Title

Modeling transcriptional regulation of the cell cycle using a novel cybernetic-inspired approach.

Journal

Biophysical Journal, 123(2)

Authors

Raja, Rubesh

Khanum, Sana

Aboulmouna, Lina

et al.

Publication Date

2024-01-16

DOI

10.1016/j.bpj.2023.12.010

Peer reviewed

Biophysical Journal
Article

Biophysical Society

Modeling transcriptional regulation of the cell cycle using a novel cybernetic-inspired approach

Rubesh Raja,[1] Sana Khanum,[1] Lina Aboulmouna,[2] Mano R. Maurya,[2] Shakti Gupta,[2] Shankar Subramaniam,[2,3,*] and Doraiswami Ramkrishna[1,*]

[1]The Davidson School of Chemical Engineering, Purdue University, West Lafayette, Indiana; [2]Department of Bioengineering, University of California San Diego, La Jolla, California; and [3]Departments of Computer Science and Engineering, Cellular and Molecular Medicine, San Diego Supercomputer Center, and the Graduate Program in Bioinformatics and Systems Biology, University of California San Diego, La Jolla, California

ABSTRACT Quantitative understanding of cellular processes, such as cell cycle and differentiation, is impeded by various forms of complexity ranging from myriad molecular players and their multilevel regulatory interactions, cellular evolution with multiple intermediate stages, lack of elucidation of cause-effect relationships among the many system players, and the computational complexity associated with the profusion of variables and parameters. In this paper, we present a modeling framework based on the cybernetic concept that biological regulation is inspired by objectives embedding rational strategies for dimension reduction, process stage specification through the system dynamics, and innovative causal association of regulatory events with the ability to predict the evolution of the dynamical system. The elementary step of the modeling strategy involves stage-specific objective functions that are computationally determined from experiments, augmented with dynamical network computations involving endpoint objective functions, mutual information, change-point detection, and maximal clique centrality. We demonstrate the power of the method through application to the mammalian cell cycle, which involves thousands of biomolecules engaged in signaling, transcription, and regulation. Starting with a fine-grained transcriptional description obtained from RNA sequencing measurements, we develop an initial model, which is then dynamically modeled using the cybernetic-inspired method, based on the strategies described above. The cybernetic-inspired method is able to distill the most significant interactions from a multitude of possibilities. In addition to capturing the complexity of regulatory processes in a mechanistically causal and stage-specific manner, we identify the functional network modules, including novel cell cycle stages. Our model is able to predict future cell cycles consistent with experimental measurements. We posit that this innovative framework has the promise to extend to the dynamics of other biological processes, with a potential to provide novel mechanistic insights.

SIGNIFICANCE Cellular processes such as cell cycle are complex, involving multiple players interacting at multiple levels, and explicit modeling of such systems is challenging. The availability of longitudinal measurements provides an opportunity to reverse engineer for novel regulatory models. We develop a novel framework, inspired by using a goal-oriented cybernetic approach, to implicitly model transcriptional regulation by constraining the system to inferred temporal goals. A preliminary causal network based on information theory is used as a starting point, and our framework is used to distill the network to temporally based networks containing essential molecular players. The strength of this approach is its ability to dynamically model the RNA temporal measurements. The approach developed paves the way for inferring regulatory processes in many complex cellular processes.

INTRODUCTION

Cellular processes involve a complex network of molecular interactions associated with response to external stimuli, signal transduction, chromatin modifications, transcriptional regulation, and a host of regulatory mechanisms. Painstaking biochemical analyses of these complex regulatory processes have provided some insights but are largely incomplete. Further, there are few data on the kinetics of regulatory processes even at a coarse-grained level, making quantitative modeling of the cellular processes difficult (1–3). The ability to infer all the regulatory players and processes is out of the scope of currently available experiment methods and it is

Submitted February 20, 2023, and accepted for publication December 12, 2023.

*Correspondence: shsubramaniam@ucsd.edu or ramkrish@purdue.edu

Editor: Jianhua Xing.

https://doi.org/10.1016/j.bpj.2023.12.010

essential to decipher an approach that will be able to account for the regulatory processes in an implicit manner. Longitudinal time-series data from high-throughput measurements provide details about potential regulation but incorporating such vast data into mathematical models has proved to be a challenge. Previously established modeling approaches, including steady-state stoichiometric models (4), kinetic models (1,5,6), dynamic flux balance analysis models (7), and discrete Boolean models (8), have faced these challenges. Constraint-based approaches such as flux balance analysis (4) and network decomposition approaches such as elementary modes (9) can aid in modeling regulatory mechanisms. Most of these represent static or steady-state situations and some with pseudo-steady-state perspectives that are very useful for reactions with disparate rates. However, the dynamic complexity of eukaryotic cell behavior makes pseudo-steady-state assumptions inapplicable.

The cybernetic method (10), which inherits its name from optimizing models with goals, where regulatory processes are inferred implicitly through parameters, has been successful in dynamically modeling bacterial metabolism where the objective is cell growth optimality (10–13). Recently, we extended this approach to modeling macrophage lipid metabolism with the goal of maximal expression and activation of inflammatory cytokine tumor necrosis factor alpha (14,15). With this single objective function, we were able to incorporate regulatory mechanisms involved in prostaglandin metabolism. In all these examples, only one stage with a single well-defined objective function is assumed, making it difficult to extend it to multi-stage processes with a distinct objective function associated with each stage.

The mammalian cell cycle is an exemplary cellular process (16) where cells replicate through a complex set of events across multiple stages and it illustrates the complexity associated with such biological processes, namely stage-specific regulation, and stage-specific phenotype endpoints, i.e., biological objectives, for each stage. It is important to understand that the stages mentioned here represent periods of time within which the cell is expected to have a distinct objective based on their phenotypic endpoints. Although stages are not the same as cell cycle phases, they can coincide with each other. The cell cycle goes through different experimentally defined phases (G1, S, G2, and M), where, during the G1 phase, the cell increases its cellular contents and grows, followed by the S phase where the chromosomes are duplicated, with growth and preparation for the mitosis in the G2 phase followed by mitosis (cell division) in the M phase before returning to the G1 phase (3,17,18). Extant dynamical modeling approaches are not capable of addressing multiple stages using distinct and multiple objective functions (18–20).

We report in this manuscript a novel framework featuring a cybernetic-inspired method (CIM) that accounts for stage-specific modeling with distinct objectives for each stage constituting an implicit representation of regulatory processes during that stage. In the cell cycle modeling using a set of fine-grained time-series transcriptomic measurements (3), we develop stage-specific models with specific objectives and model their dynamical behavior controlled by explicitly unspecified regulatory processes. It is appropriate that we model the transcriptomic data as the stages in the cell cycle are transcriptionally driven. The biological objective of such a system can be stated as the optimization of cell cycle where the objective is different at each stage depending on the specific functional differences across each stage. During each stage, the objective is mathematically represented to maximize the weighted production rates of all the transcripts where the weights are expected to implicitly depend on the functional importance of the genes during that stage. The difference in objective is contained in the different weights that are used in different stages. The weights are chosen to fit experimental data, a strategy that remains true to the cybernetic mechanism representing regulatory processes during that stage.

For most biological systems such as the cell cycle, there is limited availability of information relating to the numerous interactions among the myriad components and the amount of data needed to estimate the model parameters (21,22). The longitudinal RNA sequencing (RNA-seq) data in the form of time series is a great resource for developing and validating our CIM framework for dynamical modeling. The temporal molecular data have significant correlations both at the same time point and at distinct time points reflecting cause-effect relationships. To infer this notion of correlation and crosstalk between different RNA transcripts, we explored the power of information-theoretic approaches (23) by relating the components that have a higher degree of mutual information. Mutual information, an information-theoretic approach, detects nonlinear correlations between datasets and can be used to formulate the causal interactions (24,25). Moreover, causality is fundamental to the cybernetic approach in order that the system tweaks the relevant variables to realize objectives. Specifically, we used the time-delayed mutual information (TDMI) method to identify the causal relations and temporal correlations between any two RNA transcripts within the network, and, by setting a threshold, we selected a preliminary network to model with our CIM framework (Fig. 1). Mutual information thus plays a very key enabling role in the success of the model framework.

Further, the stages with distinct objective functions during cell cycle are not pre-defined. Although it is natural to match the stages with experimentally derived temporal regimes such as cell cycle phases, each of these phases may subsume other undefined stages. To correctly identify stages, we used the CIM to simultaneously work as change-point detector while fitting the RNA-seq data. This approach provided key descriptions of regulatory interactions, especially in identifying the key stage-specific genes. Since matching observations of system variables with their model counterparts is the manner in which the weights (and other model parameters) are determined, they represent the system choice arising from the built-in biological objectives.

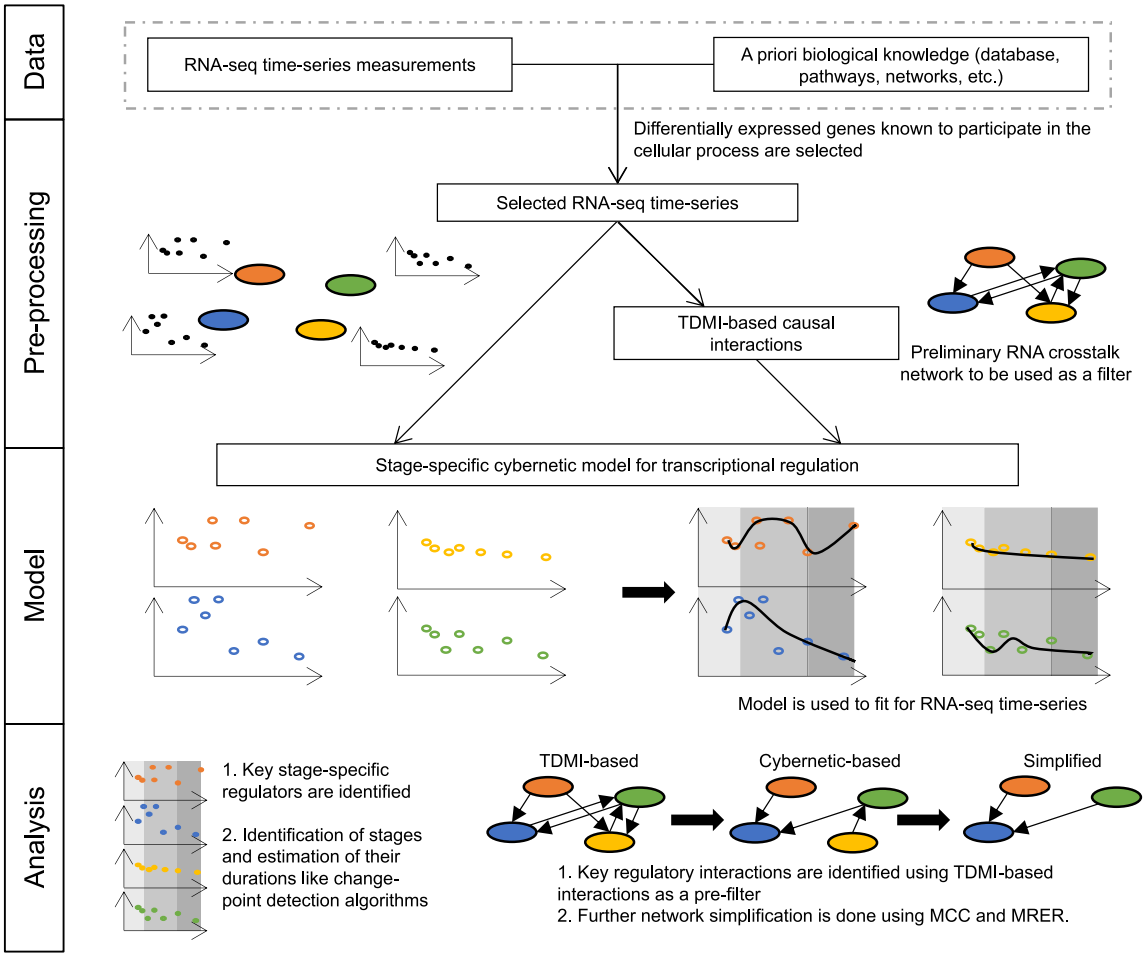

FIGURE 1 Schematic for CIM. Using the selected subset of RNA-seq measurements, a preliminary RNA crosstalk network is developed using the time-delayed mutual information (TDMI). The stage-specific cybernetic model is then applied on the data and the TDMI network. By optimizing the parameters, the data are fitted and the key species and regulatory interactions are identified. This approach also identifies the stage durations. Further network simplification is carried out using the maximal clique centrality (MCC) and the maximal RNA expression rate (MRER) (all data and networks shown in this figure are for representation and not real).

MATERIALS AND METHODS

Identification of causal interactions using TDMI

The advantages of working with information-theoretic approaches (26) are that they facilitate quantifying the interactions between datasets and avoid assuming a functional form of their relationship (27). Specifically, we employed mutual information (MI), an information-theoretic approach that quantifies the linear and nonlinear interactions between the variables (28–31).

The MI $I(X;Y)$ for random variables X and Y, given random samples $\{x_1,\ldots,x_S\}$ and $\{y_1,\ldots,y_S\}$ (S denotes the number of samples), with joint probability function $p(x_i,y_j)$ and marginal probability functions $p_x(x_i)$ and $p_y(y_j)$, is

$$I(X;Y) = \sum_{j=1}^{S}\sum_{i=1}^{S} p(x_i,y_j)\,log\left(\frac{p(x_i,y_j)}{p_x(x_i)p_y(y_j)}\right) = <\lambda(X;Y)>_{p(X,Y)} \quad (1)$$

$$\lambda(x_i,y_j) \equiv log\left(\frac{p(x_i,y_j)}{p_x(x_i)p_y(y_j)}\right)$$

where $<\ .\ >$ denotes the average operator. $I(X;Y)$ is the average of the difference between the log likelihood of two variables $\log p(x_i,y_j)$ and single variable $\log p_x(x_i)$, $\log p_y(y_j)$ with respect to joint probability density $p(x_i,y_j)$.

In our approach, the series X_t and Y_t represent RNA time-series data. The goal was to calculate mutual information $I(X_t;Y_t)$ for different combinations of RNA time-series measurements; e.g., $I(RNA_1,RNA_2)$. The above MI definition assumes that the samples x_i are sampled independently; i.e., $\{x_1,\ldots,x_n\}$ belongs to an independent set (the same is true for y_j samples as well). However, for time-series data, the assumption of independence will be invalid due to temporal correlations between data points at various times.

Galka et al. developed the innovation approach to MI for temporally correlated time series (32). Although the authors derived an $I(X_t;Y_t)$ formula for a Gaussian distributed innovation set that forms an independent and identically distributed (iid) sequence, here, we developed an approach valid for all distributions. For time series, $\lambda(x_i,y_j)$ in Eq. 1 was redefined by pairing the time points as

$$\lambda(X_t;Y_t) = log\,p((x_1,y_1),..,(x_{Nt},y_{Nt})) - log\,p_x(x_1,..,x_{Nt}) - log\,p_y(y_1,..,y_{Nt}) \quad (2)$$

Eq. 2 had high-dimensional joint distributions with complicated structure. To simplify the structure of these distributions, we described these correlations by the corresponding optimal predictors of X_t and Y_t using a time-series prediction model: auto-regression (AR) or vector auto-regression (VAR). The "lag" hyperparameters for the AR and VAR models were selected based on minimization of Akaike information criterion (AIC). The calculated residuals (ϵ_t) based on the expected model (E) are called innovations.

$$\begin{aligned}\epsilon_t(x|x) &= x_t - (E_{AR}(x_t|x_{t-1}, x_{t-2}, \dots))\\ \epsilon_t(y|y) &= y_t - (E_{AR}(y_t|y_{t-1}, y_{t-2}, \dots))\\ (\epsilon_t(x|x,y), \epsilon_t(y|x,y))^{\dagger} &= (x_t, y_t)^{\dagger} - (E_{VAR}\\ ((x_t, y_t)^{\dagger}|(x_{t-1}, y_{t-1})^{\dagger}, (x_{t-2}, y_{t-2})^{\dagger}\dots,))\end{aligned} \tag{3}$$

Since the probabilities of original series are same as the probabilities of their innovations (32), Eq. 2 becomes

$$\begin{aligned}\lambda(X_t; Y_t) &= \log[p((\epsilon_1(x|x,y), \epsilon_1(y|x,y)), \dots,\\ &(\epsilon_{Nt}(x|x,y), \epsilon_{Nt}(y|x,y)))] - \log[p_x(\epsilon_1(x|x), \dots, \epsilon_{Nt}(x|x))]\\ &- \log[p_y(\epsilon_1(y|y), \dots, \epsilon_{Nt}(y|y))]\end{aligned} \tag{4}$$

For optimal predictors of time series, the innovations are white noise, and they are independent. Moreover, if the time points are assumed to form an independent and identically distributed sequence, the joint probability of the innovation time series deciphers as the product of marginal probability densities. Eq. 4 became

$$\begin{aligned}\lambda(X_t; Y_t) &= \sum_{t=1}^{Nt} \log\left\{\frac{p(\epsilon_t(x|x,y), \epsilon_t(y|x,y))}{p_x(\epsilon_t(x|x))p_y(\epsilon_t(y|y))}\right\}\\ &= N_t \log\left\{\frac{p(\epsilon(x|x,y), \epsilon(y|x,y))}{p_x(\epsilon(x|x))p_y(\epsilon(y|y))}\right\}\end{aligned} \tag{5}$$

The MI per time point was then calculated using Eqs. 1 and 5 as follows:

$$\begin{aligned}I(X_t; Y_t) &= \frac{1}{N_t}\sum_{x,y} p(\epsilon(x|x,y), \epsilon(y|x,y))\lambda(X_t; Y_t)\\ &= \sum_{x,y} p(\epsilon(x|x,y), \epsilon(y|x,y))\log\left\{\frac{p(\epsilon(x|x,y), \epsilon(y|x,y))}{p_x(\epsilon(x|x))p_y(\epsilon(y|y))}\right\}\end{aligned} \tag{6}$$

The probabilities in the above equation were estimated using kernel density estimators (KDEs) after finding the innovations using AR and VAR models (33). We then normalized the MI value by dividing it with the average of the entropies: $I_{norm}(X_t; Y_t) = 2I(X_t; Y_t)/(H(x_t) + H(y_t))$.

To identify the causal interactions between any two species within the network, we used the TDMI. For calculating TDMI, a relative time delay (τ) was introduced between the series X_t and Y_t, and then the MI formula was applied to this τ-shifted time series; i.e., X_t and $Y_{t-\tau}$. Thus, the TDMI calculation entails evaluating MI as a function of τ. Further, at the maximum MI value ($TDMI^{max}$), the sign of τ can be used to infer the direction of causality between X_t and Y_t processes from the information transfer perspective (34,35).

$$TDMI(X_t; Y_t, \tau) = \begin{cases} I_{norm}(X_{t+\tau}; Y_t)\ for\ \tau > 0;\\ I_{norm}(X_t; Y_t)\ for\ \tau = 0;\\ I_{norm}(X_t; Y_{t+\tau})\ for\ \tau < 0;\end{cases} \tag{7}$$

CIM for transcription regulation

Cybernetic models in the past have been predominantly used in modeling metabolic regulation where two sets of control variables for enzyme activation and synthesis control the dynamics based on a system objective. Here, we specifically modeled RNA expression levels where the underlying regulation is implicitly incorporated in the cybernetic model using the new control variables u and v, which are different from the past and will be defined below. The transcriptional regulation can be affected by the expression of RNAs, histones, transcription factors, or chromatin modifications and topological constraints associated with the state. To incorporate the effect of crosstalk or interactions across RNA players due to these multilevel intermediates, we defined a lumped species g called the regulator of gene expression. Thus, in our formulation, the RNA expression depends on the species g and g depends on the indirect interactions (proxied through other RNAs). Here are the model equations:

$$\frac{dRNA_j}{dt} = v_j k_j^r g_j - \gamma_j RNA_j \tag{8}$$

$$\frac{dg_j}{dt} = \alpha + u_j \max\left(\sum_i k_{i,j}^g RNA_i, 0\right) - \beta g_j \tag{9}$$

In Eq. 8, the parameter k_j^r is the RNA_j expression rate constant and γ_j is the RNA_j degradation rate. In Eq. 9, the three parameters denote the basal priming rate α, the interaction parameter for RNA_j by RNA_i $k_{i,j}^g$, and the decrease of priming level by the rate constant β. The interaction parameter was set to be nonzero only for possible interactions pre-determined using a priori knowledge or a data-driven approach. They are either activating (+) or repressing (−) rate constants allowing the term $\sum_i k_{i,j}^g RNA_i$ in Eq. 9 to become negative for some conditions. We allowed a basal g_j level (α/β) even when the above term becomes negative. We implemented this constraint by using $\max(\sum_i k_{i,j}^g RNA_i, 0)$ in Eq. 9. The control variables u and v are regulating the level and strength of priming for the RNA transcription, respectively, and are defined based on the cybernetic objective that is either intuitively described or phenotypically described from the experimental data.

We defined the objective of this system to be to maximize $\sum_{j=1}^n (w_j k_j^r g_j)$, that is, to maximize the sum of weighted production rates of the RNAs. This objective form is similar to those used in the original cybernetic models and means that the RNA production is optimized based on the functional requirement decided by the weights. The control variables, v_j and u_j, are computed by solving the optimal control problem resulting in the proportional and matching laws (10), respectively, as follows:

$$v_j = \frac{w_j k_j^r g_j}{\max\limits_{i=1,2,\dots,n}\left(w_i k_i^r g_i\right)}, u_j = \frac{w_j k_j^r g_j}{\sum_{i=1}^n \left(w_i k_i^r g_i\right)} \tag{10}$$

The expressions for the control variables, although appearing simplistic, cause the Ordinary Differential Equation (ODE) system (Eqs. 8 and 9) to be nonlinear.

In this formulation, we attached time/age to the cell transition. If the values t_1, t_2, t_3, t_4, t_5, and t_6 represent the times at which the cell's objective changes for $G1 \rightarrow S_1$, $S_1 \rightarrow S_2$, $S_2 \rightarrow S_3$, $S_3 \rightarrow G2$, and $G2 \rightarrow M$ transitions and cell division, respectively, then the weights are as follows:

$$w_j = \begin{cases} w_{G1,j}(0 < t \le t_1) \\ w_{S1,j}(t_1 \le t < t_2) \\ w_{S2,j}(t_2 \le t < t_3) \\ w_{S3,j}(t_3 \le t < t_4) \\ w_{G2,j}(t_4 \le t < t_5) \\ w_{M,j}(t_5 \le t < t_6) \end{cases} \tag{11}$$

The fits are attained based on the global minimization of the sum of the squares of normalized fit errors (SSEs) where the RNA-specific fit errors are normalized by dividing it with their maximum value. To obtain only the key regulatory interactions, we enforce the parameter $k^g_{i,j}$ to become zero if its absolute value falls below 0.1 during the global minimization, whereas the maximal absolute value is set as 100. These thresholds provide us with sparse matrix for $k^g_{i,j}$ to highlight important edges of the network.

$$SSE = \sum_j \left(\frac{RNA_j - data(RNA_j)}{\max\left(data(RNA_j)\right)} \right)^2 \tag{12}$$

Application of CIM framework to mouse cell cycle measurements

Transcriptomic time-series measurements for synchronized cell cycle of Cf-1 mouse embryonic fibroblast primary cells (E13 embryos) were available from our prior work (3). To synchronize the cell cycle, cells were incubated in a starvation medium (0.5% fetal calf serum) for 36 h and then serum was added to reach 20% to re-initiate the cell cycle (3). RNA-seq data were measured at 96 different time points with 0.5- or 1-h intervals covering more than one full cell cycle (3). Of the 4248 genes differentially expressed (more than twofold up or down as compared to t = 0) at one or more time points, we selected 63 canonical cell cycle genes and 23 more transcription factor genes (a total of 86) based on our prior model (3).

RESULTS

Modeling transcriptomic regulation during the cell cycle

We illustrate the versatility of our CIM framework by applying it to model transcriptional regulation during cell cycle progression in a mammalian cell. Although our CIM approach has the ability to infer the regulatory interactions on unbiased choice of molecules, it is important to start with a model network based on a priori knowledge or by time-series analysis techniques to avoid model overparameterization (Fig. 1) (36,37). This will also enable the approach to relate to known biology. Using the initial set of 86 RNA transcripts as nodes (see section "materials and methods" for selection criteria), we evolved the network by introducing mechanistic causality using our longitudinal measurements and MI.

Causal interaction network development using TDMI

We refined our preliminary network model using the TDMI (see section "materials and methods"). Our approach was applied to the time-series data of the 86 selected genes to calculate the TDMI for the range of delays (τ) between −20 and 20 h. In Fig. 2, we show a sample calculation for TDMI between the two nodes (genes) in the network, Tgfb1 and Ets1 (Fig. 2 *A* and *B*). We used the innovation approach (32) for TDMI calculation in which the residuals were calculated using AR and VAR models (see section "materials and methods"). Then, we estimated the probability functions based on mono-variate and bi-variate KDE models (Fig. 2 *C–E*) (33). The TDMI were estimated across different τ values (see section "materials and methods"; Fig. 2 *F*). We can see that the TDMI peaks at a positive lag value of 4 h (Fig. 2 *F*), implying that Ets1 is the cause and Tgfb1 the effect. This calculation was repeated for every pair of the selected genes where we estimated the maximum value of TDMI both for positive and negative τ values as $TDMI^{max}_{+}$ and $TDMI^{max}_{-}$, respectively. We selected the top 10 percentile of the interactions from the maximal value of TDMI (combined set of $TDMI^{max}_{+}$ and $TDMI^{max}_{-}$) (Figs. 2 *G* and S1). Therefore, if $TDMI^{max}_{+}$ and $TDMI^{max}_{-}$ between two species are within the top 10 percentile, then they have bidirectional interaction. The interaction is similarly bidirectional if the TDMI peak occurs at $\tau = 0$ ($TDMI^{max}_{+} = TDMI^{max}_{-}$) and the value is within the top 10 percentile. These interactions among the RNA transcripts (nodes) are visualized as networks using Cytoscape (Fig. 2 *G*). Although these selected interactions are highly possible, this TDMI-based approach does not provide any mechanistic proof for these interactions. So, we need a mechanistic model such as the CIM approach to further validate the network.

Stage-specific CIM for transcriptional regulation

We developed a CIM framework to model transcriptional regulation incorporating stage-varying and multifactorial regulation (see section "materials and methods"). During cell cycle, the objective or goal of the system depends on the current phenotype of the cell and needs to be redefined each time the cell undergoes a major transition (stage change) owing to transcriptional remodeling. Although a single multi-weighted objective function describes regulation for short periods within a single stage, for modeling long-time-series measurements, we had to incorporate stage-specific objective functions. We mathematically defined that the objective for each stage is to maximize

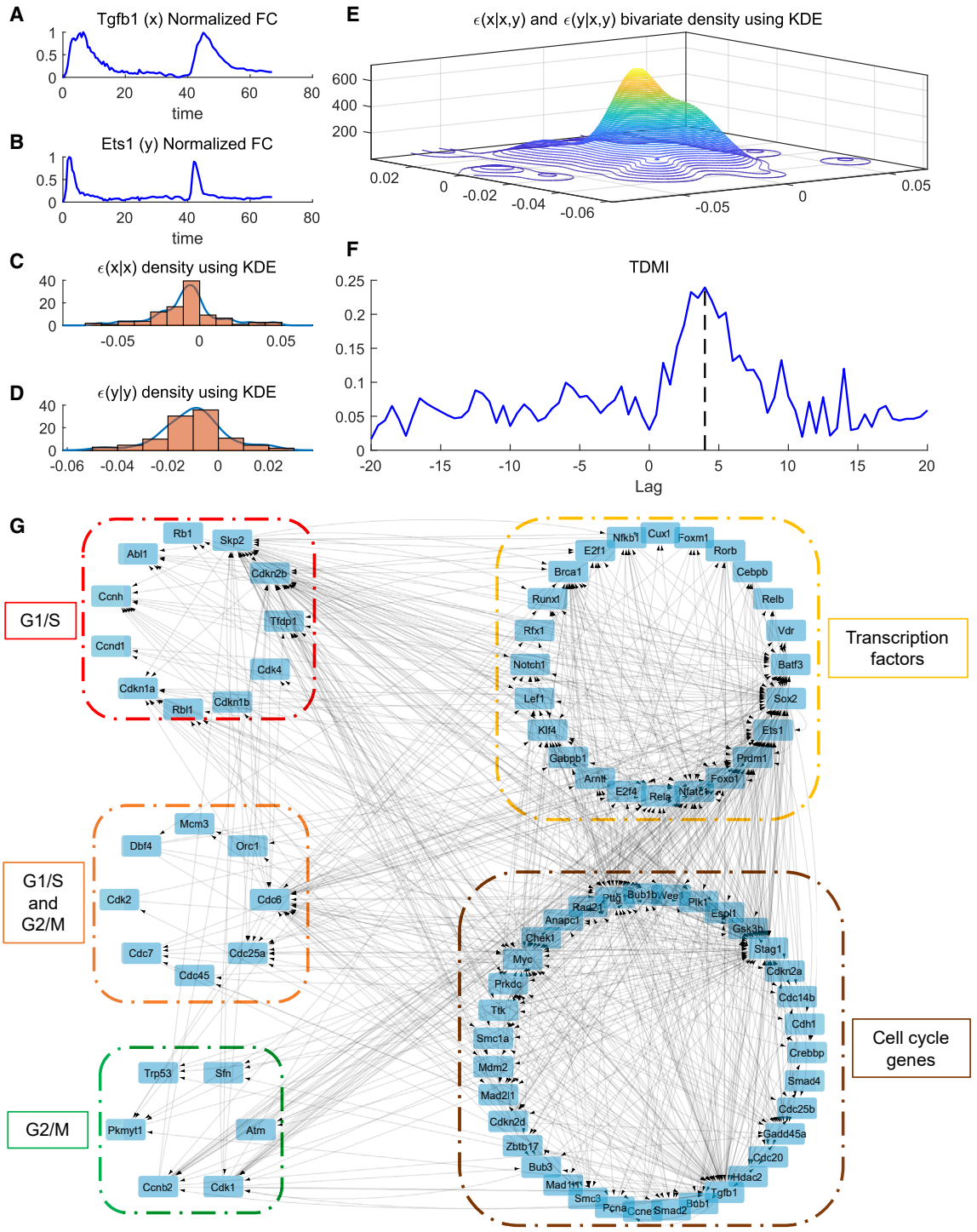

FIGURE 2 Estimating TDMI using the innovation approach and using it to generate a causal interaction network. (*A* and *B*) Time-series RNA transcriptomics measurements for Tgfb1 and Ets1. (*C* and *D*) Mono-variate KDE for probability densities of residuals. (*E*) Bi-variate KDE for probability density of residuals. (*F*) TDMI calculated between time series of Tgfb1 and Ets1 for different lags. (*G*) Interaction network development for the cell cycle model using TDMI. Network construction is based on TDMI threshold of top 10 percentile. The boxed sections are color coded based on category mentioned.

the sum of weighted production rates of the RNAs where the weights are computationally fitted (see section "materials and methods"). The importance of an RNA during a particular stage can be inferred based on its weight for the stage. The higher the weight for RNA, the higher the functional necessity for it during a particular stage and vice versa, and thus this approach connects the mathematical objective to biological objective. During cell cycle, we incorporated the changes in the objective functions by fitting for different weights for each stage. Each stage was modeled with its own

w_j represented by $w_{k,j}$, where $k = \{G1, S_1, S_2, S_3, G2, M\}$ represents the sequential objective-changing stages of the cell. It is important to note that, although most cell cycle phases (experimentally defined) can be modeled using a specific single objective (e.g., $G1$, $G2$, M), some stages will require special modifications based on the biological process. Here, during S phase, there is competition between transcription events and DNA replication, and there is a global anti-correlation between replication and transcription timing making it impossible to model using a single objective (38). This observation is an important part of the model development here, which necessitates a stage-specific objective approach and thus a significant departure from cybernetic modeling of microbial metabolism. Therefore, S phase can have multiple stages and here we used three stages to model S phase (represented as S_1, S_2, S_3), with each stage having its own objective (Fig. 3).

Model captures the experimental measurements

We applied our CIM to the selected 86 genes. Since we already identified the most likely interactions based on TDMI-based time-series analysis, we can prevent the overparameterization problem in our approach by allowing only those interactions to have a nonzero interaction parameter ($k^g_{i,j}$) value. We had time-series measurements for 67 h covering the period exceeding one cell cycle. After the first cell division, we assume that some cells of fraction f_{G0} could transit to G0 phase or cell cycle arrest (Fig. 3 *A* and *B*), whereas the remaining cells of fraction $(1 - f_{G0})$ continued to the second cell cycle with objective-changing time points to be $t_7, t_8,$ etc. (Fig. 3 *A* and *B*). We used the following strategy to validate our model. For the first cell cycle period, we fitted for all the intrinsic RNA parameters (k^r_j, γ_j, α, β, $k^g_{i,j}$), initial point for regulator ($g_j(0)$), stage-specific weights representing their objectives ($w_{G1,j}$, $w_{S1,j}$, $w_{S2,j}$,

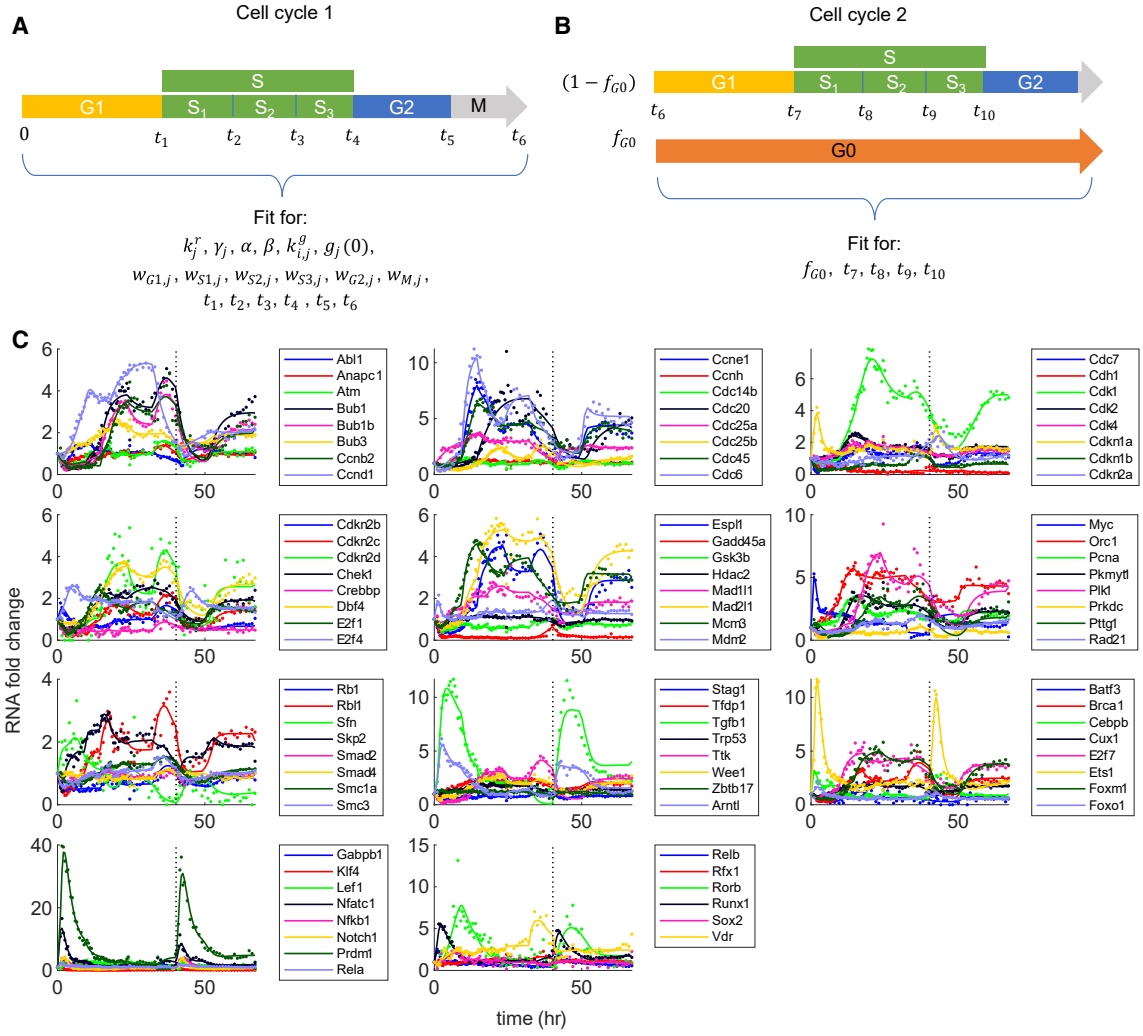

FIGURE 3 Approach for self-validating the CIM. Different stages in the stage-specific cybernetic model for more than one cell cycle period are shown. The S phase has three sub-phases (represented as S1, S2, S3), with each sub-phase having its own objective. (*A*) During cell cycle 1, all model parameters, including the RNA intrinsic parameters, are fitted. (*B*) During cell cycle 2, fraction of cells (f_{G0}) will go to cell cycle arrest (G0 phase). By only fitting for the new parameters, we validate our model. (*C*) Model fits to all 86 canonical cell cycle and transcription factor RNAs. The dots represent experimental data and lines represent model fits. The vertical black dotted line represents cell division.

$w_{S3,j}$, $w_{G2,j}$, $w_{M,j}$), and the respective time points for objective change ($t_1, t_2, t_3, t_4, t_5, t_6$). We solved the model to attain global minimization of the sum of the squares of normalized fit errors using the MATLAB ODE function "ode15s" and optimization functions "lsqnonlin" and "patternsearch" starting with 100 different initial conditions. For the time period after the cell division, we modeled the two fractions of cells representing G0 stage and remaining cells in their second cell cycle separately and added them based on their respective fractions. Here, we assumed that the G1 and G0 stages have the same objectives and therefore equal weights ($w_{G0,j} = w_{G1,j}$). While fitting the period after cell division, we used the same RNA intrinsic parameters and weights and fitted only for the f_{G0} and the transition times (t_7, t_8) (Fig. 3). Fig. 3 *C* shows the overall fits for the full duration of 67 h. Tables S1 and S2 show the overall fitted model parameters. The model described the first cell cycle and was also able to fit the time period after cell division using the same parameters, thereby providing a biologically feasible solution using our approach (this partially validates the computational model).

Network reconstruction using CIM

We reconstructed the network based on the interaction parameter $k^g_{i,j}$ (sparce matrix) by creating edges only between the nodes with a nonzero $k^g_{i,j}$ value. This resulted in only 153 interactions within this network of 86 nodes (Figs. 4 *A* and S2). This is far less than what was predicted using the TDMI approach (731 interactions). The positive $k^g_{i,j}$ represents activation and negative $k^g_{i,j}$ represents repression. The absolute value of $k^g_{i,j}$ represents the strength of interaction. These interactions are identified based on a mechanistic model that is nonlinear and multifactorial, thereby making these interactions highly probable. Comparing this network with a canonical protein-protein interaction network including second neighbor

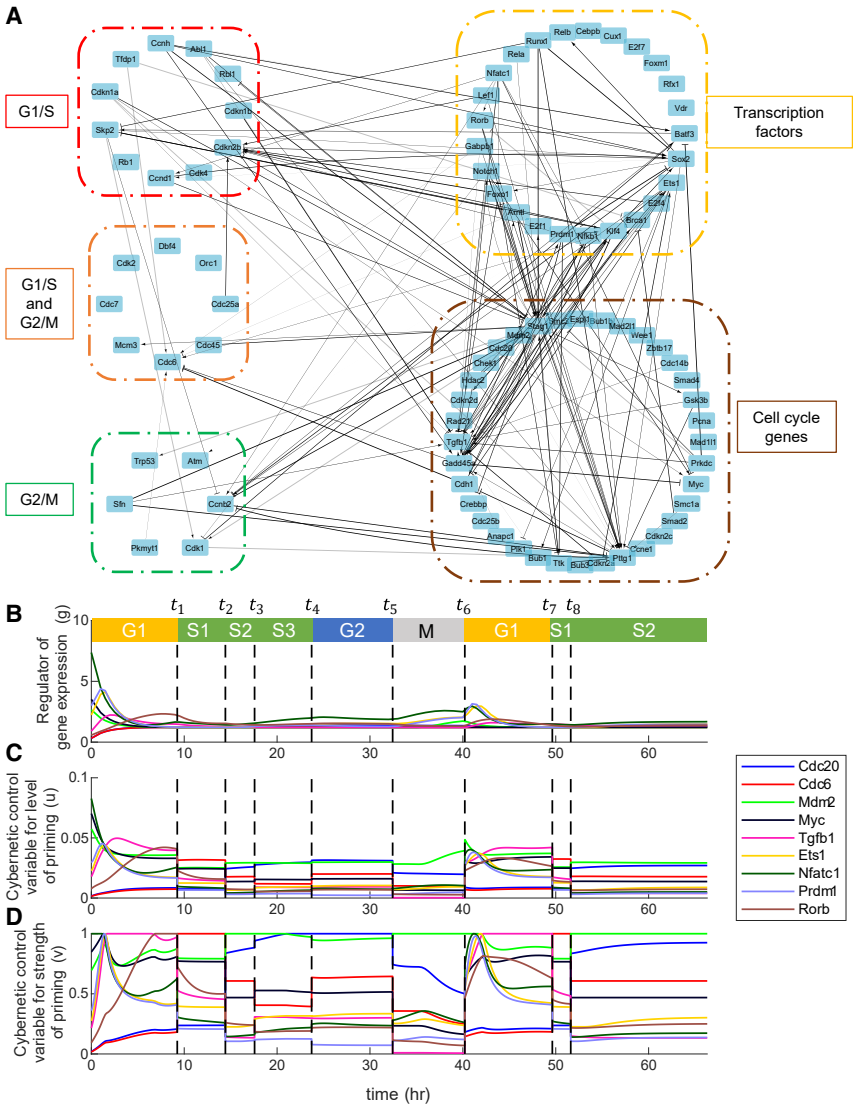

FIGURE 4 The reconstructed network based on CIM and the identification of the key regulators. (*A*) The activating interactions are shown as arrows and repressing interactions are shown as dashes. The thickness of the interaction lines depends on the absolute value of the interaction parameter, $k^g_{i,j}$. The boxed sections are color coded based on category mentioned. The dynamics of regulator of gene expression g (*B*) and control variables u (*C*) and v (*D*) for the key genes. The black dashed lines represent the change points. Each region is color coded based on its respective stage.

interactions shows that 90% of these interactions are observed in the literature (Fig. S3).

Novel interactions not present on string database

The cell cycle is an extensively studied system, but a lot is still unknown. We identified several interactions not previously identified in Sting database using our CIM model (Figs. 5 and S3): Cdk1 → Batf3, Batf3 → Ccnb2, Gabpb1 → Pttg1, Ccnh → Batf3, Prkdc → Batf3, Klf4 → Rorb, Nfatc1 → Rorb, Gabpb1 → Tgfb1, Rorb → Stag1, Stag1 → Trp53, Gabpb1 → Cdc6, Rorb → Pttg1, Abl1 → Batf3, and Cdc14b → Tgfb1. The validity of these interactions should be experimentally tested in the future.

Distinct stages across cell cycle and their durations

Since our approach involves changing from one stage to the next stage, it implies the involvement of change points in our biological process and in our modeling strategy. We can use approaches such as change-point detection (CPD) algorithms well known in the signal processing community to decipher the change points (39). We can then validate these using biological intermediate endpoints such as the those in each phase of the cell cycle. Previously, for the cell cycle system, a model-free CPD algorithm based on singular spectral analysis was used to detect changes in the time series of cell cycle genes (3). They calculated the time of phase change for the cell cycle phases as the time at which they identify significant individual time-series change points detected out of 63 cell cycle genes (3). Based on this criterion, the durations for G1, S, and G2/M phases of the cell cycle were estimated to be 14.5, 10, and 4 h, respectively (3). These CPD algorithms are only based on time series and lack any mechanistic insights. Here, we used the CIM to identify these change points by determining the times at which the objective needs change. The time points for objective change are simultaneously fitted along with other parameters. Our model predicted the duration for each of these stages $\{G1, S_1, S_2, S_3, G2, M\}$ during first cell cycle (Table 1). For the time after cell division, our model predicted duration for the second cycle $\{G1, S_1\}$ to be $\{9.62, 1.78\}$ respectively. The plots of g, u, and v of specific key regulators are shown in Fig. 4 *B–D*. The key regulators are selected based on the criterion that at least at one time point the value of v peaks to 1 for them. For understanding the importance of each RNA transcript, we define a new estimate called maximal RNA expression rate (MRER; $\mathrm{Max}(v_j k_j^r g_j)$). This is the maximum value among the RNA expression rates.

Network simplification and analysis

We analyze the reduced network to identify the key central molecules whose regulations are the most important for the cell cycle. We had grouped the RNA transcripts into groups such as with transcription factor genes, checkpoint genes (G1/S and G2/M), and other canonical cell cycle genes (Figs. 2 *G* and 4 *A*). For identifying key central nodes, we used the following two properties. One is a network property called maximal clique centrality (MCC), which analyzes and identifies the key nodes giving importance to

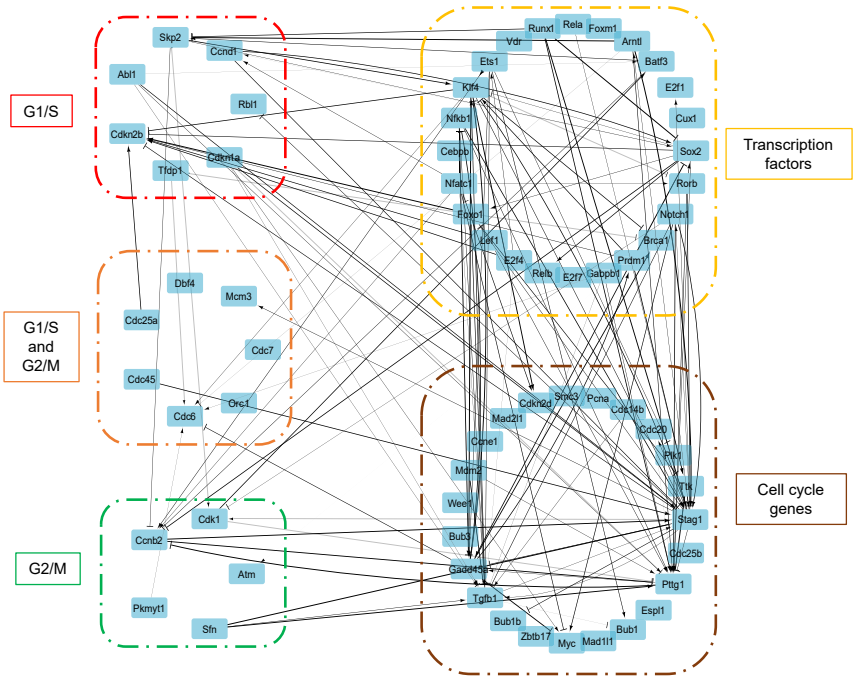

FIGURE 5 Development of simplified network based on MCC and MRER. The activating interactions are shown as arrows and repressing interactions are shown as dashes. The thickness of the interaction lines depends on the absolute value of the interaction parameter, $k_{i,j}^g$. The boxed sections are color coded based on category mentioned. The network shown here is for a union of nodes with MCC $\geq$ 7 and MRER $>$ 1. The boxed sections are color coded based on the category mentioned.

TABLE 1 Predicted phase durations across the first and second cell cycles

Phase duration	Cell cycle 1 (h)	Cell cycle 2 (h)
G1	9.25	9.62
S1	5.18	1.78
S2	3.14	>15.38
S3	6.14	NA
G2	8.72	NA
M	7.77	NA

NA, not applicable.

number and extent of interactions (40). The top nodes based on MCC were Stag1, Tgfb1, and Pttg1. Another property from the CIM is called MRER (see previous section), which helps identify key nodes in a specific-rate manner. Using thresholds for these properties, we were able to eliminate the nodes in the network while retaining the important nodes. We illustrate a simplified network based on the threshold MCC $\geq 7 \cup$ MRER > 1 (Fig. 5). This approach reduced the nodes to 66 (from 86) and interactions to 126 (from 153). This reduced network highlights the key nodes and their interactions based on MCC and MRER combined and could be used to understand the major regulatory features of the cell cycle.

DISCUSSION

Cybernetic methods were originally developed to address unknown regulatory events in a single-stage cellular process with a defined goal. However, several biological systems such as the mammalian cell cycle involve multiple stages with one or more distinct objective functions at each stage. The traditional cybernetic approach is not designed to capture this complexity. Here, we develop a novel strategy for capturing this complexity, and we call it a CIM. To account for each stage and how the stage proceeds in terms of a causal evolution, we introduced TDMI, which uses temporal correlation to provide insights into the causality of regulatory events. The preliminary regulatory network from TDMI is then used as a starting point in the CIM to obtain biologically realistic models. Such pre-processing techniques have greater importance when solving inverse biological problems. The CIM approach significantly reduced the number of interactions to describe the system (compare Figs. 2 *G* and 4 *A*). Although TDMI is one way to identify causality and determine concomitant interactions, we could have used other techniques for time-series analysis, such as Granger causality, to identify these interactions. Knowledge embedded in databases such as Kyoto Encyclopedia of Genes and Genomes pathways (41), Reactome pathway database (42), and ConsensusPathDB (43) can be used for a comprehensive list of legacy pathways to ensure the validity of the initial network. Although extracting an initial network directly from such databases is possible, our approach using TDMI (or other time-series analysis techniques) can better formulate this initial network incorporating causal connections when large longitudinal data are available. In our case, we utilized a TDMI threshold of the top 10 percentile. This resulted in a network that had a 90% overlap with StringDB and was considered good. If there are many false positives, a more stringent threshold, such as the top five percentile, may be considered. Alternatively, if too few connections are inferred, a less stringent threshold, such as the top 20 percentile, may be considered.

Since we are introducing objective functions at distinct stages, we can use stages as defined by experimental measurements if available. However, given the dynamical nature of the system, we may have changes of state that are distinct from the experimentally determined stages. This motivated us to use CIM for CPD, which provides insights into stages of the system, in addition to providing an opportunity to introduce novel objective functions at each of these intermediate stages. We have already shown that the S phase shows differential regulation because of the balance between transcription and replication events and requires multiple stage changes to model them. Since we are modeling regulation within the transcription process, our stage changes are expected to occur before we phenotypically observe them, and this time delay can be attributed to lag between RNA expression and protein production. The dynamical evolution for the stages provides insights into the regulatory mechanisms associated with the causality leading to the global endpoint of the biological system (in this case, cell cycle).

We observe that the stage durations during the second cell cycle are not same as the first. In fact, during the second cell cycle, we did not identify any change point after 51.61 hours (representing S_2 stage), implying that the cell cycle has slowed down or stopped. The reason for cell cycle arrest could be multifaceted, with one of them being attributed to *Cdc6*. *Cdc6*, a gene involved in initiating DNA replication, is low during the second cell cycle compared with the first cell cycle, and this could contribute to stopping the cell cycle. For a cell cycle model, our approach should have an ability to show a cyclic behavior if the objectives are repeated. Toward testing whether the model will yield cyclic solutions for the cell cycle, we repeated the pattern of objectives based on the first cell cycle and this resulted in a cyclic behavior (Fig. S4). Each of these stage objectives are represented by their weights ($w_{G1,j}$, $w_{S1,j}, w_{S2,j}$, $w_{S3,j}$, $w_{G2,j}$, $w_{M,j}$; Fig. S5). When comparing for weight difference across adjacent stages (Fig. S6), we observed that the stages $S_2, S_3,$ and $G2$ had highly similar objectives. It is during these stages that the cell proceeds to division, and having similar objectives is interesting and warrants further studies.

The CIM approach has several quantitative aspects that can be used to infer key regulators in multiple ways. One way is by using the stage-specific weights that provided us with the quantitative relevance of key players within each stage. From the estimated weights (Fig. S5), *Cdc20* is important during the $S_2, S_3,$ and $G2$ stages as its protein is required for nuclear

movement before anaphase and chromosome separation. We can also identify key regulators from the regulator of gene expression and the control variables (Fig. 4 *B–D*). For $G1$ phase, *Tgfb1*, *Nfatc1*, *Myc*, *Ets1*, *Rorb* and *Prdm1* are key regulators (Fig. 4 *B–D*). Tgfb1 protein is a multifunctional protein that controls proliferation, differentiation, and other functions in many cell types and is known to participate in regulation of the G1/S checkpoint. For S_1 phase, *Cdc6* is a key regulator as its protein is known to be involved in the initiation of DNA replication and also participates in checkpoint controls that ensure DNA replication is completed before mitosis is initiated (Fig. 4 *B–D*). During S_2, S_3, and M phases, *Mdm2* is a key regulator since its protein is known to inhibit p53/TP53- and p73/TP73-mediated cell cycle arrest and thereby prevent cell cycle arrest. Here, *Cdc20* is a key regulator for S_3 and $G2$ phases. For understanding the phase-specific importance for each RNA transcript, we used MRER (Fig. S7).

We further simplified the network based on a combination of network and dynamical property (MCC and MRER, respectively). Fig. 6 shows the part of simplified network involving only canonical cell cycle genes without the checkpoint genes. This network provides us with the genes/proteins required for the cell cycle. *Tgfb1*, *Stag1*, and *Pttg1* are key regulators with high degrees. *Tgfb1* has high MRER during $G1$ phase representing its key participation in the regulation of G1/S checkpoint. Stag1 has high MRER during M phase as it is required in the cohesion of sister chromatids after DNA replication. *Pttg1* has high MRER during S_3 and M phases because of its central role in chromosome stability, p53/TP53 pathway, and DNA repair. *Gadd45a*, known to stimulate DNA excision repair in vitro and inhibit entry of cells into S phase, negatively regulates Myc, which activates transcription of growth-related proteins. Most of these RNAs have high MRER during the later stages of cell cycle since multiple players are required during cell division. Fig. 7 shows the nodes of the simplified network of transcription factor genes and checkpoint genes. Unlike canonical cell cycle genes, most of the transcription factor genes have high MRER during the $G1$ phase, possibly because of high requirement of transcription factors after cell division for cellular growth. Other transcription factor genes have some role in the cell cycle, as shown in Fig. 7 *A*. *Vdr*, a nuclear receptor for vitamin D_3, also supports the cellular growth (44). *Batf3* can also support the cell cycle based on protein-protein interaction with CDK6, CCND2, CDK2, and TP53. The checkpoint genes show high variability based on phase-specific MRER values (Fig. 7 *B–D*), showing that they are expressed at various time points but can act as phase-specific checkpoints.

Our approach provides a highly adaptable solution compared to dynamical models that rely on steady and oscillatory states with key species to elucidate the progression of the cell cycle (19,45). These models necessitate parameters extracted from existing literature, rendering them challenging

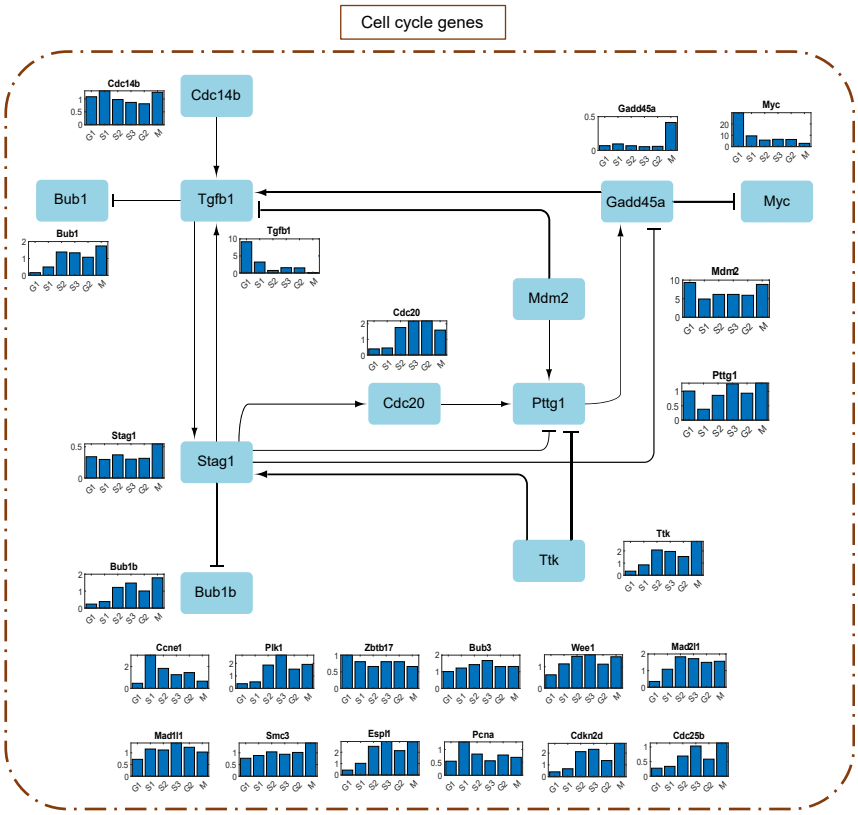

FIGURE 6 Simplified network involving only canonical cell cycle genes without the checkpoint genes and their interactions. The activating interactions are shown as arrows and repressing interactions are shown as dashes. The thickness of the interaction lines depends on the absolute value of the interaction parameter, $k^g_{i,j}$. The ones without cross-interactions within the group are shown below the network. The bar plots represent the stage-specific MRER.

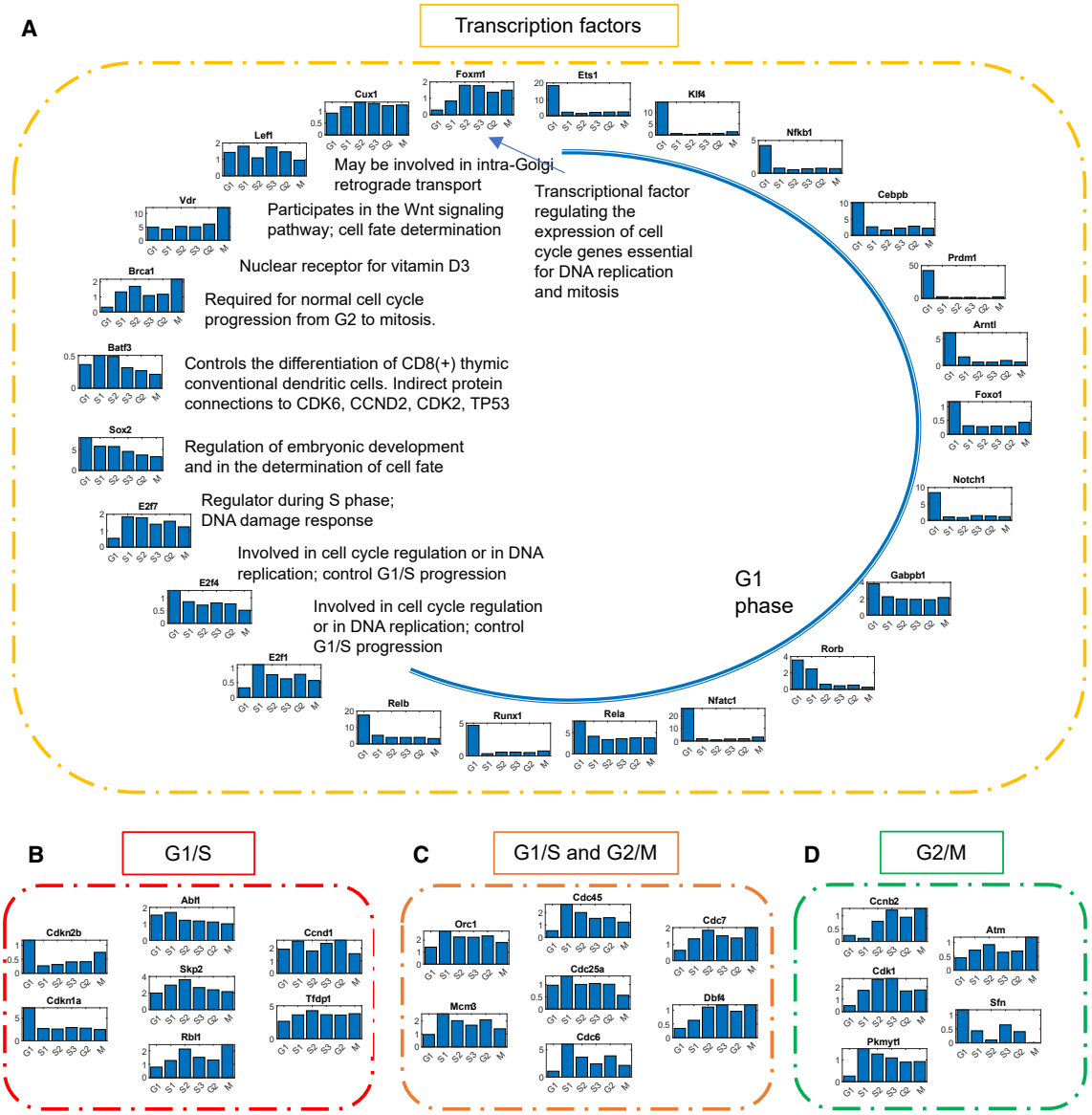

FIGURE 7 Network nodes involing transcription factor genes and checkpoint genes. Nodes of the simplified network of (*A*) transcription factor genes, (*B*) G1/S checkpoint genes, (*C*) genes considered both G1/S and G2/M checkpoints, and (*D*) G2/M checkpoint genes. The bar plots represent the stage-specific MRER.

to employ in systems lacking well-understood kinetic models and parameters. In contrast, our approach can be implemented in any system, provided that temporal data are available. Although we have evaluated our approach using the cell cycle system, it is not restricted to this system alone.

Although there were several attempts to develop transcriptional regulatory networks (25,46–50), these approaches were not extended to dynamically model longitudinal measurements. Our approach differs from ARACNE (25), Scribe (48), and LEAP (49) as we use CIM to construct our model network, which is not comparable to the information theory techniques used by these approaches. However, our intermediate network based on the TDMI threshold can be compared to those obtained from the above-mentioned methods. There are also other approaches for network inference, such as those based on statistical significance of the edge coefficients (51–53).

These eukaryotic systems show complex regulatory patterns that simple linear models cannot capture. The ability of the cybernetic approach to address this regulatory complexity lies in the nonlinearity enforced by the cybernetic control variables as well as the segregation of the system into multiple stages with varying control objectives. In our approach, we have used "time" to characterize all transitions (stage change) because of its inherent simplicity. However, we recognize that such transitions may be related more deeply to intracellular variables, a potential that can only be realized when sufficient understanding of the system

becomes available. Genetic perturbation experiments, such as gene silencing, have the potential to affect the durations of different stages. If we can determine the stage durations based on the current species states, our approach can be expanded to predict outcomes for these experiments. As we have modeled a synchronous cell cycle here, all cells may be assumed to behave similarly. On the other hand, most biological experiments are asynchronous, necessitating some form of averaging to be superimposed on the model where we can model cell behavior based on averaging weights over the stage-specific number density of cells.

The complexity within cellular processes due to the multitude of regulatory interactions is often difficult to infer using previous mathematical models. The cybernetic approach of mapping the cellular regulation to objectives that can be mathematically formulated can overcome this difficulty. Based on the observation that the cellular processes are multi-staged, we developed a novel approach that incorporates the stage-specific objectives. The implications of our proposed approach go beyond modeling the mammalian cell cycle processes. We can formulate complex cellular processes, where only sparse measurements are available, albeit with knowledge of intermediate endpoints, in terms of the CIM, enabling us to infer multiple unknown regulatory processes. For instance, we can consider a developmental process of lineage specification of a specialized tissue from pluripotent stem cells, where targeted measurements are available across stages of development (pseudo-time) as an exemplary problem for this approach. Although such investigations are in progress, it must be evident that the cybernetic approach by comprehensive addition of regulatory intervention has a higher potential to discover new biological phenomena.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2023.12.010.

AUTHOR CONTRIBUTIONS

R.R. wrote the manuscript. M.R.M., S.S., and D.R. revised the manuscript. R.R., S.S., and D.R. designed the research. R.R. performed the research. R.R. analyzed the data. S.K., L.A., M.R.M., S.G., S.S., and D.R. contributed analytical tools.

ACKNOWLEDGMENTS

This work was supported by NIH grants R01 LM012595 (S.S.), OT2 OD030544 (S.S.), R01 HL106579 (S.S.), R01 HL108735 (S.S.), U2C DK119886 (S.S.), The Joan And Irwin Jacobs Endowed Professorship (S.S.), Center for Social Inclusion (CSOI), A National Science Foundation Science And Technology Center under grant agreement CCF-0939370 (S.S. and D.R.), and The Harry Creighton Peffer Endowed Professorship (D.R.).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Maurya, M. R., and S. Subramaniam. 2007. A Kinetic Model for Calcium Dynamics in RAW 264.7 Cells: 1. Mechanisms, Parameters, and Subpopulational Variability. *Biophys. J.* 93:709–728.
2. Maurya, M. R., and S. Subramaniam. 2007. A Kinetic Model for Calcium Dynamics in RAW 264.7 Cells: 2. Knockdown Response and Long-Term Response. *Biophys. J.* 93:729–740.
3. Masnadi-Shirazi, M., M. R. Maurya, …, S. Subramaniam. 2019. Time varying causal network reconstruction of a mouse cell cycle. *BMC Bioinf.* 20:294.
4. Gianchandani, E. P., A. K. Chavali, and J. A. Papin. 2010. The application of flux balance analysis in systems biology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2:372–382.
5. Bornheimer, S. J., M. R. Maurya, …, S. Subramaniam. 2004. Computational modeling reveals how interplay between components of a GTPase-cycle module regulates signal transduction. *Proc. Natl. Acad. Sci. USA.* 101:15899–15904.
6. Gupta, S., M. R. Maurya, …, S. Subramaniam. 2009. An Integrated Model of Eicosanoid Metabolism and Signaling Based on Lipidomics Flux Analysis. *Biophys. J.* 96:4542–4551.
7. Hanly, T. J., and M. A. Henson. 2011. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol. Bioeng.* 108:376–385.
8. Chandrasekaran, S., and N. D. Price. 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA.* 107:17845–17850.
9. Trinh, C. T., A. Wlaschin, and F. Srienc. 2009. Elementary mode analysis: A useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.* 81:813–826.
10. Ramkrishna, D., and H.-S. Song. 2018. Cybernetic Modeling for Bioreaction Engineering. Cambridge University Press.
11. Dhurjati, P., D. Ramkrishna, …, G. T. Tsao. 1985. A cybernetic view of microbial growth: Modeling of cells as optimal strategists. *Biotechnol. Bioeng.* 27:1–9.
12. Kompala, D. S., D. Ramkrishna, …, G. T. Tsao. 1986. Investigation of bacterial growth on mixed substrates: Experimental evaluation of cybernetic models. *Biotechnol. Bioeng.* 28:1044–1055.
13. Song, H.-S., D. Ramkrishna, …, J. K. Fredrickson. 2013. Dynamic modeling of aerobic growth of Shewanella oneidensis. Predicting triauxic growth, flux distributions, and energy requirement for growth. *Metab. Eng.* 15:25–33.
14. Aboulmouna, L., S. Gupta, …, D. Ramkrishna. 2018. A Cybernetic Approach to Modeling Lipid Metabolism in Mammalian Cells. *Processes.* 6:126.
15. Aboulmouna, L., R. Raja, …, D. Ramkrishna. 2020. Cybernetic modeling of biological processes in mammalian systems. *Curr. Opin. Chem. Eng.* 30:120–127.
16. Fischer, M., A. E. Schade, …, J. A. DeCaprio. 2022. Coordinating gene expression during the cell cycle. *Trends Biochem. Sci.* 47:1009–1022.
17. Xu, J. 2005. Preparation, Culture, and Immortalization of Mouse Embryonic Fibroblasts. *In* Current Protocols in Molecular Biology.
18. Tyson, J. J., and B. Novák. 2015. Models in biology: Lessons from modeling regulation of the eukaryotic cell cycle. *BMC Biol.* 13:1–10.
19. Csikász-Nagy, A., D. Battogtokh, …, J. J. Tyson. 2006. Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophys. J.* 90:4361–4379.
20. Tyson, J. J. 2021. Cell Cycle Regulation. Bifurcation Theory. *In* Case Studies in Systems Biology. P. Kraikivski, ed Springer International Publishing, Cham, pp. 41–57.

21. Tyson, J. J., T. Laomettachit, and P. Kraikivski. 2019. Modeling the dynamic behavior of biochemical regulatory networks. *J. Theor. Biol.* 462:514–527.
22. Maurya, M. R., and S. Subramaniam. 2010. Chapter 8 - Computational Challenges in Systems Biology. *In* Systems Biomedicine. E. T. Liu and D. A. Lauffenburger, eds Academic Press, San Diego, pp. 175–223.
23. Farhangmehr, F., M. R. Maurya, …, S. Subramaniam. 2014. Information theoretic approach to complex biological network reconstruction: application to cytokine release in RAW 264.7 macrophages. *BMC Syst. Biol.* 8:77.
24. Zhang, X., X. M. Zhao, …, L. Chen. 2012. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics.* 28:98–104.
25. Margolin, A. A., I. Nemenman, …, A. Califano. 2006. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinf.* 7:S7.
26. Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27:379–423.
27. Little, D. Y., and L. Chen. 2009. Identification of coevolving residues and coevolution potentials emphasizing structure, bond formation and catalytic coordination in protein evolution. *PLoS One.* 4:e4762.
28. Numata, J., O. Ebenhöh, and E. W. Knapp. 2008. Measuring correlations in metabolomic networks with mutual information. *Genome Inform.* 20:112–122.
29. de Siqueira Santos, S., D. Y. Takahashi, …, A. Fujita. 2013. A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings Bioinf.* 15:906–918.
30. Kvålseth, T. 2017. On Normalized Mutual Information: Measure Derivations and Properties. *Entropy.* 19:631.
31. Batina, L., B. Gierlichs, …, N. Veyrat-Charvillon. 2011. Mutual information analysis: A comprehensive study. *J. Cryptol.* 24:269–291.
32. Galka, A., T. Ozaki, …, O. Yamashita. 2006. Whitening as a tool for estimating mutual information in spatiotemporal data sets. *J. Stat. Phys.* 124:1275–1315.
33. Botev, Z. I., J. F. Grotowski, and D. P. Kroese. 2010. Kernel density estimation via diffusion. *Ann. Stat.* 38:2916–2957.
34. Li, S., Y. Xiao, …, D. Cai. 2018. Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information. *Phys. Rev. E.* 97, 052216.
35. Li, S., J. Xu, …, D. Cai. 2017. The characterization of hippocampal theta-driving neurons-A time-delayed mutual information approach. *Sci. Rep.* 7:5637.
36. Han, H., J. W. Cho, …, I. Lee. 2018. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46:D380–D386.
37. Nueda, M. J., S. Tarazona, and A. Conesa. 2014. Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics.* 30:2598–2602.
38. Meryet-Figuiere, M., B. Alaei-Mahabadi, …, C. Kanduri. 2014. Temporal separation of replication and transcription during S-phase progression. *Cell Cycle.* 13:3241–3248.
39. Aminikhanghahi, S., and D. J. Cook. 2017. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* 51:339–367.
40. Chin, C. H., S. H. Chen, …, C. Y. Lin. 2014. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8:S11.
41. Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28:27–30.
42. Jassal, B., L. Matthews, …, P. D'Eustachio. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res.* 48:D498–D503.
43. Kamburov, A., U. Stelzl, …, R. Herwig. 2013. The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Res.* 41:D793–D800.
44. Samuel, S., and M. D. Sitrin. 2008. Vitamin D's role in cell proliferation and differentiation. *Nutr. Rev.* 66:S116–S124.
45. Gérard, C., and A. Goldbeter. 2009. Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. *Proc. Natl. Acad. Sci. USA.* 106:21643–21648.
46. Hu, Z., P. J. Killion, and V. R. Iyer. 2007. Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.* 39:683–687.
47. Babu, M. M., N. M. Luscombe, …, S. A. Teichmann. 2004. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14:283–291.
48. Qiu, X., A. Rahimzamani, …, S. Kannan. 2020. Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe. *Cell Syst.* 10:265–274.e11.
49. Specht, A. T., and J. Li. 2017. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics.* 33:764–766.
50. Qiu, X., Y. Zhang, …, J. S. Weissman. 2022. Mapping transcriptomic vector fields of single cells. *Cell.* 185:690–711.e45.
51. Pradervand, S., M. R. Maurya, and S. Subramaniam. 2006. Identificaton of signaling components required for the prediction of cytokine release in RAW 264.7 macrophages. *Genome Biol.* 7:R11.
52. Gupta, S., M. R. Maurya, and S. Subramaniam. 2010. Identification of crosstalk between phosphoprotein signaling pathways in RAW 264.7 macrophage cells. *PLoS Comput. Biol.* 6:e1000654.
53. Bonneau, R., D. J. Reiss, …, V. Thorsson. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* 7:R36.