasonally variable field settings in *Arabidopsis thaliana*. G3 (Bethesda) 2019;9(4):1131-9.

[56] Liu T, Liu XL, Chen YM, Wu RL. A unifying differential equation model for functional genetic mapping of circadian rhythms. Theor Biol Med Model 2007;4:5.

[57] Li N, McMurry T, Berg A, et al. Functional clustering of periodic transcriptional profiles through ARMA(p,q). PLoS ONE 2010;5(4):e9894.

[58] Kim B-R, Littell RC, Wu RL. Clustering the periodic pattern of gene expression using Fourier series approximations. Curr Genom 2006;7:197–203.

[59] Kim B-R, Zhang L, Berg A, Fan J, Wu RL. A computational approach to the functional clustering of periodic gene expression profiles. Genetics 2008;180:821–34.

[60] Kim BR, McMurry T, Zhao W, Berg A, Wu RL. Wavelet-based functional clustering for high-dimensional dynamic gene expression patterns. J Comp Biol 2010;17:1067–80.

[61] Fu GF, Luo J, Berg A, et al. A dynamic model for functional mapping of biological rhythms. J Biol Dyn 2010;4:1–10.

[62] Fu GF, Wang Z, Li JH, Wu RL. A mathematical framework for functional mapping of complex systems using delay differential equations. J Theor Biol 2011;289:206–16.

[63] Wei K, Wang Q, Gan JW, et al. Mapping genes for drug chronotherapy. Drug Discov

Today 2018;23:1883-8.

[64] Wu RL, Lin M. Functional mapping – how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 2006;7:229–37.

[65] Sun LD, Wu RL. Mapping complex traits as a dynamic system. Phys Life Rev 2015;13:155-85.

[66] Li Z, Sillanpää MJ. Dynamic quantitative trait locus analysis of plant phenomic data. Trends Plant Sci 2015;20:822-833.

[67] Vijesh N, Chakrabarti SK, Sreekumar J. Modeling of gene regulatory networks: a review. J Biomed Sci Eng 2013;6:223-231.

[68] Wang YX, Huang H. Review on statistical methods for gene network reconstruction using expression data. J Theor Biol 2014;362:53-61.

[69] Huynh-Thu V, Sanguinetti G. Gene regulatory network inference: an introductory survey. Methods Mol Biol 2019;1883:1-23.

[70] Yaghoobi H, Haghipour S, Hamzeiy H, Asadi-Khiavi M. A review of modeling techniques for genetic regulatory networks. J Med Signals Sens 2012;2(1):61-70.

[71] Jiang LB, Shi H, Sang M, et al. A computational model for inferring QTL control networks underlying developmental covariation. Front Plant Sci 2019;10:1557.

[72] Zhang EE, Kay SA. Clocks not winding down: unravelling circadian networks. Nat Rev

Mol Cell Biol 2010;11:764–76.

[73] von Neumann JV, Morgenstern O. Theory of Games and Economic Behavior. Princeton University Press, 1944.

[74] Nash JF. Equilibrium points in N-person games. Proc Natl Acad Sci U S A 1950;36:48-9.

[75] Smith JM, Price GR. Logic of animal conflict. Nature 1973;246(5427):15-8.

[76] Fisher RA. The Genetic Theory of Natural Selection, Oxford, Clarendon Press, 1930.

[77] Ma CX, Casella G, Wu RL. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. Genetics 2002;161:1751–62.

[78] Wu RL, Ma CX, Lin M et al. A general framework for analyzing the genetic architecture of developmental characteristics. Genetics 2004;166:1541-51.

[79] Wu C, Li G, Zhu J, Cui YH. Functional mapping of dynamic traits with robust *t*-distribution. PLoS ONE 2011;6(9):e24902.

[80] Liu G, Li M, Wen J, Du Y, Zhang Y-M. Functional mapping of quantitative trait loci associated with rice tillering. Mol Genet Genom 2010;284:263-71.

[81] Li Z, Henrik R. Hallingbäck S, et al. Functional multi-locus QTL mapping of temporal trends in scots pine wood traits. G3 2014;3:2365-79.

[82] Kwak IY, Moore CR, Spalding EP, Broman KW. Mapping quantitative trait loci underlying function-valued traits using functional principal component analysis and multi-trait mapping. G3 (Bethesda) 2016;6:79–86.

[84] Lyra DH, Virlet N, Sadeghi-Tehran P, et al. Functional QTL mapping and genomic prediction of canopy height in wheat measured using a robotic field phenotyping platform. J Exp Bot 2020;71(6):1885-98.

[85] Zhao W, Chen YQ, Casella G, Cheverud JM, Wu RL. A non-stationary model for functional mapping of complex traits. Bioinformatics 2005;21:2469–77.

[86] Zhao W, Hou W, Littell RC, Wu RL. Structured antedependence models for functional mapping of multivariate longitudinal traits. Stat Methods Mol Genet Biol 2005;4:Issue 1.

[87] Li N, McMurry T, Berg A, et al. Functional clustering of periodic transcriptional profiles through ARMA(p,q). PLoS ONE 2010;5(4):e9894.

[88] Yap J, Fan J, Wu RL. Nonparametric modeling of covariance structure in functional mapping of quantitative trait loci. Biometrics 2009;65:1068–77.

[89] Das K, Li JH, Wang Z et al. A dynamic model for genome-wide association studies. Hum Genet 2011;129:629–39.

[90] Yang R, Xu S. Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. Genetics 2007;176:1169–85.

[91] Klerman EB, St Hilaire MA. On mathematical modeling of circadian rhythms,

performance, and alertness. J Biol Rhyth 2007;22(2):91–102.

[92] Asgari-Targhi A, Klerman EB. Mathematical modeling of circadian rhythms. Wiley Interdiscip Rev Syst Biol Med 2019;11(2):e1439.

[93] Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 1998;9(12):3273-97.

[94] Dale JK, Maroto M, Dequeant ML, et al. Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. Nature 2003;421:275–8.

[95] Konopka T, Rooman M. Gene expression model (in)validation by Fourier analysis. BMC Syst Biol. 2010;4:123.

[96] Yang R, Su Z. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics 2010;26:i168–i174.

[97] Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. Cell 2017;169:1177-1186.

[98] Garlaschelli D, Caldarelli G, Pietronero L. Universal scaling relations in food webs. Nature 2003;423:165–8.

[99] Liu Y-Y, Slotine J-J, Barabási A-L. Controllability of complex networks. Nature 2011;423:167–73.

[100] Nacher JC, Akutsu T. Structural controllability of unidirectional bipartite networks. Sci Rep 2013;3:1647.

[101] Suweis S, Simini F, Banavar JR, Maritan A. Emergence of structural and dynamical properties of ecological mutualistic networks. Nature 2013;500:449–52.

[102] Grilli J, Adorisio M, Suweis S, et al. Feasibility and coexistence of large ecological communities. Nat Commun 2017;8:14389.

[103] Busiello DM, Suweis S, Hidalgo J, et al. Explorability and the origin of network sparsity in living systems. Sci Rep 2017;7:12323.

[104] May RM. Will a large complex system be stable? Nature 1972;238:413-4.

[105] Allesina S, Tang S. Stability criteria for complex ecosystems. Nature 2012;483:205–8.

[106] Tibshirani RJ. Regression shrinkage and selection via the Lasso. J Roy Stat Sco B 1996;58:267–288.

[107] Zou, H. Hastie T. Regularization and variable selection via the elastic net. J Roy Stat Soc B 2005;67: 301–320.

[108] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J Roy Stat Sco B 2006;68:49–67.

[109] Wang H, Leng C. A note on the adaptive group Lasso. Comput Stat Data Analy 2008;52:5277–5286.

[110] Van Voorn GAK, Kooi BW. Combining bifurcation and sensitivity analysis for ecological models. Eur Phys J Spec Top 2017;226:2101–18.

[111] Donzé A, Clermont G, Langmead CJ. Parameter synthesis in nonlinear dynamical systems: application to systems biology. J Comput Biol 2010;17(3):325-36.

[112] Miao H, Wu H, Xue H. Generalized ordinary differential equation models. J Am Stat Assoc 2014;109(508):1672-82.

[113] Ramsay JO, Hooker G, Campbell D, et al. Parameter estimation for differential equations: a generalized smoothing approach. J Roy Stat Soc Ser B-Stat Method 2007;69:741-70.

[114] Liang H, Wu H. Parameter estimation for differential equation models using a framework of measurement error in regression models. J Am Stat Assoc 2008;103(484):1570-83.

[115] Cao J, Huang JZ, Wu H. Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. J Comput Graph Stat 2012;21(1):42-56.

[116] Brunel NJB, Clairon Q, D'alche-Buc F. Parametric estimation of ordinary differential equations with orthogonality conditions. J Am Stat Assoc 2014;109(505):173-85.

[117] Li Y, Zhu J, Wang N. Regularized semiparametric estimation for ordinary differential equations. Technometrics 2015;57(3):341-50.

[118] Chen SZ, Shojaie A, Witten DM. Network reconstruction from high-dimensional

ordinary differential equations. J Am Stat Assoc 2017;112(520):1697-707.

[119] Bateson, W. *Mendel's Principles of Heredity* (Cambridge, Cambridge University Press, 1909)

[120] Cordell, H. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics* **11**, 2463–2468 (2002)

[121] Michailidis G, d'Alché-Buc F. Autoregressive models for gene regulatory network inference: sparsity, stability and causality issues. Math Biosci 2013;246(2):326-334.

[122] Zavlanos MM, Julius AA, Boyd SP, Pappas GJ: Inferring stable genetic networks from steady-state data. Automatica. 2011, 47: 1113-1122.

[123] Larvie JE, Sefidmazgi MG, Homaifar A, et al. Stable gene regulatory network modeling from steady-state data. Bioengineering (Basel) 2016;3(2):12.

[124] Dunbar RIM. Neocortex size as a constraint on group size in primates. J Hum Evol 1993;22:469–493.

[125] Callebaut W, Rasskin-Gutman D. Understanding the Development of Evolution of Natural Complex Systems Cambridge, CT: MIT Press, 2005.

[126] Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. Nat Rev Genet 2007;8(12):921-31.

[127] Newman ME. Modularity and community structure in networks. Proc Natl Acad Sci U S

A 2006;103(23):8577-8582.

[128] Mehrle D, Strosser A, Harkin A. Walk-modularity and community structure in networks. Netw Sci 2015;3(3):348–60.

[129] Serban M. Exploring modularity in biological networks. Phil Trans R Soc 2020; 375:20190316.

[130] Yoon J, Blumer A, Lee K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. Bioinformatics 2006;22(24):3106-8.

[131] Kaltenbach HM, Stelling J. Modular analysis of biological networks. Adv Exp Med Biol 2012;736:3-17.

[132] Tosh CR, McNally L. The relative efficiency of modular and non-modular networks of different size. Proc Biol Sci 2015;282(1802):20142568.

[133] Al-Anzi B, Gerges S, Olsman N, et al. Modeling and analysis of modular structure in diverse biological networks. J Theor Biol 2017;422:18-30.

[134] Didier G, Valdeolivas A, Baudot A. Identifying communities from multiplex biological networks by randomized optimization of modularity. F1000Res 2018;7:1042.

[135] Wang YQ, Xu M, Wang Z, et al. How to cluster gene expression dynamics in response to environmental signals. Brief Bioinform 2012;13:162–74.

[136] Li JZ, Bunney BG, Meng F, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc Natl Acad Sci U S A 2013;110(24):9950-5.

[137] Smith LM, Motta FC, Chopra G, et al. An intrinsic oscillator drives the blood stage cycle of the malaria parasite *Plasmodium falciparum*. Science 2020;368(6492):754-9.

[138] Rijo-Ferreira F, Acosta-Rodriguez VA, Abel JH, et al. The malaria parasite has an intrinsic clock. Science 2020;368(6492):746-53.

[139] Walker WH, Walton JC, DeVries AC, et al. Circadian rhythm disruption and mental health. Transl Psychiatry 2020;10:28.

[140] Chen CX, Jiang LB, Fu GF, et al. An omnidirectional visualization model of personalized gene regulatory networks. npj: Syst Biol Appl 2019;5:38.

[141] Sun, L.; Jiang, L.; Grant, C.N.; Wang, H.-G.; Gragnoli, C.; Liu, Z.; Wu, R. Computational Identification of Gene Networks as a Biomarker of Neuroblastoma Risk. Cancers 2020*,* 12*,* 2086.

[142] Karczewski, K., Snyder, M. Integrative omics for health and disease. Nat Rev Genet 2018;19:299–310.

[143] Ye Y, Zhang Z, Liu Y, et al. Multi-omics perspective of quantitative trait loci in precision medicine. Trends Genet 2020;36(5):318-36.

[144] Kang M, Gao J. Integration of multi-omics data for expression quantitative trait loci (eQTL) analysis and eQTL epistasis. Methods Mol Biol 2020;2082:157-71.

[145] Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. PLoS Comp Biol 2008;4:e1000117.

[146] Freeman LC. Centrality in social networks conceptual clarification. Social Networks 1978;1:215-39.

[147] Brandes U. A faster algorithm for betweenness centrality. J Math Sociol 2001;25:163-77.

[148] Hage P, Harary F. Eccentricity and centrality in networks. Social Networks 1995;17:57-63.

[149] Bonacich P. Some unique properties of eigenvector centrality. Social Networks. 2007;29:555-64.

[150] Page L. The PageRank citation ranking: Bringing order to the web. Stanford Digital Libraries Working Paper 1998;9:1-14.

[151] Li JZ, Bunney BG, Meng F, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc Natl Acad Sci U S A 2013;110(24):9950-9955.

[152] Gottlieb DJ, O'Connor GT, Wilk JB. Genome-wide association of sleep and circadian phenotypes. BMC Med Genet 2007;8:S9.

[153] Harbison ST, Kumar S, Huang W, et al. Genome-wide association study of circadian behavior in Drosophila melanogaster. Behav Genet 2019;49:60–82.

[154] Jiang LB, Liu JY, Zhu XL, et al. 2HiGWAS: A unifying high-dimensional platform to infer the global genetic architecture of trait development. Brief Bioinform 2015;16:905-11.

[155] Zimmerman DL, Núñez-Antón VA. Antedependence Models for Longitudinal Data. Chapman & Hall/CRC, Boca Raton, FL.

Box 1: Statistical procedure for SEGN inference

Let $\mathbf{y}_j = (y_j(t_1), \ldots, y_j(t_T))$ denote a vector of estimated time-varying genetic variance values explained by SNP j ($j = 1, \ldots, m$). Given these data, the likelihood function of model parameters $\boldsymbol{\phi} = (\boldsymbol{\mu}, \Sigma) \in \boldsymbol{\Phi}$ is written as

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = f(\mathbf{y}_1, \ldots, \mathbf{y}_m: \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{(S1)}$$

where $f(\cdot)$ is the m-variate t_T-dimensional longitudinal multivariate normal distribution with mean vector $\mathbf{g} = (\mathbf{g}_1, \ldots, \mathbf{g}_m)$ with $\mathbf{g}_i = (g_i(t_1), \ldots, g_i(t_T))$ and the covariance matrix of $e_i(t)$, $\boldsymbol{\Sigma}$. As described below, we will model mean-covariance structures.

The time-varying genetic variance of each SNP is modeled by a system of ODEs described by equation (1) containing the independent and dependent components. Each component is fitted by a nonparametric approach, such as B-splines, regression B-spline, penalized B-spline, local polynomials and Legendre orthogonal polynomials (LOP). Because of its advantage in orthogonality and efficient convergence, LOP has been used to model the curves of any complex form using sparse data in quantitative genetic studies [89,154]. The LOP is a solution of the Legendre differential equation, $(1 - v^2)(d^2u/dv^2) - 2v(du/\underline{dv}) + r(r + 1)u = 0$. Let $\mathbf{P}_{jR}(t) = (P_{j1}(t), \ldots, P_{jR}(t))$ denote a vector of LOP including the first R orders for SNP j at time t, and $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jR})$ denote a vector of basis values of time-invariant independent genetic variance of SNP j. Then, the independent genetic variance of SNP j is expressed as

$$g_i(t) = \mathbf{P}_{iR}^T(t)\boldsymbol{\alpha}_i \quad \text{(S2)}$$

An optimal order of LOP may be SNP-specific; i.e., each SNP may have a different LOP order. The optimal order for each SNP j can be determined via an information criterion, such as AIC or BIC. Similarly, we use LOP to model time-varying dependent genetic variances. Time-invariant

independent and dependent genetic variances, arrayed in Θ_j and $\Theta_{jj'}$, respectively, are the unknown ODE parameters to be estimated (see equation (1)).

Since the residual covariance matrix $\mathbf{\Sigma}$ contains an autocorrelative structure, its structural modeling using a parsimonious time-series approach can increase the precision of ODE parameter estimation and computational efficiency. Given its power to structure the longitudinal covariance of quantitative traits in genetic mapping studies [85,86], the structured antedependence (SAD) model, developed by Zimmerman and Núñez-Antón [155], is implemented into likelihood (S1). The SAD assumes that residual errors at time *t* are not only composed of innovation errors specifically produced at this time point, but also contain a proportion of the residual errors from the preceding time points. The size of this proportion, i.e., the degree of antedependence (ρ), decays with time lag. The first-order SAD (SAD (1)) only considers the dependence of errors at the immediate time point. The innovation error for SNP *j* is iid with mean zero and variance δ_j^2 which is assumed to be constant across time points. Two parameters, ρ_j and δ_j^2, can well model the structure of $\mathbf{\Sigma}$.

With joint ODE and SAD(1) modeling, the parameters involved in likelihood (S1) are re-written as $\boldsymbol{\phi} = \left\{\Theta_j, \Theta_{jj'}, \rho_j,\ \delta_j^2\ \right\}_{j\neq j'=1}^{m}$. We can obtain the optimal solution of these parameters by maximizing the likelihood (S1), expressed as

$$\hat{\boldsymbol{\phi}} \in \left\{\arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \mathcal{L}(\boldsymbol{\mu}, \mathbf{\Sigma})\right\}. \tag{S3}$$

We implement a hybrid algorithm of the fourth-order Runge-Kutta (RK4) algorithm and simplex approach to solve the likelihood incorporated by a system of LOP-transformed ODEs (equation (1)). Intuitively, this maximization that makes the data most probable implies an optimal topological structure and organization by which genes interact with each other to maximize the joint expression of all genes. This solution is therefore consistent with the basic principle of evolutionary game theory [75].

Table 1 Qualitative definition of epistasis and its quantitative characterization by the SEGN model.

No	Qualitative definition	Quantitative description		
		$P_{jj'}(t)$	Relation	$P_{j'j}(t)$
1	Symmetric positive epistasis	+	=	+
2	Asymmetric positive epistasis	+	≠	+
3	Directional positive epistasis toward j	+	>	0
4	Directional positive epistasis toward j'	0	<	+
5	Altruism toward j or exploitation by j	+		-
6	Altruism toward j' or exploitation by j'	-		+
7	Symmetric negative epistasis	-	=	-
8	Asymmetric negative epistasis	-	≠	-
9	Directional negative epistasis toward j	-		0
10	Directional negative epistasis toward j'	0		-
11	Coexistence	0		0

Note: $P_{jj'}(t)$ and $P_{j'j}(t)$ are the dependent genetic variances of SNP j by SNP j' and SNP j' by SNP j, respectively.

Figure Legends

Figure 1 Simulated examples showing functional mapping of circadian rhythms as a first step of SEGN reconstruction. **A** and **B**: Estimated rhythmic curves of three genotypes AA, Aa, and aa at each of two randomly chosen SNPs, showing genotypic differences in rhythmic features including phase, period, and amplitude. **C**: Rhythmic pattern of genetic variance explained by each SNP. **D**: Plot of the genetic variance of SNP 2 against SNP 1 over rhythmic cycles.

Figure 2 Real-time genetic networks reconstructed from 10 simulated SNPs. **A**: Genetic variance of each SNP changes rhythmically over time in a different way. **B**: Instantaneous SEGNs at times 15, 30, and 60 during rhythmic cycles, showing temporal changes in topological structure and organization. Circles denote SNPs as nodes, whose size is proportional to the magnitude of genetic variance explained by a given SNP. Red and blue arrowed lines denote promotion and inhibition from one SNP to the next, respectively, with strength proportional to the thickness of lines.

Figure 3 A diagram of tridimensional SEGN across multiple layers. **Top tier:** Coarse-grained genetic networks among five modules detected from a complete set of SNPs for a mapping study. **Second tier:** Some modules are furthered classified into submodules to form genetic networks at higher resolution. **Third tier:** Some submodules need to be decomposed into different sub-submodules, forming fine-grained genetic networks at the individual SNP level. Red and blue arrows denote promotion and inhibition from one SNP to the next, respectively.

Figure 4 Fourier series-based functional clustering classifies 00,000 simulated SNPs into different modules. On the basis of AIC, the optimal number of modules is determined to be 22 under the third order Fourier series approximation. Ten modules were chosen to show how different modules are expressed rhythmically in various manners. Thin curves in each plot

represent rhythmic changes of the genetic variance of individual SNPs and the thick curve is the mean fitting curve of all SNPs within a module.

Figure 5 A simulated example showing how a QTL controls GRNs. A total of 10 genes are assumed to regulate rhythmic activities through their interaction networks. A SNP is regarded as a significant QTL if its genotypes have different gene networks. **A**: Rhythmic expression curves of 10 genes varying among three QTL genotype AA, Aa, and aa. **B**: Gene networks reconstructed with 10 genes, individually for three different genotypes. Red and blue arrows denote promotion and inhibition from one gene to the next, respectively, with strength proportional to the thickness of lines.

Figure 6 A hypothetical clock demonstrating how epistatic networks intertwine with pleiotropic networks during rhythmic cycles. Assume that three genes G1, G2, and G3 regulate rhythmic phenotypes P1 and P2 and this process is controlled by three SNPs. Under the control of each SNP, genes are causally linked with phenotypes through a pleiotropic network (shown on yellow ellipses). For the same gene or phenotype, different SNPs control their expression through epistatic networks (indicated by a dotted triangle). Pleiotropic networks differ in topological structure among SNPs, indicating that different SNPs affect causal gene-phenotypic networks in different ways. SNPs affect a gene or phenotype through different networks (green for G1, purple for G2, green-blue for G3, blue for P1 and black for P2), suggesting that each gene or phenotype is encoded by a different genetic system. Red and blue arrows denote promotion and inhibition from one gene to the next, respectively.

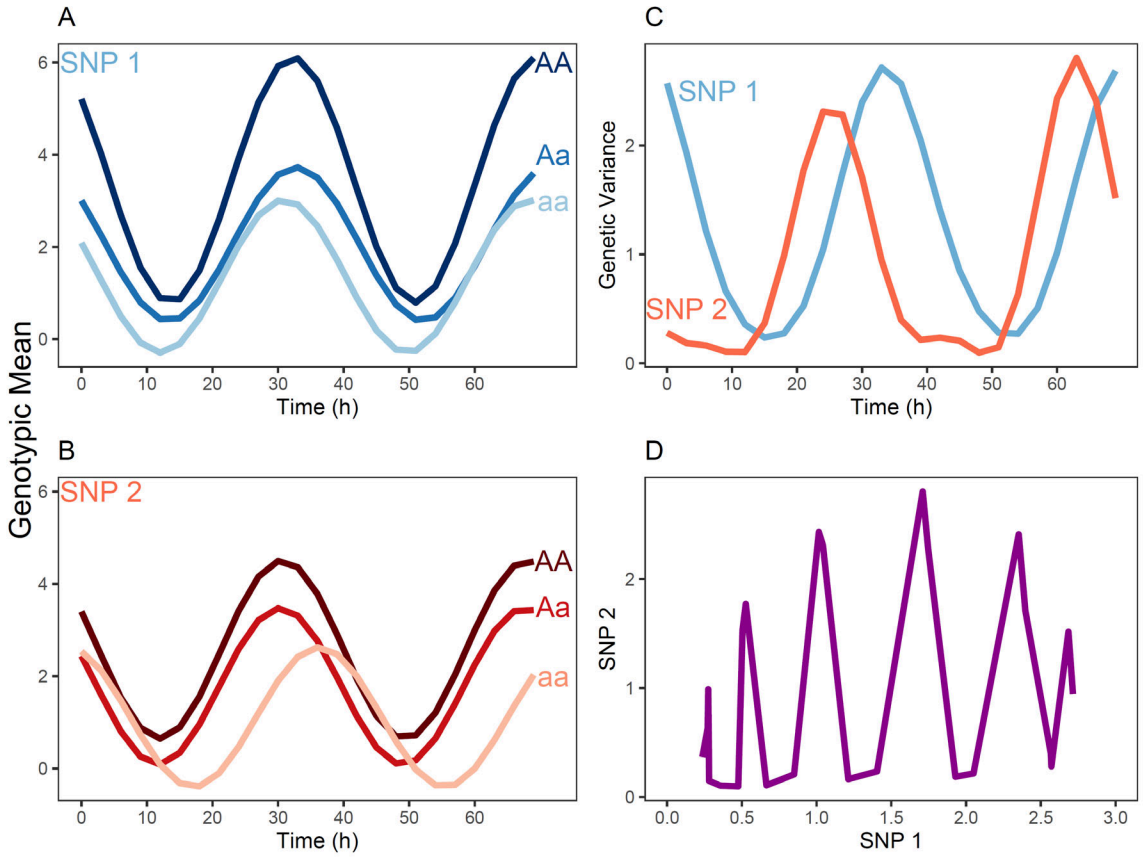
A
SNP 1
AA
Aa
aa
Genotypic Mean
Time (h)
B
SNP 2
AA
Aa
aa
Time (h)
C
SNP 1
SNP 2
Genetic Variance
Time (h)
D
SNP 2
SNP 1

Figure 1

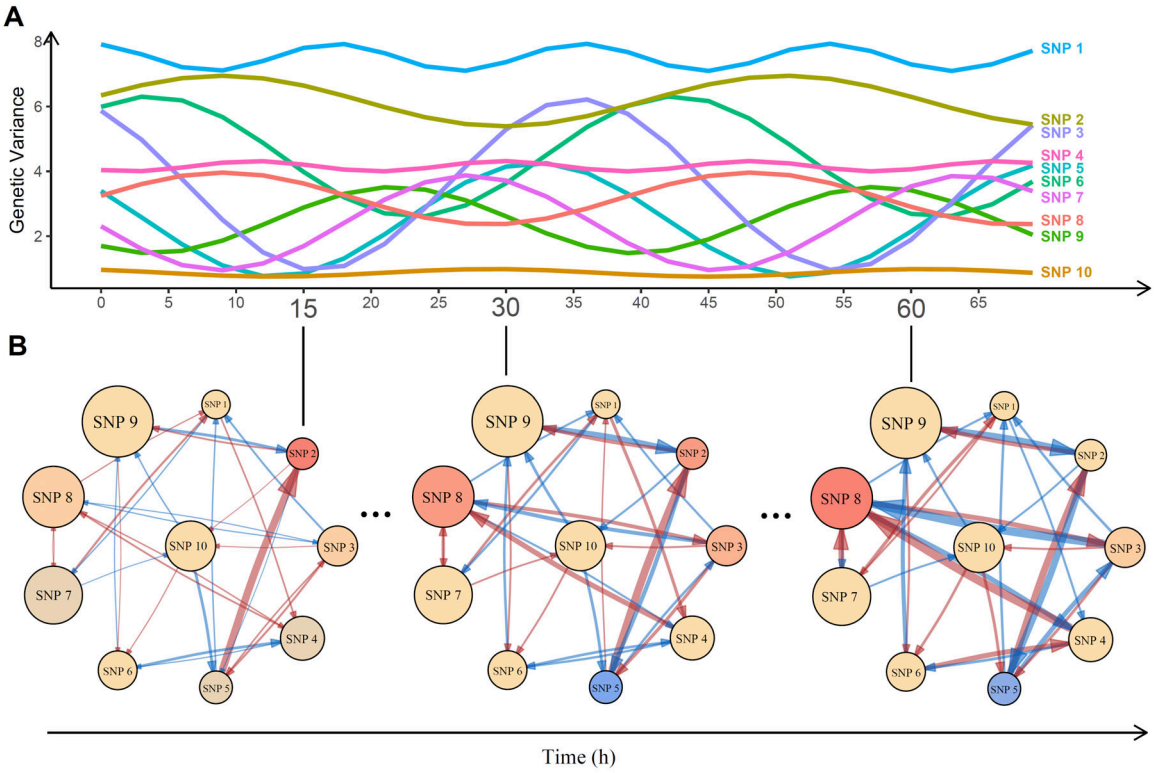
A
Genetic Variance
SNP 1
SNP 2
SNP 3
SNP 4
SNP 5
SNP 6
SNP 7
SNP 8
SNP 9
SNP 10
B
Time (h)

Figure 2

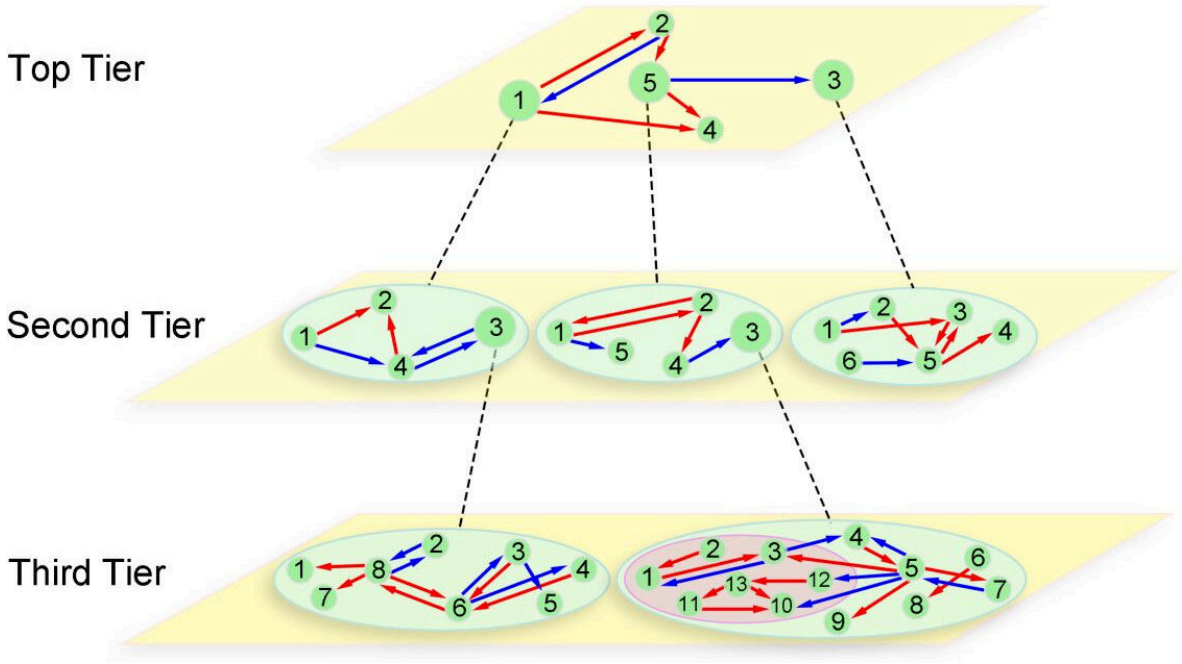

Figure 3

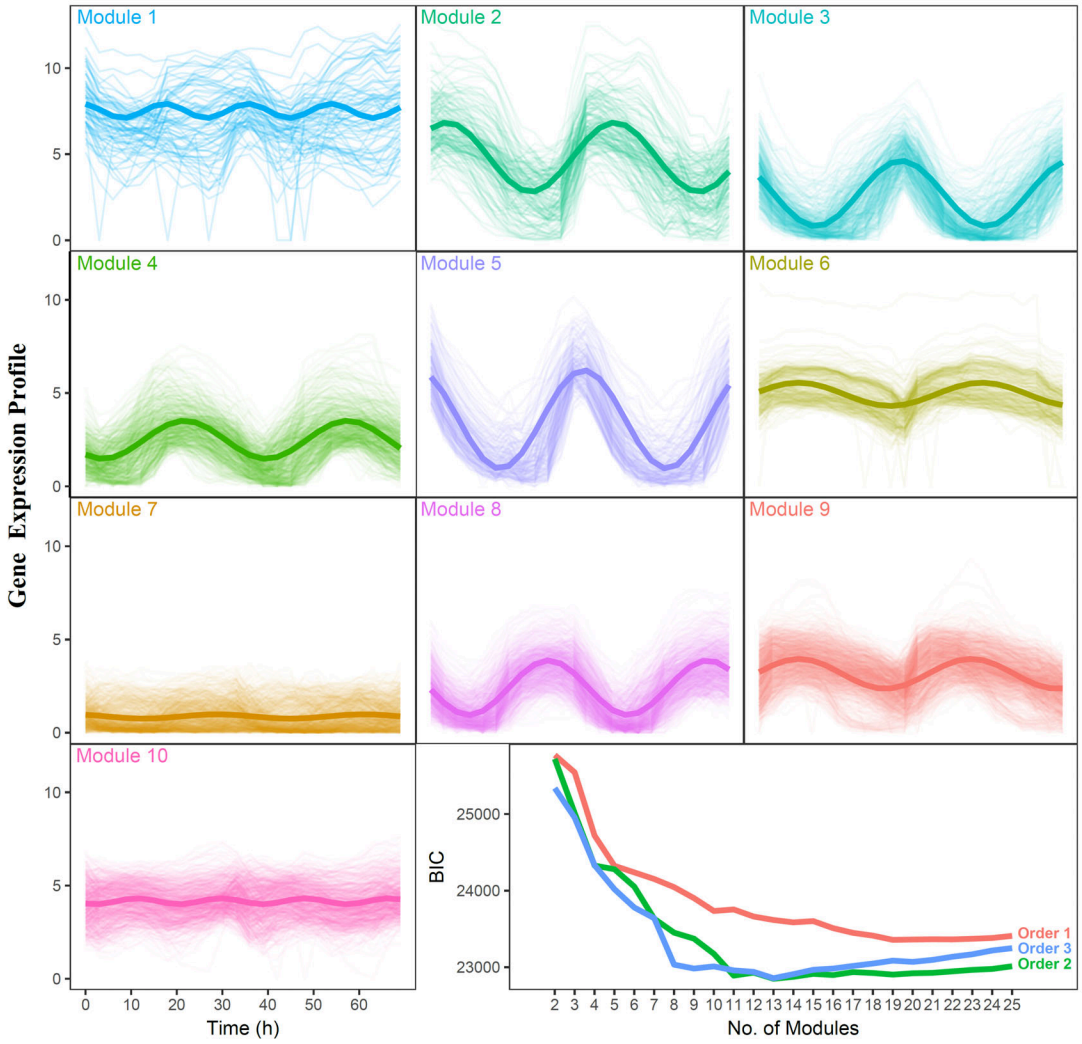
Module 1
Module 2
Module 3
Module 4
Module 5
Module 6
Module 7
Module 8
Module 9
Module 10
Gene Expression Profile
Time (h)
BIC
No. of Modules
Order 1
Order 3
Order 2

Figure 4

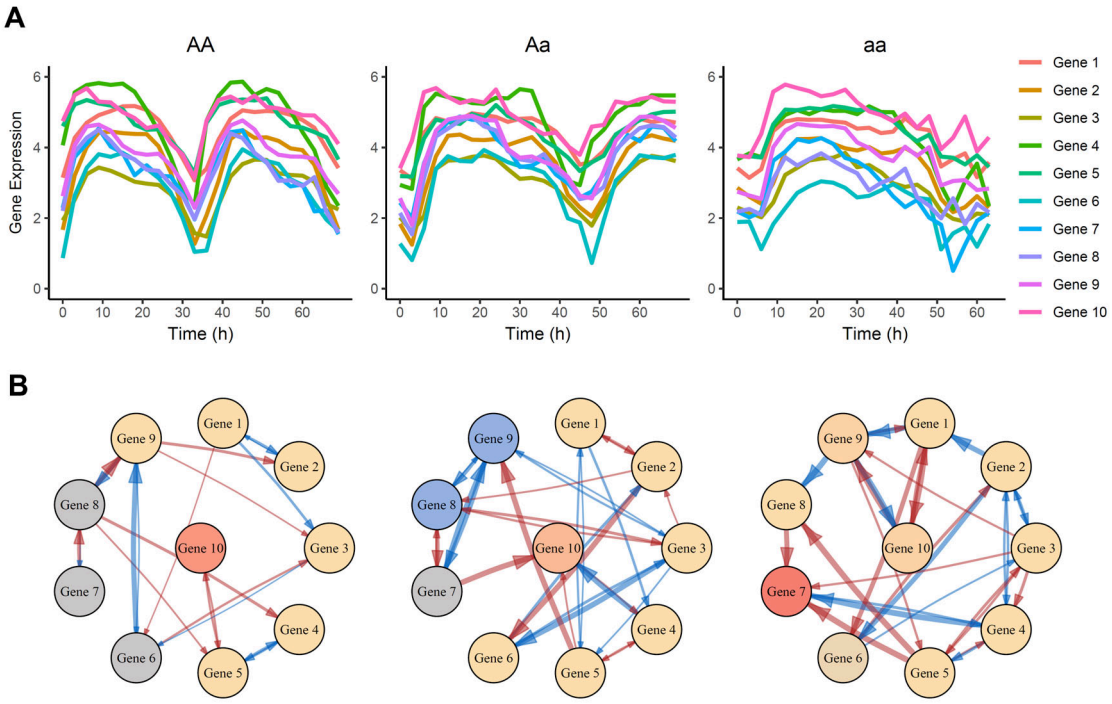
A
AA
Aa
aa
Gene Expression
Time (h)
Gene 1
Gene 2
Gene 3
Gene 4
Gene 5
Gene 6
Gene 7
Gene 8
Gene 9
Gene 10
B

Figure 5

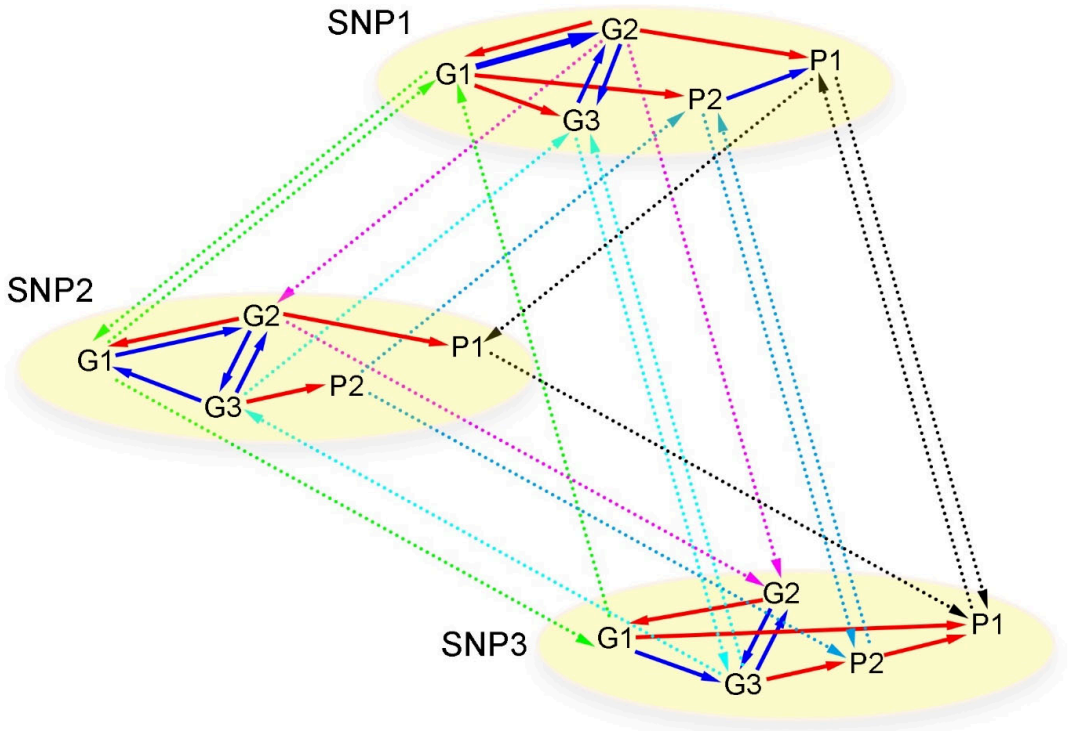

Figure 6