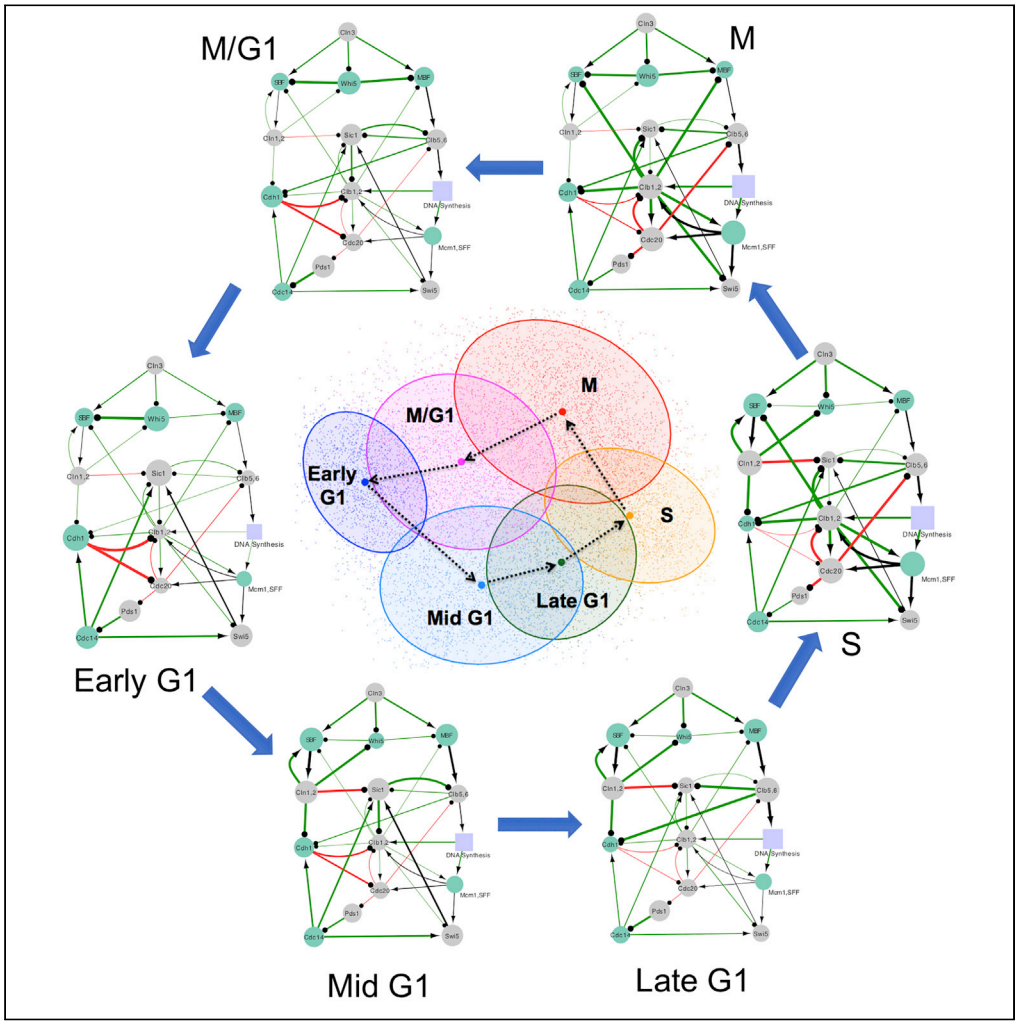
iScience

CellPress
OPEN ACCESS

Article

Random Parametric Perturbations of Gene Regulatory Circuit Uncover State Transitions in Cell Cycle



Ataur Katebi,
Vivek Kohar,
Mingyang Lu

mingyang.lu@jax.org

HIGHLIGHTS

A core gene regulatory circuit encodes irreversible cell cycle state transitions

Statistical analysis on randomly generated ODE models elucidates circuit dynamics

Steady states specify cell cycle phases, and oscillatory states, their transitions

State transition patterns inferred by delayed correlation and parameter perturbation

Katebi et al., iScience *23*, 101150
June 26, 2020
https://doi.org/10.1016/j.isci.2020.101150

iScience

CellPress
OPEN ACCESS

Article

Random Parametric Perturbations of Gene Regulatory Circuit Uncover State Transitions in Cell Cycle

Ataur Katebi,[1] Vivek Kohar,[1] and Mingyang Lu[1,2,*]

SUMMARY

Many biological processes involve precise cellular state transitions controlled by complex gene regulation. Here, we use budding yeast cell cycle as a model system and explore how a gene regulatory circuit encodes essential information of state transitions. We present a generalized random circuit perturbation method for circuits containing heterogeneous regulation types and its usage to analyze both steady and oscillatory states from an ensemble of circuit models with random kinetic parameters. The stable steady states form robust clusters with a circular structure that are associated with cell cycle phases. This circular structure in the clusters is consistent with single-cell RNA sequencing data. The oscillatory states specify the irreversible state transitions along cell cycle progression. Furthermore, we identify possible mechanisms to understand the irreversible state transitions from the steady states. We expect this approach to be robust and generally applicable to unbiasedly predict dynamical transitions of a gene regulatory circuit.

INTRODUCTION

One of the most significant challenges in biology is to elucidate the function and control mechanism of gene regulatory circuits (GRCs) that govern the decision-making of biological processes (Gerstein et al., 2012; Schwikowski et al., 2000). The rapid development of new genomics technology has allowed us to address this question by both experimental and computational systems biology approaches (Farrell et al., 2018; Moignard et al., 2015; Saelens et al., 2019). One of the critical components among them is to use mathematical modeling to simulate the dynamical behavior of GRCs (Balleza et al., 2008; Ciliberti et al., 2007; Sanchez-Osorio et al., 2013). A good model not only provides a mechanistic view of the system but also makes new predictions on the effect of a genetic perturbation like the knockdown and overexpression of one or more genes. Although mathematical modeling approaches have been successfully applied to analyze gene circuits in select cases (Csikász-Nagy et al., 2006; Goldbeter et al., 2001; Hong et al., 2012; Li and Wang, 2014; Li et al., 2004), they typically suffer from a long-standing problem inherent in almost every systems biology study—it is difficult to directly measure most circuit kinetic parameters, especially *in vivo*. Although some parameter values can be learned from published results, many others are often based on educated guesses, therefore, limiting the predictive power of mathematical modeling. The problem is amplified when circuit modeling is applied to large gene networks because such systems typically have many parameters, making conventional methods prone to over-fitting.

To address this issue, we have recently developed a mathematical modeling algorithm, named *random circuit perturbation* (RACIPE) (Huang et al., 2018a), which captures the dynamics of a GRC without the need for precise kinetic parameters (Huang et al., 2017). Instead of fine-tuning a model with a specific set of parameters, RACIPE generates an ensemble of models based on the associated chemical rate equations with distinct random kinetic parameter sets. The random models can represent the GRC in different signaling states and/or microenvironments. The simulated expressions from the ensemble of models are then subjected to statistical analysis to identify robust features of the GRC. Particularly, in several applications to simple circuit motifs and biological circuits (Chen et al., 2004; Csikász-Nagy et al., 2006; Huang et al., 2017; Jia et al., 2019a; Kohar and Lu, 2018), we found that even though the kinetic parameters are largely randomized, the stable steady-state solutions from these models form robust clusters of states that can be

[1]The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

[2]Lead Contact

*Correspondence: mingyang.lu@jax.org

https://doi.org/10.1016/j.isci.2020.101150

associated with distinct cellular states. This finding suggests that we can use the circuit topology (i.e., the connectivity and the types of regulatory interactions) as the only input and unbiasedly predict the cellular states that are allowed by the circuit. Without the need to refine a particular set of parameters, RACIPE can be extended to simulate a large gene network. In essence, RACIPE uniquely converts the nonlinear dynamics problem to a statistical data analysis problem, an approach which we believe is critical for analyzing large network systems.

So far, RACIPE has been applied, in several cases (Bocci et al., 2019; Huang et al., 2017; Jia et al., 2019b; Kohar and Lu, 2018), to understand the possible gene expression steady states and the roles of genes in the behavior of the GRCs. It remains unclear how efficient the ensemble-based approach is in dissecting the dynamical properties of a circuit, e.g., transitions between the cellular states. An archetypal system to assess RACIPE is the gene regulation of cell cycle progression because of the following reasons. First, cell cycle circuit is evolutionarily robust, as it is required by almost every cell type and organism for survival. With RACIPE, we can perturb the system to evaluate the robustness of a cell cycle GRC. Second, cell cycle dynamics involves transitions across a series of well-studied cell cycle states/checkpoints, which allows us to test whether RACIPE can capture these non-trivial dynamics using only random models. Third, there are many existing experimental evidences and data in the studies of cell cycle, which will help to validate the computational models. Fourth, a fundamental understanding of the regulatory mechanism of cell cycle will facilitate researches in areas such as cell differentiation during normal development (Dalton and Coverdell, 2015; Jakoby and Schnittger, 2004; Li and Kirschner, 2014; Myster and Duronio, 2000; Ruijtenberg and van den Heuvel, 2016; Soufi and Dalton, 2016) and tumorigenesis in cancer (Collins et al., 1997; Icard et al., 2019; Sandal, 2002). Fifth, although there are many existing computational models for studying cell cycle networks, different models usually emphasize different aspects of the system. For example, when studying a specific checkpoint, Tyson et al. proposed a two-state switching model (Barik et al., 2010; Tyson and Novak, 2014); when studying a full cycle of the progression, multiple models considered its dynamics as oscillations (Csikász-Nagy et al., 2006; Gerard and Goldbeter, 2009); yet another popular model suggested the cell cycle progression as a downhill process toward the G1 global attractor (Li et al., 2004). A RACIPE analysis could provide a holistic view of the dynamical properties of a cell cycle circuit.

In this study, we analyzed a cell cycle GRC of budding yeast with RACIPE, from which we elucidated the gene regulation during cell cycle progression and the associated state transitions. To achieve this, we first generalized RACIPE for GRCs with multiple regulatory types, including transcription, signaling (such as phosphorylation), and degradation. To carefully examine the oscillatory dynamics, which were considered as an essential component of the cell cycle dynamics (Goldbeter et al., 2001; Novák and Tyson, 2008; Tyson et al., 2003), we developed a numerical algorithm to systematically detect and characterize limit cycles for a large number of models. The generalized RACIPE allowed us to examine the sequential change of circuit dynamics in cell cycle progression. We focused on discovering the features of network dynamics unbiasedly from this statistical analysis of random models. We found that the RACIPE models of the GRC allow both the stable steady states, which form clusters associated with the cell cycle phases, and the oscillatory states, which direct the irreversible state transitions along the cell cycle progression. The model predictions are largely consistent with gene perturbation data in the literature and recent single-cell RNA sequencing (RNA-seq) data. Furthermore, we explored possible mechanisms to understand the irreversible state transitions directly from the steady states of the random models. We hope that RACIPE will facilitate systems biology network modeling and generate unbiased predictions for complex systems.

RESULTS

There are several existing GRC models for the budding yeast cell cycle in the literature (Battogtokh and Tyson, 2016; Chen et al., 2004; Gerard and Goldbeter, 2009; Li and Wang, 2014; Li et al., 2004). Based on these data, we constructed a GRC of 15 genes and 38 regulatory links (Figure 1A). The majority of the components of the circuit were reported in previous studies (Barik et al., 2016; Li et al., 2004). In addition, we made a few modifications to the circuit, as explained in the following. First, instead of including a node representing the cell size, we modeled Cln3 as the input node and the effect of the cell size by the production rate of Cln3. Furthermore, we added Whi5 and its interactions, as Whi5 was found to play a critical role as the signal integrator for cell cycle commitment (G1/S transition) (Liu et al., 2015; Palumbo et al.,

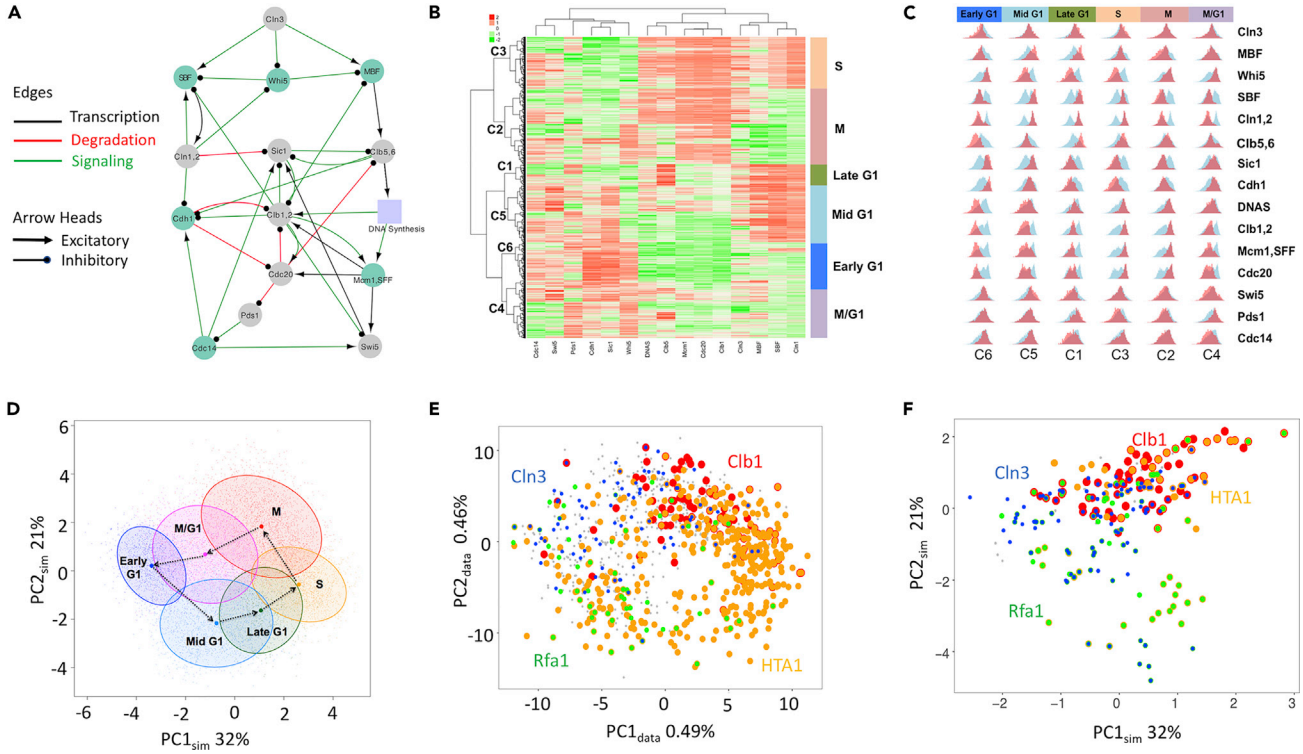


Figure 1. Stable Steady States of the Gene Regulatory Circuit Capture Cell Cycle Phases

(A) The core gene regulatory circuit of the yeast cell cycle. The nodes correspond to 14 genes and an abstract node representing DNA synthesis process, and the edges correspond to 38 regulatory interactions. These interactions can be any of the following three types: transcription (black), degradation (red), and signaling (green), where each type can be either excitatory (line and arrow head) or inhibitory (line and circle head). Nodes colored green are regulated by signaling interactions only. The node DNA sythesis is also denoted as DNAS.

(B) Hierarchical clustering analysis was performed using 12,758 activity profiles from simulated stable steady states. Six clusters of simulated stable steady states are mapped to specific cell cycle phases.

(C) The distribution of the simulated gene activity profiles of each cell cycle phase (red) and that of all the models (blue).

(D) Projection of the simulated activity profiles onto the first two principal components (PCs, $PC1_{sim}$: ~32%, $PC2_{sim}$: ~21%). Each projected cluster that is associated with a cell cycle phase is presented by an ellipse whose center is illustrated by a dot.

(E) Projection of the single-cell RNA-seq data onto the first two PCs ($PC1_{data}$: 0.49%, $PC2_{data}$: 0.46%) obtained from the sequencing data. Cln3 expression peaks at the G1 phase (blue), Rfa1 peaks (green) across the late G1 and S phases, HTA1 peaks (orange) at S, and Clb1 peaks (red) at M.

(F) Projection of the single-cell expression data of 15 representative genes (after KNN smoothing) onto the first two PCs from the simulated data (same as D). See Figures S6 and S7 for the application of another five smoothing/inputation methods.

2016; Qu et al., 2019) and was included in subsequent cell cycle modeling (Barik et al., 2016). Last, we added the inhibition of Cdc20 by Cdh1 as supported by multiple evidences (Hong et al., 2012; Mangla et al., 2010; Okabe and Sasai, 2007; Rudner and Murray, 2000). The final circuit contains transcription factors (SBF, Whi5, MBF, Mcm1, SFF, and Swi5), major cyclins (Cln3, Cln1,2, Clb5,6, and Clb1,2), other regulatory signaling proteins (Sic1, Cdh1, Cdc20, Pds1, and Cdc14), and a node named *DNA Synthesis* (rectangular shape, denoted as DNAS) representing the DNA synthesis process. These nodes regulate one another by transcription (black lines), degradation (red lines), or signaling (green lines) (Figure 1A). Each interaction can be either excitatory (arrow head) or inhibitory (circle head).

To elucidate the dynamics of gene regulation during cell cycle progression, we applied a generalized version of the mathematical modeling algorithm RACIPE to the cell cycle gene circuit (see Section 1-2 and 13 in Transparent Methods for the modeling details). In a previous study, the Boolean network model was applied to a similar circuit (Li et al., 2004). In comparison, our approach (1) models the activity of each gene as a continuous variable, (2) considers the differences in various regulatory types, and (3) systematically characterizes both the stable steady states and the oscillatory states (see Section 3 in Transparent Methods, Figure S8). We generated 10,000 RACIPE models, from which we found 12,758 stable steady states (~84% models) and 3,070 stable oscillatory states (~24% models). About 9% models allow the

coexistence of both dynamical modes, and another 1% not reaching stable states. The convergent tests of the RACIPE analyses are shown in Section 7 in Transparent Methods, Figures S2, and S3.

The Stable Steady States of the Circuit Capture Cell Cycle Phases

We first investigated the stable steady states from the RACIPE models. Here, the simulated activity levels of 15 genes from the 12,758 states were first log-transformed and standardized. The data are then analyzed by unsupervised hierarchical clustering analysis (R function *hclust* with distance metric as *Euclidean* and method *ward.D2*), from which we found six robust clusters, labeled as C1 ~ C6 (Figure 1B). To characterize the biological meaning of these clusters, we compared the expression patterns from the simulations with the literature data as follows. First, we computed, for each gene, the distributions of the simulated activity levels for each cluster (Figure 1B, red histograms) and those of all the data (Figure 1B, blue histograms). Second, we assigned the states of each gene to high, medium, or low by comparing the red and blue histograms (Table S1). Third, we compared the expression pattern of each cluster with the experimental evidence in the literature and associated each cluster to the best matching cell cycle phase (Table S2). The assignment of cell cycle phases for early G1, mid G1, late G1, S, M, and M/G1 is illustrated in Figures 1B and 1C.

Next, we projected the simulated steady states to the first two principal components (PCs) (~32% contribution from PC1 and ~21% from PC2). As shown in Figure 1D, the six clusters of the steady states are arranged circularly. Strikingly, the directionality of the cell cycle progression follows a series of consecutive state transitions from a cluster to the neighboring cluster along the counterclockwise direction. We then compared the simulation results with recently published single-cell RNA-seq data of budding yeast (Jackson et al., 2020). When principal-component analysis (PCA) was performed on the genome-wide gene expression data, we also observed a similar circular structure of gene expression for individual cells (Figure 1E). Representative marker genes Cln3 (Early G1), Rfa1 (Late G1), HTA1 (S), and Clb1 (M) have peak expression in the corresponding cell cycle phases. Subsequently, we projected the expression data of genes associated with the cell cycle GRC to the PCs obtained from the RACIPE simulations. We found that one of the genes (Whi5) in the GRC is not expressed at all and another gene (Cdc14) is not highly expressed in the phases it is supposed to be expressed. In those cases, we replaced the gene with another gene that (1) directly interacts with the original gene and (2) has expression consistent with previously reported microarray data (Spellman et al., 1998). See Table S3 for the detailed mapping. As shown in Figure 1F, the circular structure, although much noisier, can still be discerned. Here, we applied KNN smoothing (Wagner et al., 2018) to the single-cell RNA-seq data to alleviate the dropout effects. We also explored several imputation methods, including MAGIC (van Dijk et al., 2018), scImpute (Li and Li, 2018), SAVER (Huang et al., 2018b), and RESCUE (Tracy et al., 2019). We found that although these methods can capture the circular structure of the genome-wide transcriptomics data, there is no evident improvement over KNN in capturing the structure of the data when projecting using the GRC genes (see Section 8 in Transparent Methods and Figures S4–S7 for details). Furthermore, our finding that cell cycle progression corresponds to the state transitions in a circular structure has also been observed in budding yeast time series microarray data (Ferrezuelo et al., 2010; Spellman et al., 1998) and mouse embryonic stem cell single-cell RNA-seq data (Liu et al., 2017).

The aforementioned mapping of the cell cycle phases helps to identify the dynamical change of the gene activities along cell cycle progression from the RACIPE models. As shown in Figure 1C, the histogram of the activity of some genes (e.g., SBF) is bimodal and the gene switches the level of its activity discretely, whereas some other genes (e.g., Cdh1) have more continuous changes in the activity during the state transitions. Markedly, the three cyclin genes Cln1,2, Clb5,6, and Clb1,2 belong to the former type. Our model predicts that Cln1,2 switches from low to high activity in the mid G1 phase and remains high till the S phase, and then switches back to low activity in the M phase. By constrast Clb5,6 switches to high in the late G1 phase, whereas Clb1,2 switches to high in the S phase. The model predictions on the discrete changes of the cyclin expressions and their specific order are consistent with known experimental observations (Rustici et al., 2004; Spellman et al., 1998).

Perturbation Analysis Identifies the Role of the Genes in Cell Cycle Progression

To further validate the modeling results of the cell cycle circuit, we performed *in silico* perturbation analysis to evaluate the effects of gene knockdown (KD) on the behavior of the circuit. The outcomes of the gene perturbations were then compared with experimental evidence in the literature. Figure 2 shows the modeling results for the untreated (UT), 15 single gene KD, and 12 double gene KD. Here, the effect of

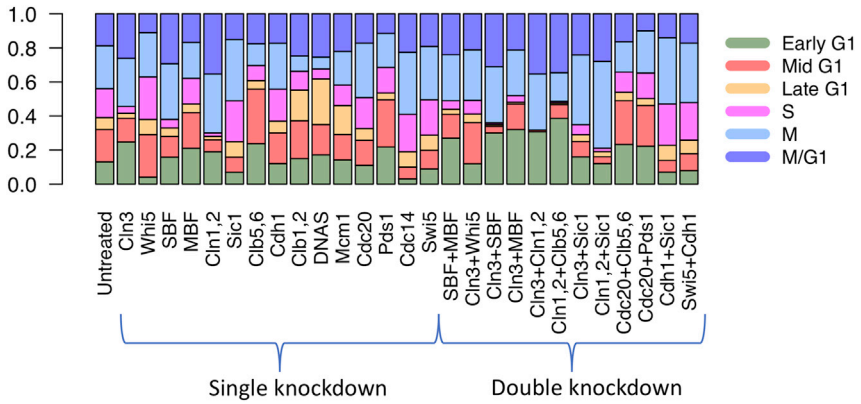


Figure 2. Perturbation Analysis

Distribution of the fraction of steady states in each cell cycle phase from the simulations of the cell cycle circuit under untreated (UT), single knockdown (KD), and double KD conditions. To assign new phases after a perturbation, we trained a neural net model using the activity profiles of the untreated condition and the classification of cell cycle phases from the previous hierarchical clustering analysis (see Section 5 in Transparent Methods for details), and then we infer the phase for any new activity profile using the trained model.

gene KD was simulated by reducing the maximum production rate of the corresponding genes by 95%. For each case, we simulated 10,000 random models by RACIPE and evaluated the distribution of the stable steady states in each cell cycle phase. To rapidly assign the cell cycle phase to every stable steady state, we trained a fast-forward neural network model using the original 10,000 RACIPE models (i.e., the UT condition, the data from Figure 1B). From our benchmark, the neural network model has over 90% accuracy in the state assignment (see Section 5 in Transparent Methods and Figure S26). The results for the KD simulations at different KD levels are shown in Figure S9.

The perturbation analysis identifies the role of each gene in stabilizing or destabilizing various cell cycle phases. The KD conditions wherein one or multiple cell cycle phases are mainly affected are summarized in Table S4. The outcomes of the gene perturbation are well supported by experimental evidence. For example, (1) the modeling results show that Cln3 KD increases the fraction of states in early G1 and M/G1 and decreases that of S. This is consistent with the experimental data where the inactivation of Cln3 was found to cause the cell size to increase, elongating early G1 (Dirick et al., 1995; Stuart and Wittenberg, 1995; Tyers et al., 1993; Wijnen et al., 2002). For the cell cycle START to take place, activation of Cdc28 by G1-cyclins Cln1, Cln2, or Cln3 is necessary (Dirick et al., 1995), and Cdc28 mutants cause a cell to stall at the START of early G1 phase (Hartwell et al., 1974). (2) Whi5 KD increases the fraction of states in mid G1 and S and decreases that in M/G1 and early G1, consistent with the finding that the inactivation of Whi5 dramatically reduces the cell size by shortening early G1 (de Bruin et al., 2004). (3) Interestingly, Whi5 KD seems to have the opposite effect of Cln3 KD. Double perturbation of Cln3 and Whi5 appears to mitigate the impact of each other by restoring the proportions in early G1 and mid G1, which is consistent with experimental findings (de Bruin et al., 2004). (4) Cln1 and Clb5 double KD reduces the proportion of stable states in mid G1 to M and increases that in M/G1 and early G1. This prediction is also supported by the experimental finding that quadruple KD of Cln1,2 and Clb5,6 arrests cells in G1 (Schwob and Nasmyth, 1993). (5) Cdc20 and Pds1 double KD decreases the proportion of states in M/G1, consistent with the findings that double mutant (Cdc20 and Pds1) cells get arrested in telophase (Lim et al., 1998). These experimental evidences are listed in Table S5.

Dynamical Changes of Circuit States along Cell Cycle Progression

As shown in previous sections, stable steady states from RACIPE models can capture the gene activity patterns of various cell cycle phases. Next, we show, from these steady states altogether, how we can recapitulate the dynamic behavior of the cell cycle circuit. We explored several existing pseudotime inference methods, including Angle (Saelens et al., 2019), ElPiGraph Cycle (Albergante et al., 2020), reCAT (Liu et al., 2017), and SoptSC (Wang et al., 2019), to assign pseudotime index to each activity profile of a steady state. However, these methods failed to infer the correct ordering of cell cycle phases, presumably because of the special circular structure and the high variation of the simulated data. Thus, we devised approaches

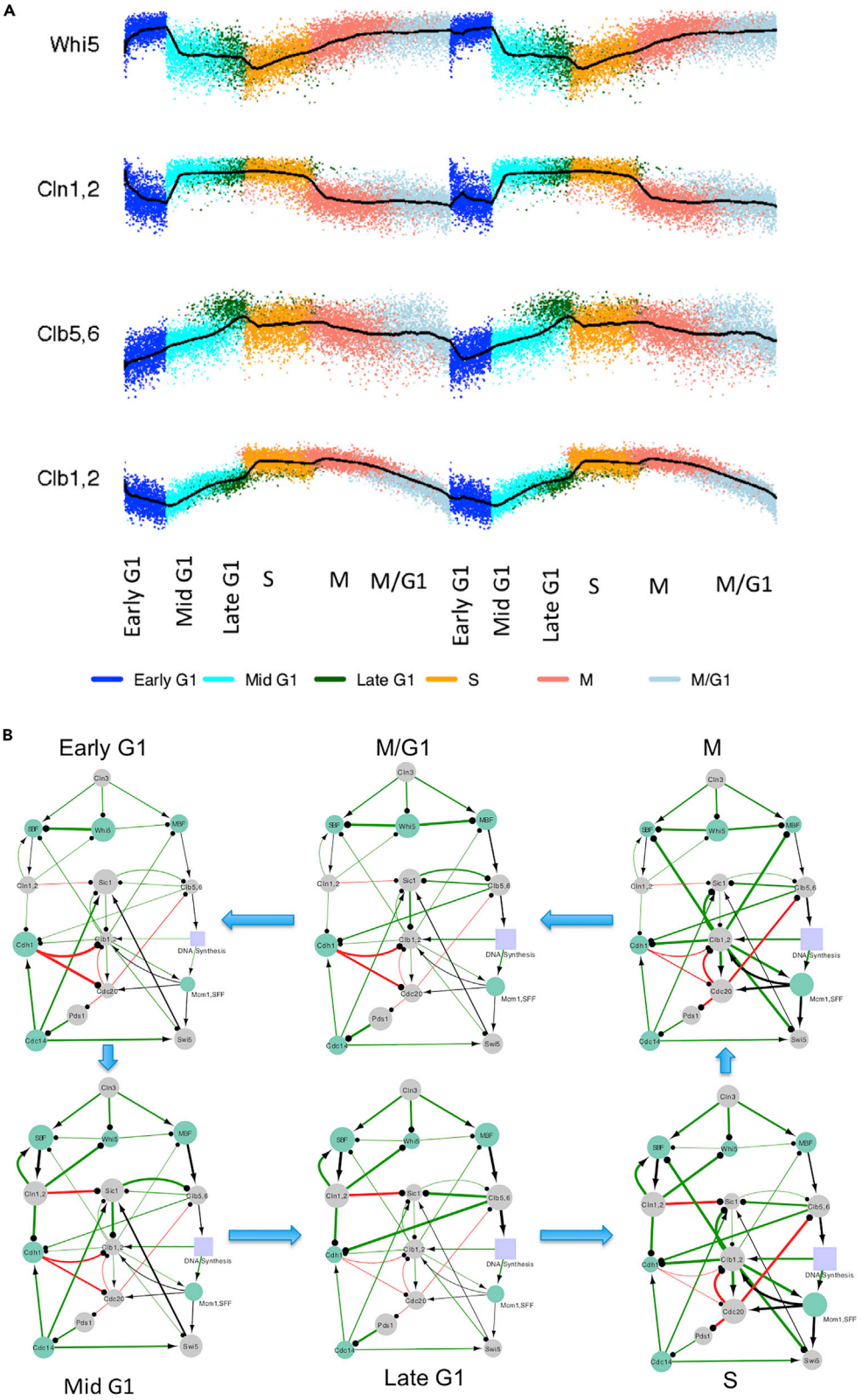

A
Whi5
Cln1,2
Clb5,6
Clb1,2
Early G1
Mid G1
Late G1
S
M
M/G1
Early G1
Mid G1
Late G1
S
M
M/G1
Early G1
Mid G1
Late G1
S
M
M/G1
B
Early G1
M/G1
M
Mid G1
Late G1
S

Figure 3. Dynamic Change of Gene Expressions during Cell Cycle Progression

(A) The activity profiles of Whi5, Cln1,2, Clb5,6, and Clb1,2 along the inferred pseudotime (with two cycles) obtained from the custom use of psupertime method. Colors annotate different cell cycle phases. The activity profiles of all nodes in the cell cycle GRC are shown in Figure S12A.

(B) A dynamic network view of cell cycle progression. The size of each node represents the level of gene activity, and the thickness of the each edge represents the activity of the regulation, defined as the proportion of models wherein the level of the regulator is higher than the threshold level of the edge.

to assign pseudotime using either our results from hierarchical clustering analysis shown in Figure 1B or a custom use of a recent pseudotime inference method named *psupertime* (Macnair and Claassen, 2019) shown in Figure S12; see Section 9 in Transparent Methods and Figures S10–S14 for details of the pseudotime analyses.

The change in gene activity for Whi5 and three cyclins (Cln1,2, the G1 phase cyclin; Clb5,6, the S phase cyclin; and Clb1,2, the M phase cyclin) along the pseudotime (obtained using the custom-use of psupertime) are shown in Figure 3A (see Figure S12A for the other genes). The profiles in Figure 3A are illustrated with two repeats to represent two cycles, and they capture the overall gene activity dynamics well. Interestingly, genes like Cln1,2 exhibit binary activity—Cln1,2 level is high during G1 and low during M and M/G1. However, genes like Whi5 and Clb1,2 exhibit more gradual changes in activity along cell cycle progression. Moreover, Whi5 and Cln1,2 have opposite phases, with Whi5 switching activity slightly ahead of Cln1,2, which is captured by their peaked negative cross-correlation at lag 0 (Figure S12C). Noticeably, Clb5,6 level increases when the Cln1,2 level is still high. When Clb5,6 level decreases, Clb1,2 level starts to rise and continues to stay elevated during M/G1, as its coordination with the other genes (specifically Sic1, Cdh1, Cdc14, and Swi5) is needed for mitotic exit (Prinz Susanne, 1999; Shirayama et al., 1999; Visintin et al., 1998). The synchrony of Sic1 and Cdh1 with Clb1,2 is further observed in their autocorrelations (Figure S12B), emphasizing the role of Clb1,2 in mitotic exit.

Coordinated activation of a subset of genes and their interactions drive the cell cycle from one phase to the next along the cell cycle progression. Using the simulated data from the stable steady states of the 10,000 RACIPE models, we can visualize the dynamic changes of the cell cycle gene circuit (Figure 3B). For each cell cycle phase, the size of a node is scaled to the average activity level of the corresponding gene from all the models in that phase, whereas the thickness of an edge is scaled to the fraction of models where the activity of the regulator is larger than the threshold of the Hill function (see Section 2 in Transparent Methods for the mathematical equations). Remarkably, this representation provides a dynamic view of gene regulation along cell cycle progression. In early G1, Whi5 has high activity, which inhibits SBF, and consequently keeps Cln1,2 activity suppressed. An increased level of Cln3 suppresses Whi5 and enables SBF release. The increased level of SBF activates Cln1,2, which induces the transition to mid G1. Furthermore, high level of Cln1,2 inhibits Whi5 and Sic1, which transitions the cell to enter late G1. This positive feedback of Cln1,2 behaves like a switch. This feedback mechanism is known as the START checkpoint, whose activation makes the cells commit to the S phase (DNA synthesis) (Johnston et al., 1977; Skotheim et al., 2008). In contrast to models wherein one network is associated with one state, here we show that a dynamic network model can describe the sequential changes of the gene regulation in all the cell cycle phases.

Oscillatory Dynamics Specify the Directionality of the Cell Cycle State Transitions

In the previous sections, we analyzed the steady states of RACIPE models and associated them with various cell cycle phases. However, the analysis so far does not explain the direction of transitions between these phases. Here, we explore how the state transitions can be understood from the oscillatory states of RACIPE models.

To facilitate the statistical analysis on oscillations, we obtained 47,637 oscillatory states from 1,60,000 RACIPE models (10,000 models from the prior analysis plus 1,50,000 models generated anew). For each oscillation, we projected the trajectory to the first two PCs of the steady-state data (same PCs as in Figure 1D). Some of the projected trajectories reside in one or a very few cell cycle phases, whereas some others can span many phases (Figure 4A). To characterize the transitions through various cell cycle phases along an oscillatory trajectory, we used the previously trained neural network model to predict the cell cycle phase for any given gene activity profile along an oscillatory trajectory. This allows us to identify the patterns of state transitions for each oscillation. For example, two oscillations are shown in Figure 4A: the one on

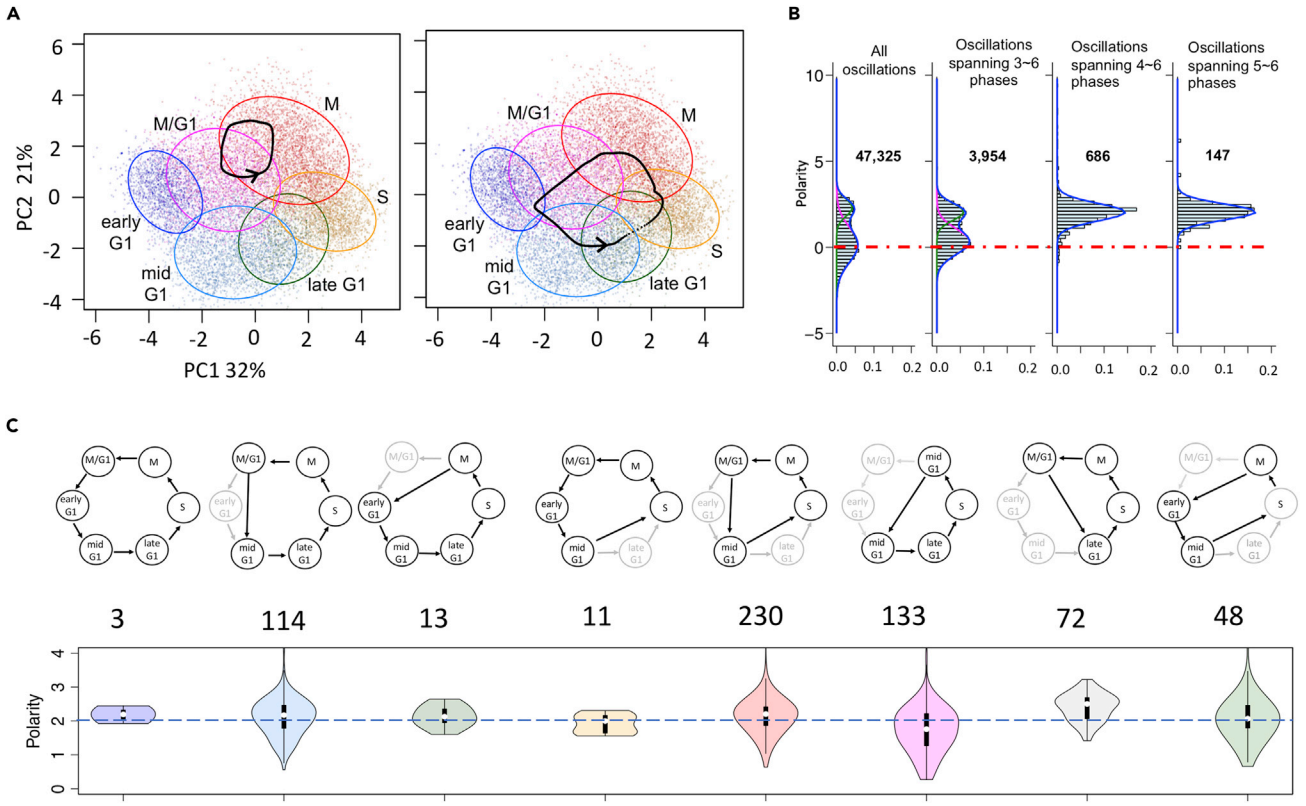


Figure 4. Statistical Analysis of Oscillatory Trajectories Elucidates Cell Cycle Dynamics

(A) Examples of the oscillatory trajectories projected on the first two principal components (PC1 and PC2) obtained from the stable states from 10,000 models. Left panel: Trajectory traveling only two cell cycle phases; right panel: Trajectory traveling all six cell cycle phases. An arrow on the projected trajectory marks the direction of oscillation.

(B) The histogram of the polarity of oscillatory trajectories of different sizes—all trajectories (n = 47,327) and trajectories spanning 3–6 phases (n = 3954), 4–6 phases (n = 686), and 5–6 phases (n = 147). The histograms are fitted by either two Gaussian distributions (first two columns, where the green curve is for the distribution with a positive mean and pink curve is for the distribution with a near zero mean) or a single Gaussian distribution (the last two columns). The horizontal dotted red line shows the zero polarity.

(C) Top panel shows eight major state transition patterns of the oscillatory trajectories. The number below each pattern indicates the number of observed oscillations. Bottom panel shows the histogram of polarity of each pattern.

the left shuffles back and forth between two phases (M → M/G1 → M), whereas the one on the right passes through all the phases (early G1 → mid G1 → late G1 → S → M → M/G1 → early G1). See Section 3 in Transparent Methods for details on detecting oscillatory states.

To quantify the directions of the oscillatory trajectories, we defined an angular-velocity-based metric, named *polarity*, to measure whether a projected oscillatory trajectory is clockwise or counterclockwise (see Section 4 in Transparent Methods). A positive polarity corresponds to a counterclockwise trajectory. Interestingly, the distribution of the polarity for all the projected trajectories is bimodal (Figure 4B, leftmost panel), which can be fit nicely by two Gaussian distributions, one with mean zero (pink curve) and the other with positive values for almost the whole distribution (green curve). However, if we considered the trajectories spanning four or more cell cycle phases, the distribution of the polarity can be fit by a single Gaussian distribution with only positive values (Figure 4B, two rightmost panels). This result suggests that the polarity for the trajectories of small size (i.e., spanning no more than three phases) is random, presumably because there is no restraint for the oscillations to travel. However, the polarity for the trajectories of large size (i.e., spanning four or more phases) is mostly positive. These oscillations only travel counterclockwise, consistent with the known sequence of cell cycle progression. Note that we found this consistent pattern of state transitions from randomly generated models because the gene circuit topology restrains the possible routes of state transitions. Figure 4C lists eight major recurring transition patterns involving four or more phases with

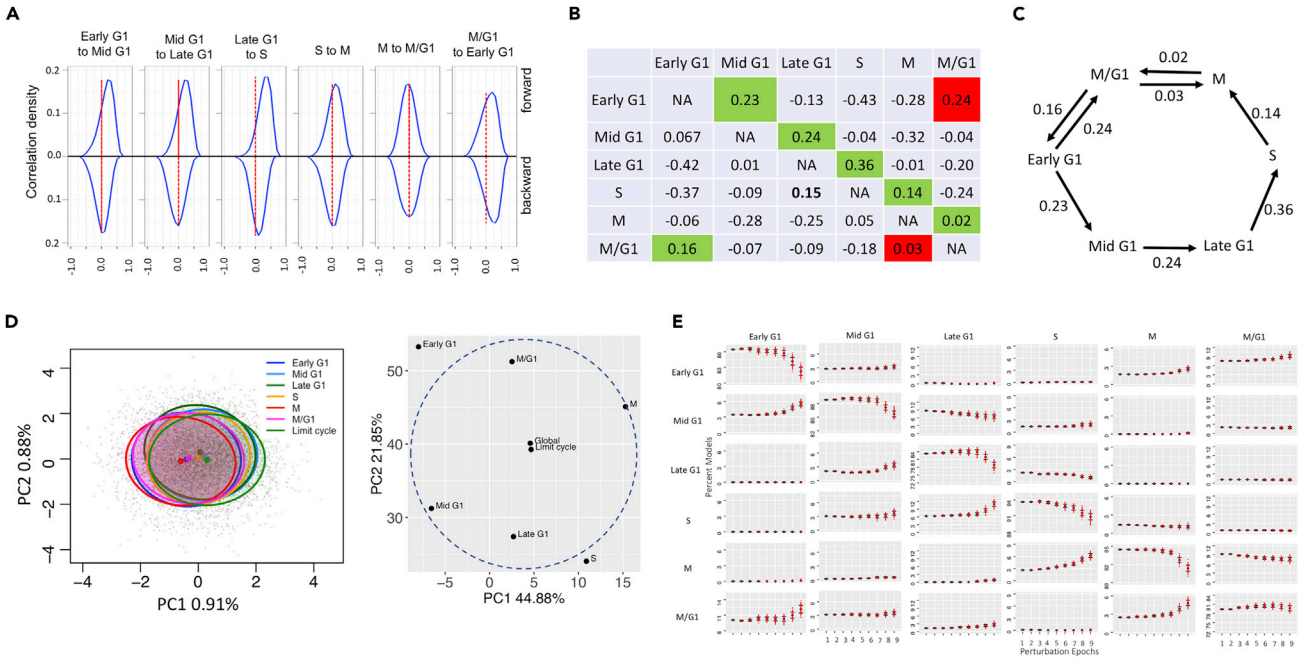
positive polarity, i.e., all the corresponding oscillations travel in a counterclockwise direction. The polarity distributions of the other transition patterns are shown in Figure S15 and Table S7.

Cell Cycle Gene Circuit Specifies Irreversible Phase Transitions

So far, we have shown that the cell cycle phases can be described by the stable steady states from the RACIPE models, whereas the directionality of the cell cycle progression can be deduced from the oscillatory states. Nevertheless, there is a remaining issue in this theory of directionality of cell cycle progression as follows. RACIPE models with random kinetic parameters (intrinsic factors for the regulatory events associated with the GRC genes) can be regarded as heterogeneous cells under different signaling or environmental conditions (extrinsic factors for regulatory events that are important to cellular functions but are not described explicitly in the small GRC, e.g., the copy number of RNA polymerase). This view is consistent with the following previous studies: (1) extrinsic factors, such as the copy number of RNA polymerase, have been shown to contribute to cell-to-cell variability and gene expression noise (Jones et al., 2014); (2) cell-to-cell heterogeneity has also been described by a population of single-cell models with different parameter values (production and degradation rates, in Llamosi et al., 2016). Thus, it makes sense that different models can correspond to different cell cycle phases. However, if the signaling conditions can be altered reversibly, the state transitions should also be reversible, which is not consistent with the fact that the cell cycle progression is irreversible. To address this issue, we propose the following two analyses to explain the irreversible state transitions in the cell cycle using the stable steady-state solutions from the RACIPE models.

First, we show in the following that the irreversible state transitions can be inferred using the steady-state gene activity profiles and the topology of the cell cycle gene circuit. The approach is based on the fact that the activity of the targets lags behind the activity of the regulators. Thus, for a given transition from state 1 to state 2, the activity correlation of the regulators in state 1 and the targets in state 2 should be higher than the activity correlation of the regulators in state 2 and the targets in state 1. Therefore, we evaluate a so-called *delayed correlation* for the forward and backward transitions between state 1 and state 2 (see Section 6 in Transparent Methods). If the median delayed correlation is significantly different between the forward and backward directions, the state transition is likely to be irreversible. As shown in Figures 5A, 5B, and S16, the delayed correlations for the forward transitions of consecutive cell cycle phases are usually higher than those for the backward transitions, except for the M to M/G1 and M/G1 to early G1 transitions where both forward and backward correlations are comparable. Based on these delayed correlations, we derive a state transition model for cell cycle progression (Figure 5C). The transition paths are overall irreversible along the direction of cell cycle progression, and the two bidirectional paths could be related to the START point of cell cycle (Kraikivski et al., 2015). Note that this approach is generally applicable to infer the paths of state transitions for a GRC, examples of which are shown in Figure 6 for a repressilator circuit (see Section 10 in Transparent Methods) and an induced toggle switch circuit (see Section 11 in Transparent Methods).

Second, we propose a putative model to explain how irreversible state transitions occur by changing the kinetic parameters. In the analysis, we consider seven groups of models—each of the first six groups corresponds to the models that allow only one stable steady state of the corresponding cell cycle phase and the seventh group corresponds to the models that generate oscillatory state spanning three or more cell cycle phases. We first performed PCA on the high-dimensional kinetic parameter profiles from the selected RACIPE models (# models—early G1: 635, mid G1: 689, late G1: 365, S: 721, M: 1,304, M/G1: 805, and limit cycles: 3,899). Unlike the gene activity profiles, the parameters from the different cell cycle phases and the oscillatory states overlap, and the magnitude of the parameters is not significantly different from each other (Figure 5D, left panel). However, if PCA is performed on the average parameter profiles of each group, the projected parameter profiles follow a similar spatial arrangement as the projected gene activity profiles (Figure 5D right panel). Interestingly, the average projected parameter profile for the oscillatory states is near the center and at the same time, close to the average projected parameter profile of all the models.This pattern persists when we considered only individual type of parameters (Figure S17). Next, starting from the models of a cell cycle phase, we perturbed the kinetic parameters and evaluated the changes in the model proportions across the cell cycle phases. We found that if the parameters are perturbed toward the global average, the models shift toward the next cell cycle phase for most cases, except for early G1 (Figures 5E and S18), in which models are increased in both mid G1 and M/G1. Similar observations were found if we perturbed the parameters toward the



	Early G1	Mid G1	Late G1	S	M	M/G1
Early G1	NA	0.23	-0.13	-0.43	-0.28	0.24
Mid G1	0.067	NA	0.24	-0.04	-0.32	-0.04
Late G1	-0.42	0.01	NA	0.36	-0.01	-0.20
S	-0.37	-0.09	**0.15**	NA	0.14	-0.24
M	-0.06	-0.28	-0.25	0.05	NA	0.02
M/G1	0.16	-0.07	-0.09	-0.18	0.03	NA

Figure 5. The Irreversible Direction of the Cell Cycle Inferred from Stable Steady States

(A–C) A model for cell cycle progression inferred from the delayed correlations. (A) Densities of the forward (top) and backward (bottom) delayed correlations of gene expressions between successive cell cycle phases. Densities for all possible pairs are shown in Figure S16. The vertical dotted red line along zero correlation helps identify the skewness of the distributions. (B) Median delayed correlations for every two cell cycle phases. Medians along the direction of the cell cycle are predominantly larger (green), except for M to M/G1 and M/G1 to early G1 where the backward correlations are also comparable (red). (C) A model for cell cycle progression derived from the delayed correlations.

(D and E) A scheme for cell cycle progression inferred from parameter perturbation. (D) Left panel shows the parameter vectors of every one-state model in each cell cycle phase and oscillatory state projected on the first two principal components in the parameter space. Right panel shows the averages of the parameter vectors of all the one-state models in each cell cycle phase and oscillatory state. (E) Mean and standard deviation (SD) of the percent models in each cell cycle phase when simulated with the parameters perturbed toward the global parameter average (shown as mean ± SD). Perturbed models from one cell cycle phase diffuse to the next cell cycle phase with one exception, early G1. In early G1, the perturbed models diffuse to both mid G1 and M/G1.

average of the oscillatory states (Figure S19), or toward the parameters of any randomly selected oscillatory model (Figure S20). Note that the pattern of diffusion of the models along the cell cycle direction is not observed when the kinetic parameters are perturbed randomly (Figure S21). For details of the parameter perturbation, see Section 12 in Transparent Methods. Our modeling results demonstrate that the irreversible state transition during the cell cycle progression can be achieved by relaxing the kinetic parameters toward the global average, the average of the oscillatory states, or a random model of the oscillatory states. This approach has also been successfully applied to a repressilator (details are in Section 10 in Transparent Methods, Figures S22–S24).

In short, we have shown through two different approaches how irreversible state transitions can be derived from the steady states of random models. Our simulation approach also sheds light on possible mechanisms to achieve irreversible state transitions by changing the signaling states (represented here by the kinetic parameters) of the circuit.

DISCUSSION

In this study, we used the budding yeast cell cycle as a model system to elucidate dynamic behaviors of GRCs using a computational systems biology approach. We generalized and applied our recently developed mathematical modeling algorithm, named *random circuit perturbation* (RACIPE), to analyze the dynamic features of a budding yeast cell cycle gene circuit using an ensemble of chemical rate equation models with random kinetic parameters. From an extensive statistical analysis on a large set of simulated data, we found that the clusters of steady-state gene activities can be associated with specific cell cycle phases, whereas the oscillatory states specify the path of the state transitions

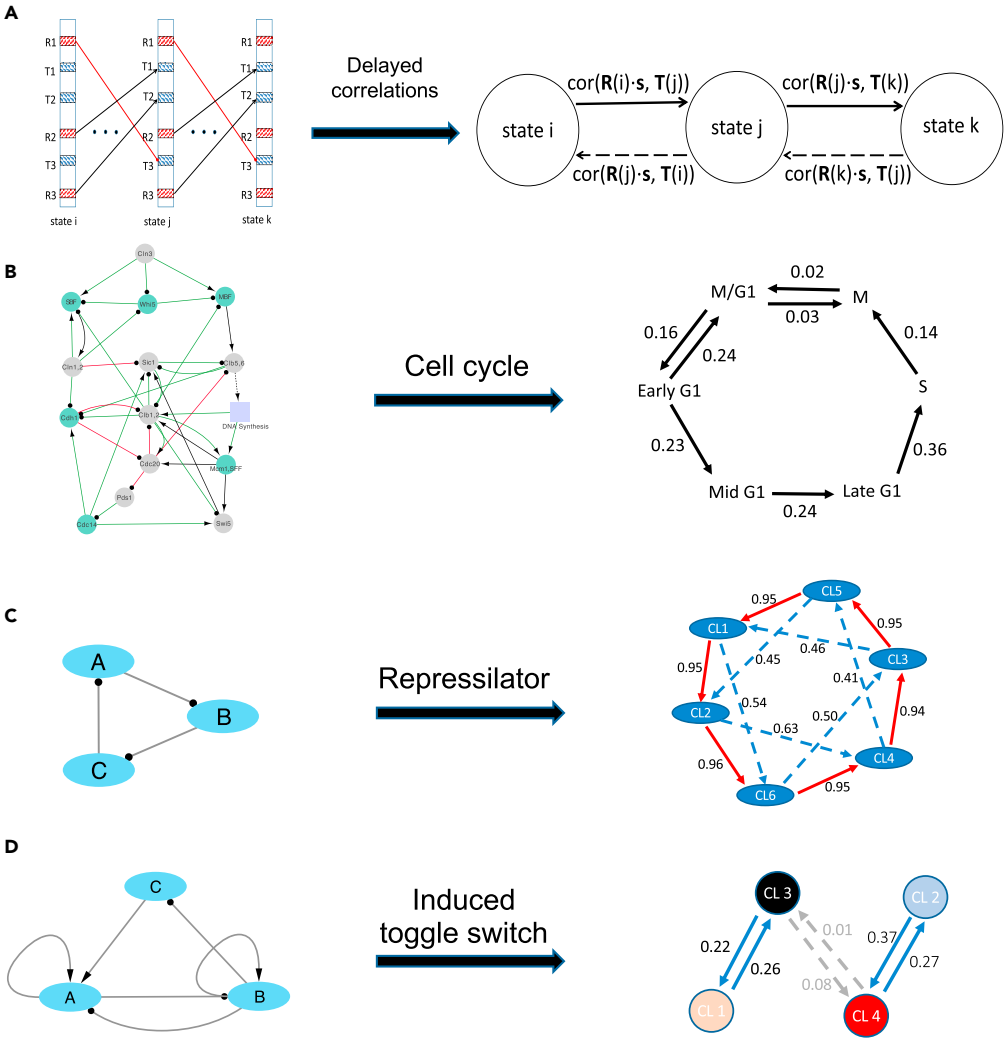


Figure 6. A Method to Predict State Transitions Using the Steady-State Gene Activity Profiles of RACIPE Models and the Circuit Topology

(A) Schematic of the method. Left panel shows three expression/activity vectors of three states i, j, and k, respectively. A black line with an arrow head indicates an excitatory link, whereas a red line with a circle head indicates an inhibitory link, and s(links) is the sign vector expressing each interaction type (+1: excitatory, −1: inhibitory). Right panel illustrates the definition of the delayed correlation for forward transitions (thick black line pointed right) and backward transitions (black dotted line pointed left). Panels B to D show three applications of the delayed correlations to infer state transitions. (B) The cell cycle circuit (left) and delayed correlation-based transitions between six cell cycle phases (right). (C) A repressilator circuit (left) and predicted transitions between six clusters (CL1 ~ CL6) obtained from the RACIPE models. Details are in Figure S22. (D) A three-node induced toggle switch circuit (left) and predicted transitions between four clusters (CL1 ~ CL4). Details are in Figure S25.

during the cell cycle. These cell cycle-specific clusters from circuit simulations are consistent with budding yeast single-cell RNA-seq data. Finally, we elucidate the mechanism of the irreversible state transitions using RACIPE-generated models. Our analyses demonstrate the role of the circuit topology in determining the cell cycle dynamics and its function. The unique combination of modeling and statistical data analysis can be a powerful approach to analyzing the behavior of complex dynamical systems in biology. Note that, because of the evolutionary conservation, most of the genes and interactions in the circuit of the budding yeast cell cycle have counterparts in mammalian systems (Table S6) (Bähler, 2005; Csikász-Nagy et al., 2006). Thus, the analysis, although focused on budding yeast, can be extended to study the mammalian cell cycle.

Cell cycle is an ideal testing ground for mathematical modeling, because (1) there are rich datasets and literature on gene regulation in cell cycle, (2) substantial efforts have been made to model the dynamics of cell cycle GRCs, and (3) cell cycle involves transitions of multiple cell cycle phases. Thus, the dynamics are non-trivial. Elucidating cell cycle gene regulation can contribute to a better understanding of the role of the cell cycle in sundry biological processes, such as cell differentiation (Dalton and Coverdell, 2015; Jakoby and Schnittger, 2004; Li and Kirschner, 2014; Myster and Duronio, 2000; Ruijtenberg and van den Heuvel, 2016; Soufi and Dalton, 2016), DNA damage (Branzei and Foiani, 2008; Hustedt and Durocher, 2017; Murray and Carr, 2018; Shaltiel et al., 2015) and metabolism (Kalucka et al., 2015; Kaplon et al., 2015; Roy et al., 2017), and in diseases, such as cancer (Collins et al., 1997; Icard et al., 2019; Sandal, 2002). Compared with previous modeling studies (Chen et al., 2004, 2000; Csikász-Nagy et al., 2006; Gerard and Goldbeter, 2009; Li and Wang, 2014; Li et al., 2004; Lv et al., 2015), we have generated a large number of models and performed extensive data analysis on all these models. Our results illustrate that models with different sets of parameters can be associated with either one of the cell cycle phases and checkpoints, or the oscillatory dynamics across multiple cell cycle phases. Strikingly, we show that the cell cycle dynamics can be well explained even without carefully selected kinetic parameters, again demonstrating the crucial role of the cell cycle gene circuit topology, other than the detailed kinetic parameters, in determining the circuit's function.

To study the cell cycle gene circuit, we have improved the original RACIPE method (Huang et al., 2017) and further developed statistical methods to analyze the simulated data from a large set of models. Primarily, RACIPE is generalized to model three additional types of regulation (phosphorylation, signaling, and degradation) in addition to transcriptional regulation. Furthermore, the method is extended for high throughput analysis of not only stable steady states but also oscillatory dynamics. To carefully characterize different dynamical features from many models, we developed numerical algorithms to automatically detect oscillatory trajectories and characterize the state transition patterns using machine learning. The advances in RACIPE enable us to identify robust dynamical features of the cell cycle directly from randomly generated models.

An important question we focused on is to understand how the cell cycle gene circuit controls the irreversible state transitions. As shown in our simulated data, we have investigated it from three different aspects. First, we systematically mapped the patterns of state transitions from all the RACIPE-generated oscillatory trajectories. From this, we identified not only a consistent directionality of the state transitions but also recurring transition patterns. The statistical analysis indicates the crucial role of the circuit topology, instead of a specific set of kinetic parameters, in determining the irreversible nature of the state transitions during the cell cycle. Second, we developed a numerical method to infer the directions of state transitions directly from the steady-state gene expression/activity profiles and the circuit topology. Here we achieved this by considering that the expression of the targets occurs later than the activation of the regulators and employed a metric based on delayed correlations. We have applied this approach on several classic GRCs, including the cell cycle circuit, and successfully reproduced the patterns of state transitions. The success of this method demonstrates again how GRC drives state transitions. Third, we examined the relationship between the cell cycle progression and the signaling states of the GRC by perturbing the model parameters belonging to each cell cycle phase and assessing how the models diffuse to the other phases. We perturbed the parameters either toward the global parameter average (i.e., making the parameters more balanced) or relaxed them toward the limit cycle parameters. We found that the perturbed models transition to the next cell cycle phase in both cases. This finding suggests that an irreversible state transition of the cell cycle would occur by systematically changing the signaling state of the circuit.

Overall, we have shown the power of RACIPE in gene circuit modeling. The paths of the cell cycle state transitions that we inferred from either the delayed correlations or the parameter perturbations suggest that an ensemble of random models provides a holistic view of the dynamical behaviors identified from previous cell cycle models, including state switching (Chen et al., 2004, 2000; Csikász-Nagy et al., 2006), oscillations (Csikász-Nagy et al., 2006; Gerard and Goldbeter, 2009; Li and Wang, 2014; Lv et al., 2015), and G1 global attractor (Li et al., 2004). In particular, irreversible state transitions during cell cycle progression have been elucidated in previous studies by various approaches, including those using the directionality of an oscillatory trajectory (Chen et al., 2004, 2000; Csikász-Nagy et al., 2006), phase transitions in a bifurcation from steady states to oscillatory states (Csikász-Nagy et al., 2006; Gerard and Goldbeter, 2009), and the directionality of flux in a landscape view of a detailed cell cycle circuit (Li and Wang, 2014) and a reduced state network model (Lv et al., 2015). In contrast, our approach does not require specifying the kinetic parameters

of the GRC model but derives the state transition patterns purely from statistical analysis of an ensemble of ordinary differential equation (ODE) models with randomly generated kinetic parameters. Using an extensive statistical analysis of the oscillatory states from a large number of random models, we found that the irreversible directionality of cell cycle state transitions was only observed in large oscillatory trajectories, whereas the directionality is random for small ones. Moreover, using the two new analyses based on delayed correlations and parameter perturbations, our approach explains not only the global irreversible state transitions along cell cycle progression but also the local reversible transitions before the start checkpoint. Our combined mathematical modeling and statistical analysis strengthens the view that cell cycle progression is driven by interplay between steady states and oscillatory dynamics.

Limitations of the Study

We propose a generalized mathematical modeling approach to computing the dynamics of a GRC and an analytical framework to find the states of the GRC and the direction of the state transitions. Although we did not use a detailed description in the mathematical formula for modeling the dynamics of a GRC (see Equation S1 ~ Equation S7 in Section 2 in Transparent Methods), the modeling still works reasonably well in finding the states of cell cycle GRC and a few other circuits. There are still several issues that need to be addressed in the future. First, the state assignment is currently based on the hierarchical clustering analysis of the steady state activity profiles from the random models. The quality and robustness of the clustering might affect the performance of the subsequent statistical analysis. For instance, we have noticed a few distinct features for the M/G1 phase, which might suggest a hybrid nature of the phase. From the performance of neural net model in assigning phases, we observed that the M/G1 phase was incorrectly assigned to the early G1 and M phases, each being at 7.59% and 6.62%, respectively, whereas the early G1 phase was incorrectly assigned to M/G1 by 11.99%. This state assignment problem for M/G1 suggests that the gene activities of M/G1 are similar to those of both the early G1 and M phases. In addition, we know that the mitotic exit network gets activated during the M/G1 phase (Prinz Susanne, 1999; Shirayama et al., 1999; Visintin et al., 1998), which makes the activity profile of the dividing cells very similar to that of early G1 (shown in Figure 3B). Second, although we have tested several cell cycle circuits with minor variations, a systematic approach for circuit refinement is needed for better modeling. One potential approach is to apply RACIPE on different versions of a circuit by adding/removing genes/links and analyzing the circuits iteratively. During the process, the modeling results from different circuits can be compared against literature or genomics data (such as gene expression data). Thus, RACIPE can not only be used for modeling existing gene circuits but also can be utilized in circuit refinement.

Resource Availability

Lead Contact

Mingyang Lu, Assistant Professor, The Jackson Laboratory, Bar Harbor, ME 04609, Email: mingyang.lu@jax.org

Materials Availability

This study does not generate any new unique reagents.

Data and Code Availability

The generalized RACIPE is available on GitHub at https://github.com/arkatebi/cellcycle.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101150.

ACKNOWLEDGMENTS

The study is supported by a startup fund from The Jackson Laboratory, by the National Cancer Institute of the National Institutes of Health under Award Number P30CA034196, and by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM128717.

AUTHOR CONTRIBUTIONS

A.K. performed the experiment, interpreted and analyzed the data, and wrote the manuscript. V.K. performed the experiment, interpreted and analyzed the data and wrote the manuscript. M.Y. designed the experiment, interpreted and analyzed the data, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interest.

Received: November 4, 2019
Revised: March 5, 2020
Accepted: May 5, 2020
Published: June 26, 2020

REFERENCES

Albergante, L., Mirkes, E., Bac, J., Chen, H., Martin, A., Faure, L., Barillot, E., Pinello, L., Gorban, A., and Zinovyev, A. (2020). Robust and scalable learning of complex intrinsic dataset geometry via ElPiGraph. Entropy *22*, 296.

Bähler, J. (2005). Cell-cycle control of gene expression in budding and fission yeast. Annu. Rev. Genet. *39*, 69–94.

Balleza, E., Alvarez-Buylla, E.R., Chaos, A., Kauffman, S., Shmulevich, I., and Aldana, M. (2008). Critical dynamics in genetic regulatory networks: examples from four kingdoms. PLoS One *3*, e2456.

Barik, D., Ball, D.A., Peccoud, J., and Tyson, J.J. (2016). A stochastic model of the yeast cell cycle reveals roles for feedback regulation in limiting cellular variability. PLoS Comput. Biol. *12*, e1005230.

Barik, D., Baumann, W.T., Paul, M.R., Novak, B., and Tyson, J.J. (2010). A model of yeast cell-cycle regulation based on multisite phosphorylation. Mol. Syst. Biol. *6*, 405.

Battogtokh, D., and Tyson, J.J. (2016). A bistable switch mechanism for stem cell domain nucleation in the shoot apical meristem. Front. Plant Sci. *7*, 674.

Bocci, F., Tripathi, S.C., Vilchez Mercedes, S.A., George, J.T., Casabar, J.P., Wong, P.K., Hanash, S.M., Levine, H., Onuchic, J.N., and Jolly, M.K. (2019). NRF2 activates a partial epithelial-mesenchymal transition and is maximally present in a hybrid epithelial/mesenchymal phenotype. Integr. Biol. *11*, 251–263.

Branzei, D., and Foiani, M. (2008). Regulation of DNA repair throughout the cell cycle. Nat. Rev. Mol. Cell Biol. *9*, 297–308.

Chen, K.C., Calzone, L., Csikasz-Nagy, A., Cross, F.R., Novak, B., and Tyson, J.J. (2004). Integrative analysis of cell cycle control in budding yeast. Mol. Biol. Cell *15*, 3841–3862.

Chen, K.C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., Tyson, J.J., and Solomon, M.J. (2000). Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol. Biol. Cell *11*, 369–391.

Ciliberti, S., Martin, O.C., and Wagner, A. (2007). Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Comput. Biol. *3*, e15.

Collins, K., Jacks, T., and Pavletich, N.P. (1997). The cell cycle and cancer. Proc. Natl. Acad. Sci. U S A *94*, 2776–2778.

Csikász-Nagy, A., Battogtokh, D., Chen, K.C., Novák, B., and Tyson, J.J. (2006). Analysis of a generic model of eukaryotic cell-cycle regulation. Biophys. J. *90*, 4361–4379.

Dalton, S., and Coverdell, P.D. (2015). Linking the cell cycle to cell fate decisions. Trends Cell Biol. *25*, 592–600.

de Bruin, R.A.M., McDonald, W.H., Kalashnikova, T.I., Yates, J., III, and Wittenberg, C. (2004). Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. Cell *117*, 887–898.

Dirick, L., Böhm, T., and Nasmyth, K. (1995). Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of Saccharomyces cerevisiae. EMBO J. *14*, 4803–4813.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360*, eaar3131.

Ferrezuelo, F., Colomina, N., Futcher, B., and Aldea, M. (2010). The transcriptional network activated by Cln3 cyclin at the G1-to-S transition of the yeast cell cycle. Genome Biol. *11*, R67.

Gerard, C., and Goldbeter, A. (2009). Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. Proc. Natl. Acad. Sci. U S A *106*, 21643–21648.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Goldbeter, A., Gonze, D., Houart, G., Leloup, J.-C., Halloy, J., and Dupont, G. (2001). From simple to complex oscillatory behavior in metabolic and genetic control networks. Chaos Interdiscip. J. Nonlin. Sci. *11*, 247–260.

Hartwell, L.H., Culotti, J., Pringle, J., and Reid, B. (1974). Genetic control of the cell division cycle in yeast. Science *183*, 46–51.

Hong, C., Lee, M., Kim, D., Kim, D., Cho, K.-H., and Shin, I. (2012). A checkpoints capturing timing-robust Boolean model of the budding yeast cell cycle regulatory network. BMC Syst. Biol. *6*, 129.

Huang, B., Jia, D., Feng, J., Levine, H., Onuchic, J.N., and Lu, M. (2018a). RACIPE: a computational tool for modeling gene regulatory circuits using randomization. BMC Syst. Biol. *12*, 74.

Huang, B., Lu, M., Jia, D., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2017). Interrogating the topological robustness of gene regulatory circuits by randomization. PLoS Comput. Biol. *13*, e1005456.

Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J.I., Raj, A., Li, M., and Zhang, N.R. (2018b). SAVER: gene expression recovery for single-cell RNA sequencing. Nat. Methods *15*, 539–542.

Hustedt, N., and Durocher, D. (2017). The control of DNA repair by the cell cycle. Nat. Cell Biol. *19*, 1–9.

Icard, P., Fournel, L., Wu, Z., Alifano, M., and Lincet, H. (2019). Interconnection between metabolism and cell cycle in cancer. Trends Biochem. Sci. *44*, 490–501.

Jackson, C.A., Castro, D.M., Saldi, G.-A., Bonneau, R., and Gresham, D. (2020). Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. Elife *9*, e51254.

Jakoby, M., and Schnittger, A. (2004). Cell cycle and differentiation. Curr. Opin. Plant Biol. *7*, 661–669.

Jia, D., George, J.T., Tripathi, S.C., Kundnani, D.L., Lu, M., Hanash, S.M., Onuchic, J.N., Jolly, M.K., and Levine, H. (2019a). Testing the gene expression classification of the EMT spectrum. Phys. Biol. *16*, 025002.

Jia, D., Lu, M., Jung, K.H., Park, J.H., Yu, L., Onuchic, J.N., Kaipparettu, B.A., and Levine, H. (2019b). Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic

pathways. Proc. Natl. Acad. Sci. U S A *116*, 3909–3918.

Johnston, G.C., Pringle, J.R., and Hartwell, L.H. (1977). Coordination of growth with cell division in the yeast Saccharomyces cerevisiae. Exp. Cell Res. *105*, 79–98.

Jones, D.L., Brewster, R.C., and Phillips, R. (2014). Promoter architecture dictates cell-to-cell variability in gene expression. Science *346*, 1533–1536.

Kalucka, J., Missiaen, R., Georgiadou, M., Schoors, S., Lange, C., Bock, K.D., Dewerchin, M., and Carmeliet, P. (2015). Metabolic control of the cell cycle. Cell Cycle *14*, 3379–3388.

Kaplon, J., van Dam, L., and Peeper, D. (2015). Two-way communication between the metabolic and cell cycle machineries: the molecular basis. Cell Cycle *14*, 2022–2032.

Kohar, V., and Lu, M. (2018). Role of noise and parametric variation in the dynamics of gene regulatory circuits. NPJ Syst. Biol. Appl. *4*, 40.

Kraikivski, P., Chen, K.C., Laomettachit, T., Murali, T.M., and Tyson, J.J. (2015). From START to FINISH: computational analysis of cell cycle control in budding yeast. NPJ Syst. Biol. Appl. *1*, 15016.

Li, C., and Wang, J. (2014). Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. Proc. Natl. Acad. Sci. U S A *111*, 14130–14135.

Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. U S A *101*, 4781–4786.

Li, V.C., and Kirschner, M.W. (2014). Molecular ties between the cell cycle and differentiation in embryonic stem cells. Proc. Natl. Acad. Sci. U S A *111*, 9503–9508.

Li, W.V., and Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat. Commun. *9*, 1–9.

Lim, H.H., Goh, P.-Y., and Surana, U. (1998). Cdc20 is essential for the cyclosome-mediated proteolysis of both Pds1 and Clb2 during M phase in budding yeast. Curr. Biol. *8*, 231–237.

Liu, X., Wang, X., Yang, X., Liu, S., Jiang, L., Qu, Y., Hu, L., Ouyang, Q., and Tang, C. (2015). Reliable cell cycle commitment in budding yeast is ensured by signal integration. Elife *4*, e03977.

Liu, Z., Lou, H., Xie, K., Wang, H., Chen, N., Aparicio, O.M., Zhang, M.Q., Jiang, R., and Chen, T. (2017). Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat. Commun. *8*, 22.

Llamosi, A., Gonzalez-Vargas, A.M., Versari, C., Cinquemani, E., Ferrari-Trecate, G., Hersen, P., and Batt, G. (2016). What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. PLoS Comput. Biol. *12*, e1004706.

Lv, C., Li, X., Li, F., and Li, T. (2015). Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. PLoS Comput. Biol. *11*, e1004156.

Macnair, W., and Claassen, M. (2019). psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels. bioRxiv. https://doi.org/10.1101/622001.

Mangla, K., Dill, D.L., and Horowitz, M.A. (2010). Timing robustness in the budding and fission yeast cell cycles. PLoS One *5*, e8906.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. *33*, 269–276.

Murray, J.M., and Carr, A.M. (2018). Integrating DNA damage repair with the cell cycle. Curr. Opin. Cell Biol. *52*, 120–125.

Myster, D.L., and Duronio, R.J. (2000). Cell cycle: to differentiate or not to differentiate? Curr. Biol. *10*, R302–R304.

Novák, B., and Tyson, J.J. (2008). Design principles of biochemical oscillators. Nat. Rev. Mol. Cell Biol. *9*, 981–991.

Okabe, Y., and Sasai, M. (2007). Stable stochastic dynamics in yeast cell cycle. Biophys. J. *93*, 3451–3459.

Palumbo, P., Vanoni, M., Cusimano, V., Busti, S., Marano, F., Manes, C., and Alberghina, L. (2016). Whi5 phosphorylation embedded in the G1/S network dynamically controls critical cell size and cell fate. Nat. Commun. *7*, 11372.

Prinz Susanne, A.A. (1999). Dual control of mitotic exit. Nature *402*, 133.

Qu, Y., Jiang, J., Liu, X., Wei, P., Yang, X., and Tang, C. (2019). Cell cycle inhibitor Whi5 records environmental information to coordinate growth and division in yeast. Cell Rep. *29*, 987–994.e5.

Roy, D., Sheng, G.Y., Herve, S., Carvalho, E., Mahanty, A., Yuan, S., and Sun, L. (2017). Interplay between cancer cell cycle and metabolism: challenges, targets and therapeutic opportunities. Biomed. Pharmacother. *89*, 288–296.

Rudner, A.D., and Murray, A.W. (2000). Phosphorylation by Cdc28 activates the Cdc20-dependent activity of the anaphase-promoting complex. J. Cell Biol. *149*, 1377–1390.

Ruijtenberg, S., and van den Heuvel, S. (2016). Coordinating cell proliferation and differentiation: antagonism between cell cycle regulators and cell type-specific gene expression. Cell Cycle *15*, 196–212.

Rustici, G., Mata, J., Kivinen, K., Lió, P., Penkett, C.J., Burns, G., Hayles, J., Brazma, A., Nurse, P., and Bähler, J. (2004). Periodic gene expression program of the fission yeast cell cycle. Nat. Genet. *36*, 809–817.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. *37*, 547.

Sanchez-Osorio, I., Ramos, F., Mayorga, P., and Dantan, E. (2013). Foundations for modeling the dynamics of gene regulatory networks: a multilevel-perspective review. J. Bioinform. Comput. Biol. *12*, 1330003.

Sandal, T. (2002). Molecular aspects of the mammalian cell cycle and cancer. Oncologist *7*, 73–81.

Schwikowski, B., Uetz, P., and Fields, S. (2000). A network of protein-protein interactions in yeast. Nat. Biotechnol. *18*, 1257–1261.

Schwob, E., and Nasmyth, K. (1993). CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in Saccharomyces cerevisiae. Genes Dev. *7*, 1160–1175.

Shaltiel, I.A., Krenning, L., Bruinsma, W., and Medema, R.H. (2015). The same, only different – DNA damage checkpoints and their reversal throughout the cell cycle. J. Cell Sci. *128*, 607–620.

Shirayama, M., Tóth, A., Gálová, M., and Nasmyth, K. (1999). APCCdc20 promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. Nature *402*, 203.

Skotheim, J.M., Di Talia, S., Siggia, E.D., and Cross, F.R. (2008). Positive feedback of G1 cyclins ensures coherent cell cycle entry. Nature *454*, 291–296.

Soufi, A., and Dalton, S. (2016). Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. Development *143*, 4301–4311.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell *9*, 3273–3297.

Stuart, D., and Wittenberg, C. (1995). CLN3, not positive feedback, determines the timing of CLN2 transcription in cycling cells. Genes Dev. *9*, 2780–2794.

Tracy, S., Yuan, G.-C., and Dries, R. (2019). RESCUE: imputing dropout events in single-cell RNA-sequencing data. BMC Bioinformatics *20*, 388.

Tyers, M., Tokiwa, G., and Futcher, B. (1993). Comparison of the Saccharomyces cerevisiae G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. EMBO J. *12*, 1955–1968.

Tyson, J.J., Chen, K.C., and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr. Opin. Cell Biol. *15*, 221–231.

Tyson, J.J., and Novak, B. (2014). Control of cell growth, division and death: information processing in living cells. Interface Focus *4*, 20130070.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018).

Recovering gene interactions from single-cell data using data diffusion. Cell *174*, 716–729.e27.

Visintin, R., Craig, K., Hwang, E.S., Prinz, S., Tyers, M., and Amon, A. (1998). The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. Mol. Cell *2*, 709–718.

Wagner, F., Yan, Y., and Yanai, I. (2018). K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. bioRxiv. https://doi.org/10.1101/217737.

Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Res. *47*, e66.

Wijnen, H., Landman, A., and Futcher, B. (2002). The G1 cyclin Cln3 promotes cell cycle entry via the transcription factor Swi6. Mol. Cell. Biol. *22*, 4402–4418.

iScience, Volume *23*

Supplemental Information

Random Parametric Perturbations of Gene Regulatory Circuit Uncover State Transitions in Cell Cycle

Ataur Katebi, Vivek Kohar, and Mingyang Lu

Supplemental Information

Table of contents

Transparent Methods

Supplemental Figures

Supplemental Tables

Supplemental References

Transparent Methods

1. Generalized RACIPE

We have recently developed a computational algorithm, named *ra*ndom *ci*rcuit *per*turbation (RACIPE), for modeling the dynamics of a gene regulatory circuit without the need for specific kinetic parameters(Huang et al., 2017). Compared to the conventional modeling methods, RACIPE uses the topology of a circuit as the only input for modeling. Instead of fitting kinetic parameters against experimental data (such as gene expression profiles), RACIPE generates an ensemble of mathematical models with distinct random kinetic parameters. The modeling results are then derived from the statistics of the dynamical behavior of the whole ensemble. Details of the RACIPE framework are described in one of our previous papers(B. Huang et al., 2018).

In the original RACIPE, we mainly focused on modeling gene circuits with transcription factors involving transcriptional regulation and analyzing the stable steady state solutions from a random model. However, the yeast cell cycle circuit contains not only transcriptional regulation but also signaling regulation by kinases and the regulation of degradation by specialized protein complexes, such as anaphase-promoting complex widely known as APC. The cell cycle is also frequently modeled by oscillatory dynamics. To adapt the RACIPE-based approach, we have generalized the original method to allow modeling circuits with the additional types of regulatory interactions and developed a numerical method to systematically detect oscillatory states.

2. Modeling multiple interaction types

In the generalized RACIPE, we model the level of gene X in a given GRC using the following rate equations for different situations. In the original RACIPE (B. Huang et al., 2018), transcriptional regulation of a gene Y to a gene X is modeled by

$$\frac{dX}{dt} = gH^S(Y) - kX \quad , \qquad \text{(S1)}$$

where g is the basal production rate of X, k is the degradation rate of X, and H^S is the shifted Hill function. Huang et al. defined the term basal production rate (g) as the production rate of a gene without any regulator bound to the promoter, the maximum production rate (G) as the production rate when the activators are abundant and inhibitors are scarce (B. Huang et al., 2018). For a gene X regulated by n1 activators (no matter how many inhibitors are presented), the basal production rate can be expressed in terms of maximum production rate, as shown below:

$$g = \frac{G}{\prod_{i=1}^{n1} \lambda_{Y_i}} \quad , \qquad \text{(S2)}$$

where, λ_{Y_i} is the fold change of activator Y_i. Noticeably, g is independent to the fold change for inhibitors -- when X is controlled by only inhibitors, g = G. In Equation S1, the shifted Hill function, as originally reported in (Lu et al., 2014), can be expressed as

$$H^S(Y, Y_0, n_Y, \lambda_Y) = \lambda_Y + \frac{1-\lambda_Y}{1+\left(\frac{Y}{Y_0}\right)^{n_Y}} \quad , \qquad \text{(S3)}$$

where Y_0 is the threshold level of Y, n_Y is the Hill coefficient, and λ_Y is the fold change. H^S is one when Y equals to zero, and converges to λ_Y when Y is infinity. Thus, H^S can be applied to

either excitatory interactions ($\lambda_Y > 1$) or inhibitory interactions ($0 < \lambda_Y < 1$). This definition of the shifted Hill function gives us additional convenience to handle some modeling complexity. First, it's much more convenient to randomize λ than multiple production rate constants. Second, for genes independently regulated by multiple regulators, we can naturally model the production rate in the form of product of multiple shifted Hill terms (shown in Equation S5 and Equation S7).

For a gene X that is regulated by a gene S via signaling interaction, e.g., phosphorylation event, we model the *activity* of gene X (the level of the active form, instead of the level of gene expression) by

$$\frac{dX}{dt} = \eta(gH^S(S) - kX) . \qquad \text{(S4)}$$

Here, the rate equation is based on a phenomenological model that (1) uses the shifted Hill function to capture the ultra-sensitivity of signaling interaction (Huang and Ferrell, 1996); (2) uses a constant scaling factor η to account for faster time scale of signaling regulation (a similar approach has been used in a recent study (Kraeutler et al., 2010)); (3) describes the degradation/disassociation term with linear approximation. For modeling the circuits in this study, we set η as ten (Steinway et al., 2014). For a gene X that is regulated by n genes (S_i) via signaling interactions only

$$\frac{dX}{dt} = \eta(g\prod_{i=1}^{n} H^s(S_i) - kX) . \qquad \text{(S5)}$$

For a gene X whose protein degradation is regulated by gene Z, we model the level of gene X by

$$\frac{dX}{dt} = g - kXH^S(Z) . \qquad \text{(S6)}$$

When a combinatorial interaction of multiple types is involved, e.g., transcriptional regulation by n1 regulators (Y_i), signaling regulation by n2 regulators (S_j) and degradation regulation by n3 regulators (Z_l), and in the situation with at least one transcriptional or degradation regulation (n1+n3 > 0), we model the activity of gene X by

$$\frac{dX}{dt} = g\prod_{i=1}^{n1} H^s(Y_i)\prod_{j=1}^{n2} H^s(S_j) - kX\prod_{l=1}^{n3} H^s(Z_l) . \qquad \text{(S7)}$$

We do not include the scaling factor η here, as transcription and/or degradation regulations are usually the rate-limiting steps.

With the above generalized formalism, we model the activity level of each node in the cell cycle GRC (shown in Figure 1A) accordingly. The detailed rate equations for each node in the cell cycle GRC are listed in Section 13.

3. Detecting both stable states and oscillatory states

The original RACIPE detects stable steady states of each model by first simulating ordinary differential equations (ODEs) from each of 1000 initial conditions until the system converges to a stable steady state, and then identifies all unique steady states. It is challenging to detect oscillatory dynamics robustly and efficiently with a high-throughput numerical method, as an oscillatory system never converges to a constant expression. Additionally, we need to detect the coexistence of both the stable steady states and the oscillatory states. Here, we design and

implement a fast-numerical method that automatically detects both the oscillatory states and the stable states. The steps of the numerical method are illustrated in the flowchart in Figure S8A. First, we simulate the corresponding ODEs for 10000 models, each of which starts from 1000 randomly sampled initial conditions (ICs) (Figure S8A, step 1). For each model, we simulate upto 2000 time units using a step size of 0.01. From these simulations, we identify all unique steady state gene activity profiles (the same procedure as the original RACIPE) and the rest nonsteady state gene activity profiles (the derivatives are non-zero) (Figure S8A, step 2 and step 3). In Figure S8A, step 3 and step 4 schematically show how we further identify the oscillatory states from the nonsteady states. Details of these steps (step 3 and step 4) are given below.

In order to identify the oscillatory state from the nonsteady states, we perform additional simulations of the corresponding ODEs (starting from each nonsteady state) and examine whether the nonsteady state leads to a stable oscillatory state. Figures S8B, S8D show two examples of time trajectories projected onto the first two principal components obtained from the RACIPE models. We calculate the distance between the starting nonsteady state (labeled as SS in the figure) and each state along the simulation and evaluate the repeating patterns of the distances along the simulation time. The distances start from zero and are expected to go up and down until eventually reaching to a steady oscillatory pattern (Figures S8C, S8E). We chose to evaluate a total of 100 local minima and maxima of the distance trajectory to identify the converged pattern.

Next, we compute the period of the oscillation as follows. Starting from the last minimum distance of the 100 local extrema that we obtained in the previous step, we evaluate the differences between the last minimum distance and any of the distance value of a preceding local minimum until we find the same distance value. The period of the oscillation is estimated by the time differences of the two local minima. Figure S8C shows a case where the period is the time difference between the two occurences of the same local minimum and Figure S8E shows a case where the period is the time difference between the two consecutive occurences of one of the two local minima. We calculate three more periods by searching backward further among the local minima. If the difference between each pair of the periods is less than three time steps (an adjustable parameter), we identify that the nonsteady state leads to a stable oscillatoy state and record its period. Otherwise, we discard this nonsteady state. The oscillatory trajectory of a complete cycle can be obtained by simulating the ODEs using the state of a local minimum as the initial condition and the period of the oscilation as the duration of the simulation.

For the remaining nonsteady states, we repeat the procedure described above and check for addional oscillatory states. Through the above procedure, if a nonsteady state gives rise to an oscillatory state which is different from all the previously found periods, we consider this as a new oscillatory state. There is a very rare chance that the method will discard new coexisting oscillatory states with exactly the same period. But it is less likely to affect the statistical results from a large number of models.

4. A polarity metric of limit cycles

For an oscillatory state, we project the corresponding oscillatory trajectory to the first two principal components (PCs) that are derived from all of the stable steady states from the 10,000

RACIPE models. In this two-dimensional space, we quantify the directionality of a projected oscillatory trajectory (limit cycle) by a polarity metric, defined below:

$$I = \frac{\sum_i (\mathbf{x}_i - \mathbf{x}_0) \times (\mathbf{x}_{i+1} - \mathbf{x}_i)}{(PC1_{\max} - PC1_{\min})^2 + (PC2_{\max} - PC2_{\min})^2}, \quad \text{(S8)}$$

where the denominator is the sum of the squares of each extent of the limit cycle along PC1 and PC2 ($PC1_{min}$ and $PC1_{max}$ are the minimum and the maximum of the limit cycle along PC1, and $PC2_{min}$ and $PC2_{max}$ are the minimum and the maximum along PC2) and the numerator is the sum over each magnitude of the cross product between the vector pair x_i-x_0 and x_{i+1}-x_i where x_0 is the position vector for the center of the limit cycle, x_i is the position vector for a point i, where i runs through each point along the limit cycle. The center of the limit cycle was obtained by projecting the mean expression of all the expression vectors along the oscillatory trajectory on the first two PCs. A positive polarity indicates a counterclockwise oscillation of the projected trajectory. Likewise, a negative polarity corresponds to a clockwise oscillation.

5. Identifying the state transitions in an oscillation

To further characterize the directionality of an oscillatory trajectory without the dimension reduction by PCA, we use the activity profile along the oscillatory trajectory and adopt a machine learning approach on these data to identify the cell cycle phases traveled by the trajectory. From this sequence of phases, we can obtain the pattern of state transitions for each oscillation. First, we apply *nnet*(Venables and Ripley, 2002) to train a feed-forward neural network using the activity profile of 10000 RACIPE models (a total of 12758 stable steady states), where the data are classified by their cell cycle phases obtained from the hierarchical clustering analysis (Figure 1B). *nnet* fits a neural network with single-hidden layer containing five hidden nodes. Second, we use the trained model to assign cell cycle phases to the points along the oscillatory trajectory. Here, a data point is assigned to a cell cycle phase with the highest probability from the neural network. To achieve high accuracy in the prediction, we only make the assignment to the points when the differences between the first two highest probabilities are larger than a threshold value T. The remaining points are not included in the calculation of the cell cycle phase transition pattern of the oscillatory trajectory.

We tested the performance of the neural network model by employing a 10-fold cross-validation on the data set of 12758 states. The performance was evaluated for data points that are retained after the filtering step. The first two panels of Figure S26 show the accuracy of the model on the training and testing sets for various T threshold values. It is clear that the accuracy consistently gets better as T increases. However, the coverage (defined as the percentage of retained points) decreases as T increases. We chose T as 0.25, as it provides a right balance of accuracy and coverage, both being around 91%.

After assigning the cell cycle phases to the points along a limit cycle trajectory, we further obtain the cell cycle phase transition pattern by removing the redundant phases from every consecutive repeat (which represents more cycles). With this strategy, we compute the unique state transition patterns of all the oscillatory trajectories (47637) from the 160000 RACIPE models. We find 47 unique patterns from the trajectories passing through 3 or more cell cycle phases. A list of these patterns and their frequencies can be found in Table S7.

6. Determining directionality from circuit-topology-based delayed correlations

We further developed a correlation-based method to predict the propensity of transition between two cell cycle phases using the RACIPE-generated simulated gene activity data and the circuit topology. In this method, we calculate the delayed correlation of each pair of cell cycle phases as follows. We compute the Pearson's correlations between the gene activities of two models where the first model coming from the first cell cycle phase 1 and the other model from the other phase 2. The forward correlation between phase 1 and phase 2 is cor(**T**(2)·**s**(links), **R**(1)), where **T**(2) is the activity vector of the targets belonging to the phase 2 model, **R**(1) is the activity vector of the sources/regulators belonging to the phase 1 model, and **s**(links) is the sign vector expressing each interaction type (+1: excitatory, -1: inhibitory). Thus, the backward correlations for the same model pair can be calculated as cor(**T**(1)·**s**(links), **R**(2)). To infer the direction of state transitions, we evaluate the median delayed correlations for any pair of models from the two phases – significantly larger (positive) forward correlation suggests an irreversible transition from phase 1 to phase 2. While calculating the delayed correlations for each successive pair of cell cycle phases, we considered the monostable models belonging to each cell cycle phase.

7. Convergence of RACIPE sampling

We evaluated the convergence of RACIPE simulations by inspecting the convergence of the stable steady states, local density of the stable steady-state gene activity profiles, and the 2D density maps. We considered 160000 RACIPE models, from which we drew eight samples (without replacement) at each of the three sample sizes 5000, 10000, and 15000. First, for each dataset (i.e., each sample at each sample size level), we calculated the histogram of the number of stable steady states for models (1) with no oscillatory state (Figure S2A), (2) with additional one oscillatory state (Figure S2B), and (3) with additional two oscillatory states (Figure S2C). We observed that the statistics for the number of stable states (both stable steady states and oscillatory states) converge across the samples of different sizes.

Next, we evaluated the behaviors of the stable steady states from different samples drawn from 134972 unique stable steady states obtained from 160000 RACIPE models. We calculated P(x), the k nearest neighbor density estimate (kNN DE)(Duda et al., 2000) around a reference point x representing the gene activity profile in high dimension:

$$P(x) \cong \frac{k}{NV} = \frac{k}{N.c_D.R_k^D(x)},$$

where k is the number of nearest neighbors for the density estimation, N is the total number of data points, V is the volume that encloses the k neighbors surrounding the point x. c_D is the volume of the unit sphere in D dimensions, defined as: $c_D = \frac{\pi^{D/2}}{\Gamma(\frac{D}{2}+1)}$, where Γ represents the gamma function, and $R_k^D(x)$ is the distance between the point x and its k-th nearest neighboring point. D was selected as 15 which is the number of genes in the cell cycle GRC. To calculate the kNN density estimates, we used knnDE function from the R package TDA(Fasy et al., 2015) which implements the above formulation.

To compare the convergence of the local density of the stable steady state gene activity profiles at different sample sizes, we first calculated the kNN density estimate of each gene activity profile for each of the eight datasets at each of the three sample size levels (5K, 10K, and 15K). For each of the three sample sizes 5K, 10K, and 15K, we use 100, 200, and 300 as the number of neighbors in kNN density estimation. Then, at each sample size level, we calculated the histograms (with 18 bins) of the log transformed (with base 10) kNN densities obtained from each sample. Subsequently, we calculated the mean and the standard deviation of each bin across the eight samples at a sample size level, as shown in Figure S3A. The low standard deviation for each bin across the samples of different sizes indicates the convergence of the local density of the stable steady state gene activity profiles from the RACIPE models.

To inspect the convergence of the density maps, we first projected each sample dataset onto the first two principal components obtained from the 10K models used in Figure 1. Then, we calculated the kNN density at each grid point (on an 81x78 grid). For each of the three sample sizes 5K, 10K, and 15K, we use 100, 150, and 200 as the number of neighbors in kNN density estimation. The 2D density map is found to be consistent along all the samples across each of the three sample sizes (Figure S3B). To compare the convergence of these kNN densities, we used Bhattacharyya (BC) distance (defined in Section 11) as a metric. First, we calculated the BC distance between any pair of the 2D density maps within group (distributions obtained from samples of specific size) and between groups (distributions obtained across different sample sizes). We found that overall, increasing the sample size reduces the mean and standard deviation of the BC distances. Additionally, the mean of the BC distances between 5K and 10K samples is similar to the mean of the BC distances between 5K and 15K samples, whereas the mean of the BC distances between 10K and 15K samples is noticeably smaller than the other two, suggesting that the kNN density is converged for models around 10K to 15K.

In summary, the convergence in the multiple analyses indicate that an ensemble of 10K RACIPE models are sufficient to generate good statistics. When analyzing oscillatory states, we considered all 160000 models to improve statistics.

8. Processing and analyzing yeast single cell RNA-seq data

To compare RACIPE simulations of the cell cycle gene regulatory circuit with experimental data, we analyzed a published single-cell RNA-seq (scRNA-seq) dataset of budding yeast (*Saccharomyces cerevisiae*) of diverse genotypes in 11 different environmental conditions (Jackson et al., 2020). Out of the 38,285 cells, 976 wild type cells grown in YPD condition (Yeast Extract, Peptone, Glucose – in Table 1(Jackson et al., 2020)) were considered in this analysis. We used *normalizeSCE* from the R package Scater(McCarthy et al., 2017) to normalize the data that removes cell-specific scaling biases, e.g., due to differences in sequencing coverage or capture efficiency. Out of 6,828 genes, 1,205 are not expressed in any of the 976 cells and hence were removed from further consideration. For the remaining 5,623 genes, the z-scores were calculated from the normalized log counts. To evaluate the mRNA expression patterns of the network genes, we chose and examined 15 representative genes. These 15 genes, either the network genes or their suitable replacements, were chosen because they were found to reach their cell-cycle-phase-specific peak expressions a given in Spellman et al.(Spellman et al., 1998). More details on how these genes were chosen can be found in Table

S3. Also, Whi5 is not expressed in any cells; its functionally related genes, including Rnr1 and Rnr3, were either not expressed or only expressed in a very few cells. Thus, we left Whi5 as zero expression in the following analysis.

We first performed Principal Component Analysis (PCA) on the z-scores of the 5623 genes. The projected expressions (on PC1-PC2) of the 15 genes corresponding to the circuit topology clearly marks their cell cycle phase specific expressions (Figure S4, left panel). However, because of the low coverage of the gene expressions across the cells, when the expressions of the 15 genes are projected on the PCs obtained from the RACIPE models (*i.e.*, the PCs shown in Figure 1D in the main article), the distinct cell cycle phase specific expression patterns of these genes are not as clear (Figure S4, right panel).

To improve the analysis, we used K-nearest neighbor (KNN) smoothing implemented by Wagner et al.(Wagner et al., 2018). As we already normalized the data, we turned off the median normalization and Freeman-Tukey transformations on the data in the KNN smoothing method. After some exploration, we chose the number of nearest neighbors as 15 and the number of principal components as 2 (the cell cycle is mostly captured by the first two components) for KNN smoothing. The smoothed data were again analyzed by PCA. When the z-scores of the smoothed data were projected on the PCs of all of the 5,623 genes, we observed that the genes corresponding to the circuit topology again clearly separate the cell cycle phases (Figure S5, left - note that some cells are overlapping in this case because of smoothing). This distinction is noticeable even when the scRNA-seq data of the 15 representative genes were projected on the PCs from the RACIPE models (Figure S5, right). The projected expressions of four marker genes (Cln3, Rfa1, HTA1, and Clb1) on the two different projections are shown in Figures 1E, 1F.

Furthermore, we employed four additional methods (they are all based on imputation) namely MAGIC (van Dijk et al., 2018), scImpute (Li and Li, 2018), SAVER (M. Huang et al., 2018), and RESCUE (Tracy et al., 2019). All these methods could successfully recover the circular structure of the genome-wide transcriptomics data, with KNN, SAVER, and RESCUE performing better than the other two. However, when the imputed scRNA-seq data of the 15 representative genes were projected on the PCs from the RACIPE models, none of these methods improved the outcomes over the KNN smoothing method. The results are shown in Figure S6 and Figure S7.

9. Reconstructing cell cycle dynamics using steady state gene activity profiles from RACIPE models

Based on hierarchical clustering analysis of the activity profiles, we devised a simple algorithm to reconstruct the cell cycle dynamics which is described below.

9.1. Pseudotime estimation based on hierarchical clustering analysis

In Figure 1B, we show how we mapped the clusters to distinct cell cycle phases. Here, we show how we use the clustered data to assign pseudotime for every gene activity profile and reconstruct the cell cycle dynamics. First, for the clusters of early G1 and mid G1, we compute the pairwise gene activity distances between each of the two models from the two ends of early

G1 and each of the two models at the two ends of mid G1. We bring together the two ends that give the smallest distance (by flipping the model orders if necessary) and obtain the correct order for early G1 and mid G1. Second, for every additional cluster that is associated with another cell cycle phase (in the sequence of late G1, S, M, and M/G1), we compute the pairwise gene activity distances between each of the two models from the two ends of the new cluster and the last model (after reordering) from the previous phase. We bring the side of the new cluster that produces smaller distance at the end of the ordered clusters in the same manner as performed in the previous step. This step is repeated until we order all the six phases (early G1, mid G1, late G1, S, M, and M/G1). Pseudotime is then assigned according to the ordered sequence. The gene activity profiles of the 15 genes along the pseudotime are shown in Figure S10A.

Additionally, we also inferred the cell cycle pseudotime by a custom-use of another method named psupertime (Macnair and Claassen, 2019) as described below.

9.2. Pseudotime estimation by custom-use of psupertime method

We used the psupertime() function from the R package psupertime (Macnair and Claassen, 2019). We made a custom-use of the method when the direct use of the method did not provide desired results. First, we directly used the method by providing the labeled RACIPE activity profiles to the method. The program failed to position the cell cycle phases in the correct order.

To overcome the above limitation of the program in calculating pseudotime for systems with a circular structure such as cell cycle, we calculated pseudotime in two stages. First, we calculated the pseudotime (t1) for the models belonging to the first four of the six phases: early G1, mid G1, late G1, and S. Second, we calculated the pseudotime (t2) for the models belonging to the last three phases from the first cycle and the first phase from the second cycle: S, M, M/G1, and early G1. Note that S and early G1 are overlapped between the two calculations. Then, we perform a scaling of t2 to t2′ by t2′ = a* t2 + b so that the mean and the standard deviation of the two distributions for the S phase from t1 and t2′ match. Additionally, to align the pseudotime of early G1 phase from the second cycle, we set the pseudotime for the second cycle as t1′ = t1 + c. To obtain the parameter a, we minimize the following cost function:

$$F = (\sigma_{1.early\ G1} - a\sigma_{2.early\ G1})^2 + (\sigma_{1.S} - a\sigma_{2.S})^2 \qquad (S9)$$

where $\sigma_{1.early\ G1}$ and $\sigma_{2.early\ G1}$ are the standard deviations of early G1 phase pseudotime in t1 and t2, respectively and $\sigma_{1.S}$ and $\sigma_{s.S}$ are the standard deviations of S phase pseudotime in t1 and t2, respectively. From this step, we found a to be 1.38. We obtained the parameters b and c from the following two equations:

$$b = \mu_{3.2} - \mu_{3.1} \qquad (S10)$$

where $\mu_{3.1}$ and $\mu_{3.2}$ are the means of the S phase pseudotime in t1 and t2, respectively.

$$c = \mu_{6.2} - \mu_{6.1} \qquad (S11)$$

where $\mu_{6.1}$ and $\mu_{6.2}$ are the means of the early G1 phase pseudotime in t1 and t2, respectively. From this step, we identified b and c to be 11.23 and 27.04, respectively.

To calculate the pseudotime for the S phase, we took the average of its pseudotime from t1 and that from t2′. And, for the early G1 phase, we took the average of its pseudotime from t2′

and t1′, followed by subtracting c. The pseudotime for mid G1 and late G1 comes from t1 and that for M and M/G1 comes from t2′. The above steps are illustrated in Figure S11. However, using the pseudotime obtained in the steps above, we still didn't get satisfactory results. The main issue is the ranges of the pseudotime for S and early G1 are largely decreased, thus the gene activity time dynamics have drastic changes within these two phases. We addressed this problem by assigning the order indices, instead of the inferred pseudotime values, as time. The results are shown in Figure S12A. We further calculated autocorrelation for each node in the cell cycle GRC and cross-correlation for each node pair of the four marked nodes (shown in Figure S12B and S12C**,** respectively). Compared to our former analysis (Figure S10), the time dynamics obtained from the custom-use of psupertime are smoother, suggesting it captures cell cycle dynamics better.

9.3. Other pseudotime inference methods

We also explored several other trajectory reconstruction methods. Though more than 70 pseudotime inference methods have already been published, their performances depend mostly on the dataset dimensions and the characteristics of the topology of the pseudotime. A recently published evaluation benchmark brought most of these methods in one software platform named dynverse (Saelens et al., 2019). We used the dynverse implementation of trajectory inference methods to calculate pseudotime for 10000 simulated models. For our data which contain a circular structure as the expected topology, dynverse suggested Angle (Saelens et al., 2019), ElPiGraph Cycle (Albergante et al., 2020), and reCAT (Liu et al., 2017) as the top three methods. Additionally, we provided the cell cycle state assignments of the models, early G1 as the starting state and M/G1 as the end state, and a randomly selected model from early G1 as the root node as the user input to these methods. The results of the three pseudotime and the identified trajectory are shown in Figure S13. We observe that, although the trajectory inference methods captured the circular structure, they usually failed to capture the right order of the cell cycle phases. Additionally, there are major differences between the pseudotime values inferred by the three methods.

We also tried another recent pseudotime method SoptSC (Wang et al., 2019) by using the default settings in its R implementation. In this case, we used 5000 models, as calculations using 10000 models didn't finish within three days using a single computer core. It successfully identified many correct transitions like mid G1 to late G1, late G1 to S, M to M/G1, M/G1 to earlyG1 but yielded a bifurcating trajectory instead of cyclic one. It didn't identify early G1 to mid G1, and S to M transitions and instead gave a bifurcation from M/G1 to early G1 and mid G1. The results from this method are shown in Figure S14.

10. RACIPE simulation of a repressilator circuit

Repressilator circuit consists of three genes in a loop where each gene represses the next gene in the loop (Figure S22A). Using this circuit topology as the input, we generated 10,000 RACIPE models with random parameters. Hierarchical clustering of these models gives six distinct clusters named CL1 ~ CL6 and the percent of models in these clusters are 21, 14, 21, 14, 13, and 17, respectively (Figure S22B). Projection of these clusters on the first two principal components (PC1 and PC2) places them in the following order CL1, CL2, CL6, CL4, CL3, and

CL5, respectively. Subsequently, we calculated the pairwise delayed correlations (detailed in Section 6) of the models belonging to each cluster pair. We found six dominant delayed correlations occurring between the cluster pairs CL1:CL2, CL2:CL6, CL6:CL4, CL4:CL3, CL3:CL5, and CL5:CL1, highlighted as red and bold in Figure S22D. These cluster pairs form a circular structure (Figure S22E), where the order of the clusters matches with their order on the PCA projection (Figure S22C). The directions of the state transitions are consistent with the dynamical changes in gene activity from an oscillatory state of the repressilator circuit (Figure S22F). The parameters for the simulation are production rate, MPR(G)=(18.213301, 94.944480, 75.030818), degradation rate, DNR(k)=(0.778456, 0.251355, 0.219330), threshold, TSH(A)=(48.253526, 45.761646, 47.119051), Hill coefficient, HCO(n)=(3, 5, 5), and fold change, FCH(λ)=(0.013442, 0.225939, 0.021474).

Additionally, we analyzed the simulated models belonging to each cluster by systematically perturbing their parameters toward the global average (see Section 12). The nine perturbation levels that we used are 0.005%, 0.05%, 0.50%, 1%, 2%, 3%, 5%, 10%, and 15%, respectively. The mean and standard deviation calculated at each perturbation level are shown in Figure S23. Panels on each column show the distinct state transition from the cluster at the top to a cluster at the left side. From all the columns together, we can see all the state transitions CL1 → CL2 → CL6 → CL4 → CL3 → CL5. The distribution of their negative z-scores (defined in Section 12) are shown in Figure S24.

11. RACIPE simulation of an induced toggle switch circuit

The gene circuit consists of three genes A, B, and C, where genes A and B are self-activating and also mutually inhibitory; and C is an input node that activates gene A and in turn is inhibited by gene B (Figure S25A). We generated 10000 RACIPE models for this circuit. By hierarchical clustering (R function hclust with complete linkage and (1 - Pearson correlation) as the distance function), we obtained four distinct clusters (Figure S25B) – CL1: A and C high, B low; CL2: A and C high, B low; CL3: A low, B and C high; CL4: A and B high, C low. The clusters once projected on the first two principal components occupy four distinct regions (Figure S25C). Two clusters CL1 and CL2 are the most dominant ones with 46.93% and 49.77% of the models, respectively, and the other two clusters CL3 and CL4 with 2.11% and 1.99%, respectively. Subsequently, we calculated the pairwise delayed correlations (detailed in Section 6) of the models belonging to each cluster pair. Based on the medians of the delayed correlations between cluster pairs (Figure S25D), we constructed the state transitions between these clusters (Figure S25E). We predicted transitions to occur between cluster pair (CL1 and CL3) and cluster pair (CL2 and CL4). This prediction is consistent with RACIPE simulation of the bistable models (models with two stable steady states from both CL1 and CL2). When we simulated these bistable models with CL1 states as the initial conditions and the increased production rate for gene C, we found the trajectories to travel through CL3 before they reach CL2 (Figure S25F). On the other hand, simulating the models with CL2 states as the initial conditions and the decreased production rate for gene C, we found the trajectories to travel through CL4 before they reach CL1 (Figure S25G).

Furthermore, we discretized each of the forward and backward projected trajectories by using 10 bins. Figure S25H, bottom shows the mean and standard deviation of the binned data in

both direction (red: forward, blue: backward). The line goes through the average of the data points. Note that the distance between the average forward and backward trajectories maximizes as they pass through CL3 and CL4, respectively. This suggests that CL3 is an intermediate state on the transition path from CL1 to CL2; on the other hand, CL4 is a distinct intermediate state on the transition path from CL2 to CL1. Additionally, we calculated the Bhattacharyya distance between the pair of distributions for each bin pair (bin IDs in the forward direction aligned with the bin IDs in the backward direction). For each bin pair, first we calculated Bhattacharyya coefficient, BC:

$$BC(p,q) = \sum_{x \in X} \sqrt{p(x)q(x)} \qquad (S12)$$

where p(x) and q(x) are the forward and backward probability densities for the bin pair. Then, we calculated Bhattacharyya distance, D_B for the two distributions p and q:

$$D_B = -\ln\big(BC(p,q)\big) \qquad (S13)$$

The top panel in Figure S25 shows that this distance maximizes when the corresponding forward and backward trajectories pass through CL3 and CL4 respectively.

12. Modeling the effects of parameter perturbations

In the main article, we proposed putative models to explain how irreversible state transitions occur by changing the kinetic parameters. Here, we perturb the parameters of the models in each cell cycle phase and observe their tendency to move toward the other cell cycle phases. We perturb the parameters using two schemes: directed perturbation and random perturbation.

In the directed perturbation, the perturbed parameters can be computed as $\omega = \theta + (\theta - \theta_r) D \eta$, where ω denotes the parameter set after perturbation, θ denotes the current parameter set, θ_r denotes the target parameter set (see below for details), D denotes the perturbation level, and η denotes a random number generated from a uniform distribution between 0 and 1.

First, we perturb the parameters towards the global average. Using the original stable steady state gene expression profile as the initial condition, we simulate each RACIPE model in a cell cycle phase with the perturbed kinetic parameters and identify the new stable steady state. The perturbation is performed 100 times (with different random numbers) for each model to compute the statistics (mean, m and standard deviation, s). The cell cycle phases for the new steady states are inferred using the neural net model, from which we compute the fraction (percent) of the new models in each cell cycle phase. We also calculate negative z-score by $(m_{wt} - m_{mt})/s$, where m_{wt} is the neural net predicted percent of the unperturbed models belonging a cell cycle phase, m_{mt} is the mean of the neural net predicted percent of perturbed models for 100 random perturbations, s is the standard deviation thereof. The above process is repeated at all the nine perturbation levels (0.005%, 0.05%, 0.50%, 1%, 2%, 3%, 5%, 10%, and 15%, respectively). In Figure S18**,** panel (A) shows the mean and standard deviation of percent states obtained at each perturbation level for each cell cycle phase and panel (B) shows the corresponding negative z-scores.

Second, using a similar approach, we perturb the parameters towards the average parameters of oscillatory models spanning four or more states (Figure S19). Third, we also

perturb the parameters towards the parameters from any oscillatory model spanning four or more states (Figure S20). These all three scenarios show the characteristic tendency of the RACIPE models belonging to any cell cycle phase to relax towards the next cell cycle phase.

In the random perturbation, the parameters can be computed as $\omega = \theta + cD\eta$, where ω denotes the parameter set after perturbation, θ denotes the current parameter set, c is a scaling factor expressed as the difference between the maximum and minimum values of the RACIPE-generated parameter range (Table S1, in Huang et al.(Huang et al., 2017)) divided by 200, D denotes the perturbation level, and η denotes a random number generated from a uniform distribution between -0.5 and 0.5. If any perturbed parameter exceeds the corresponding parameter range, we assign it to its nearest extreme value. In Figure S21**,** panel (A) shows the mean and standard deviation of percent states obtained at each perturbation level for each cell cycle phase and panel (B) shows the corresponding negative z-scores. Clearly, the relaxation tendency of the perturbed RACIPE models towards the next cell cycle phase is not present in this case.

13. Rate equations for the cell cycle GRC

Here, we list the detailed rate equations for the cell cycle GRC (shown in Figure 1A), which follow the generic equations (Equation S1 – Equation S7) described in Section 2.

Cln3

$$\frac{d([Cln3])}{dt} = g_{Cln3} - k_{Cln3}[Cln3] \qquad (S14)$$

SBF

$$\frac{d([SBF])}{dt} = \eta(g_{SBF}H^s([Whi5])H^s([Cln3])H^s([Cln1,2])H^s([Clb1,2]) - k_{SBF}[SBF]) \qquad (S15)$$

Whi5

$$\frac{d([Whi5])}{dt} = \eta(g_{Whi5}H^s([Cln3])H^s([Cln1,2]) - k_{Whi5}[Whi5]) \qquad (S16)$$

MBF

$$\frac{d([MBF])}{dt} = \eta(g_{MBF}H^s([Cln3])H^s([Whi5])H^s([Clb1,2]) - k_{MBF}[MBF]) \qquad (S17)$$

Cln1,2

$$\frac{d([Cln1,2])}{dt} = g_{Cln1,2}H^s([SBF]) - k_{Cln1,2}[Cln1,2] \qquad (S18)$$

Sic1

$$\frac{d([Sic1])}{dt} = gH^s([Swi5])\,H^s([Clb5,6])\,H^s([Clb1,2])\,H^s([Cdc14]) - k_{Sic1}[Sic1]H^s([Cln1,2]) \qquad (S19)$$

Clb5,6

$$\frac{d([Clb5,6])}{dt} = g_{Clb5,6}H^s([MBF])H^s([Sic1]) - k_{Clb5,6}[Clb5,6]H^s([Cdc20]) \qquad (S20)$$

Cdh1

$$\frac{d([Cdh1])}{dt} = \eta(g_{Cdh1}H^s([Cln1,2])\ H^s([Clb5,6])\ H^s([Clb1,2])\ H^s([Cdc14]) - k_{Cdh1}[Cdh1]) \qquad (S21)$$

Clb1,2

$$\frac{d([Clb1,2])}{dt} = g_{Clb1,2}H^s([Mcm1,SFF])H^s([Sic1])H^s([DNAS]) - k_{Clb1,2}\ [Clb1,2]H^s([Cdh1])H^s([Cdc20]) \qquad (S22)$$

DNA Synthesis (DNAS)

$$\frac{d([DNAS])}{dt} = g_{DNAS}H^S([Clb5,6]) - k_{DNAS}[DNAS] \qquad (S23)$$

Cdc20

$$\frac{d([Cdc20])}{dt} = g_{Cdc20}H^s([Mcm1,SFF])H^s([Clb1,2]) - k_{Cdc20}[Cdc20]H^s([Cdh1]) \qquad (S24)$$

Mcm1,SFF

$$\frac{d([Mcm1,SFF])}{dt} = \eta\left(g_{Mcm1,SFF}H^s([Clb1,2])H^s([DNAS]) - k_{Mcm1,SFF}[Mcm1,SFF]\right) \qquad (S25)$$

Pds1

$$\frac{d([Pds1])}{dt} = g_{Pds1} - k_{Pds1}[Pds1]H^s([Cdc20]) \qquad (S26)$$

Cdc14

$$\frac{d([Cdc14])}{dt} = \eta(g_{Cdc14}H^s([Pds1]) - k_{Cdc14}[Cdc14]) \qquad (S27)$$

Swi5

$$\frac{d([Swi5])}{dt} = g_{Swi5}H^s([Mcm1,SFF])\ H^s([Clb1,2])\ H^s([Cdc14]) - k_{Swi5}[Swi5] \qquad (S28)$$

Supplemental Figures

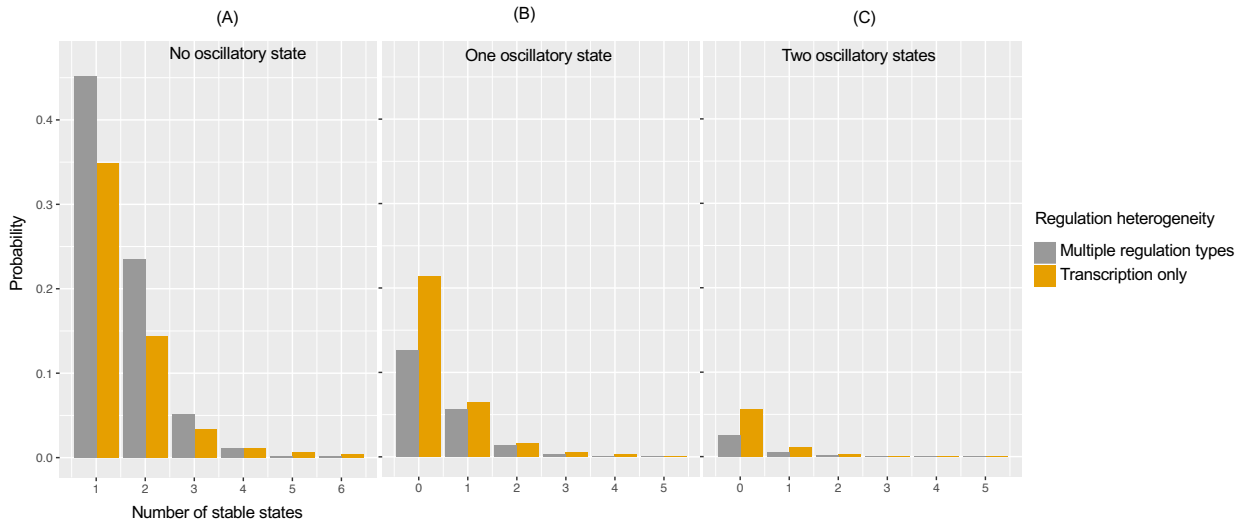


Figure S1. Convergence of the number of stable states from an ensemble of RACIPE models – multiple regulation types versus only transcription, related to Figures 1A-1D. The histograms of stable steady states and oscillatory states were calculated for 10000 models, generated for the cell cycle GRC (grey, considering multiple types of regulations presented in the GRC) and a GRC of the same topology but replacing regulation types to transcriptional regulations only (orange). Each panel shows the histogram of the number of stable steady states of each model for **(A)** models with no oscillatory state, **(B)** models with one oscillatory state, and **(C)** models with two oscillatory states. In comparison, the GRC with only transcriptional regulations has a higher chance to generate oscillations.

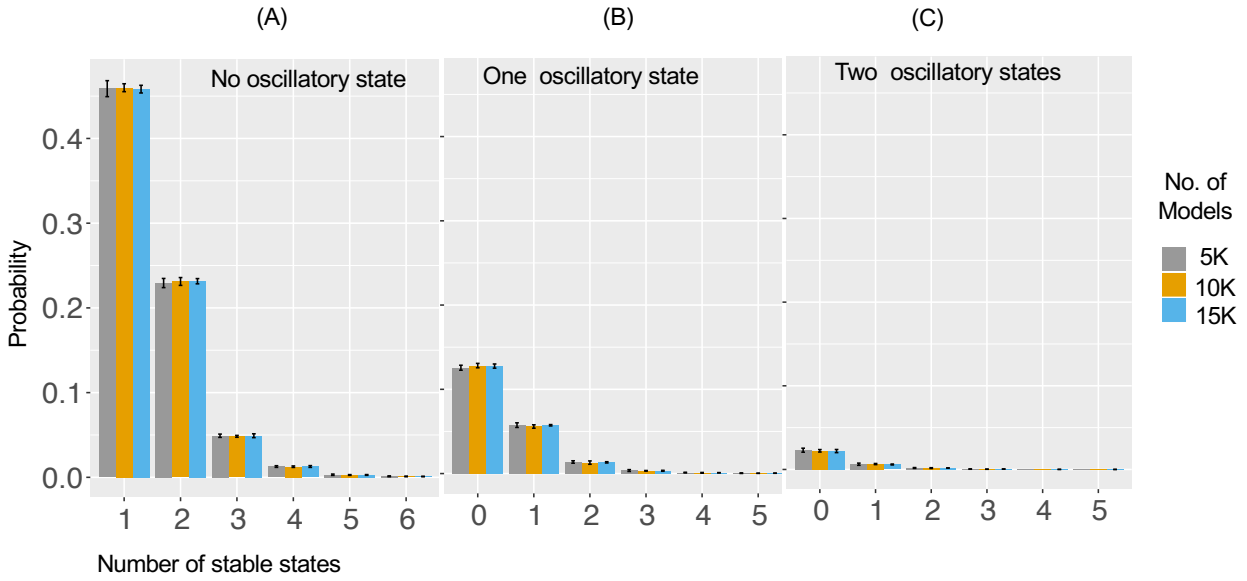


Figure S2. Convergence of the number of stable states from an ensemble of RACIPE models of different sample sizes, related to Figures 1A-1D**.** The histograms of stable steady states and oscillatory states were calculated for 5000 (5K, grey), 10000 (10K, orange), and 15000 (15K, blue) models, each with eight independent samples. Each panel shows the mean and standard deviation of the histogram of the number of stable steady states of each model for **(A)** models with no oscillatory state, **(B)** models with one oscillatory state, and **(C)** models with two oscillatory states, shown as mean ± SD.

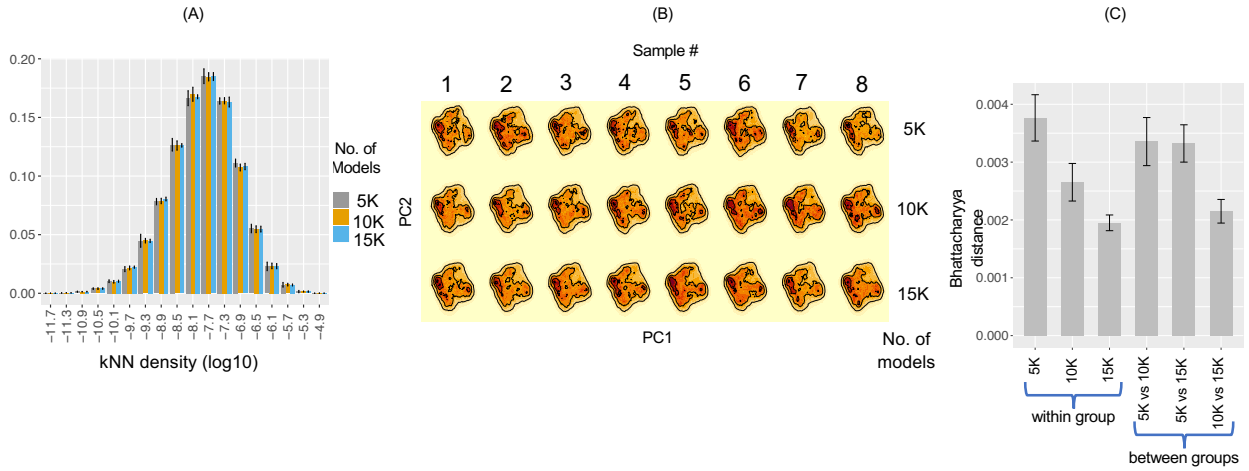


Figure S3. Convergence of the distribution of steady-state gene activity profiles, related to Figures 1A-1D**.** RACIPE was applied to the cell cycle GRC to generate 5000 (5K), 10000 (10K), and 15000 (15K) models, each with eight repeats for statistical analysis. For each case, the simulated steady-state gene activity profiles were log transformed and standardized, before k nearest neighbor (kNN) densities were estimated **(A)** Distributions of the kNN densities estimated from the high-dimensional steady-state gene activity profiles, shown as mean ± SD. **(B)** kNN densities estimated for each dataset projected on the first two principal components (PCs). The PCs were obtained from the 10K models used in Figure 1. kNN density was estimated at each grid point on an 81x78 grid and the number of neighbors (k) was used were 100, 200, and 300 for the sample sizes 5K, 10K, and 15K, respectively. **(C)** Means and standard deviations of Bhattacharyya (BC) distances between each pair of kNN densities from panel (b) within group (same sample size) and between groups (different sample sizes), shown as mean ± SD.

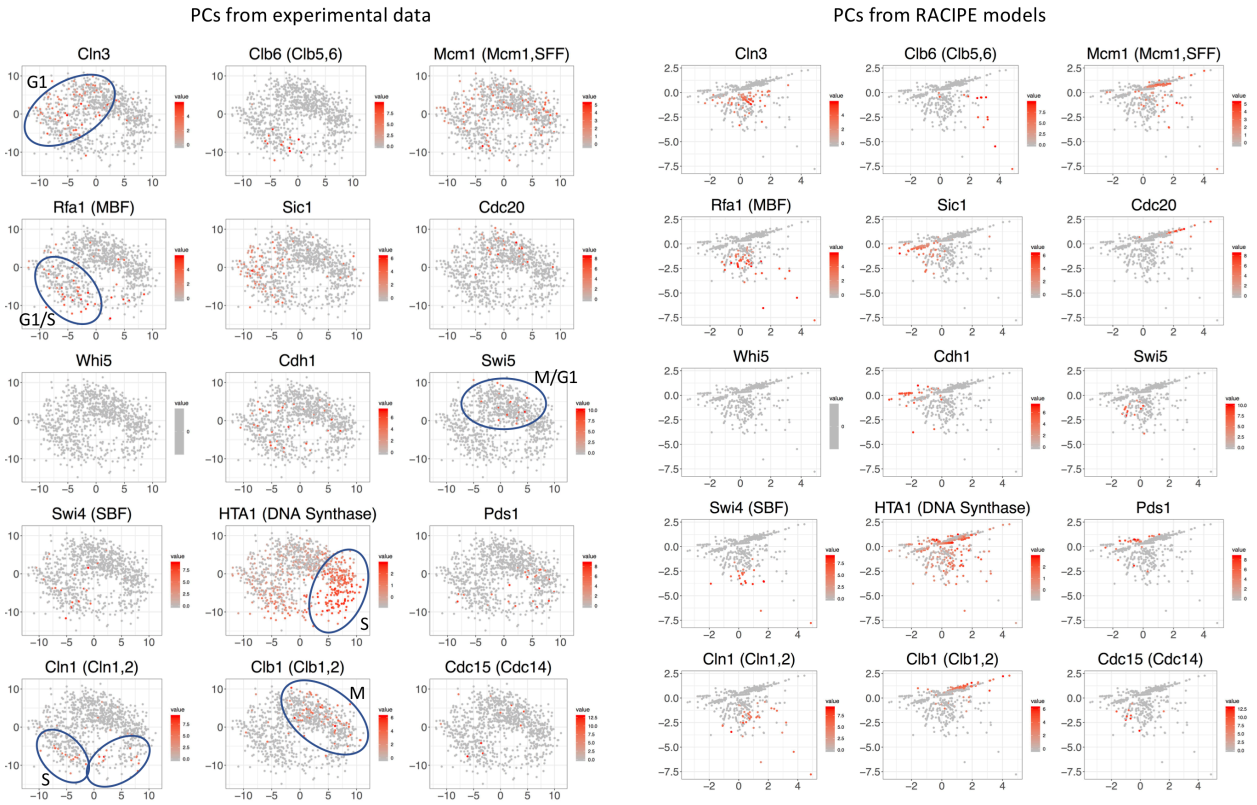


Figure S4. Single cell gene expressions (before smoothing) of the selected network genes projected onto the first two principal components (PCs), related to Figures 1E,1F**.** Left panel – the PCs obtained from the experimental data (not smoothed). Right panel - the PCs from the RACIPE models.

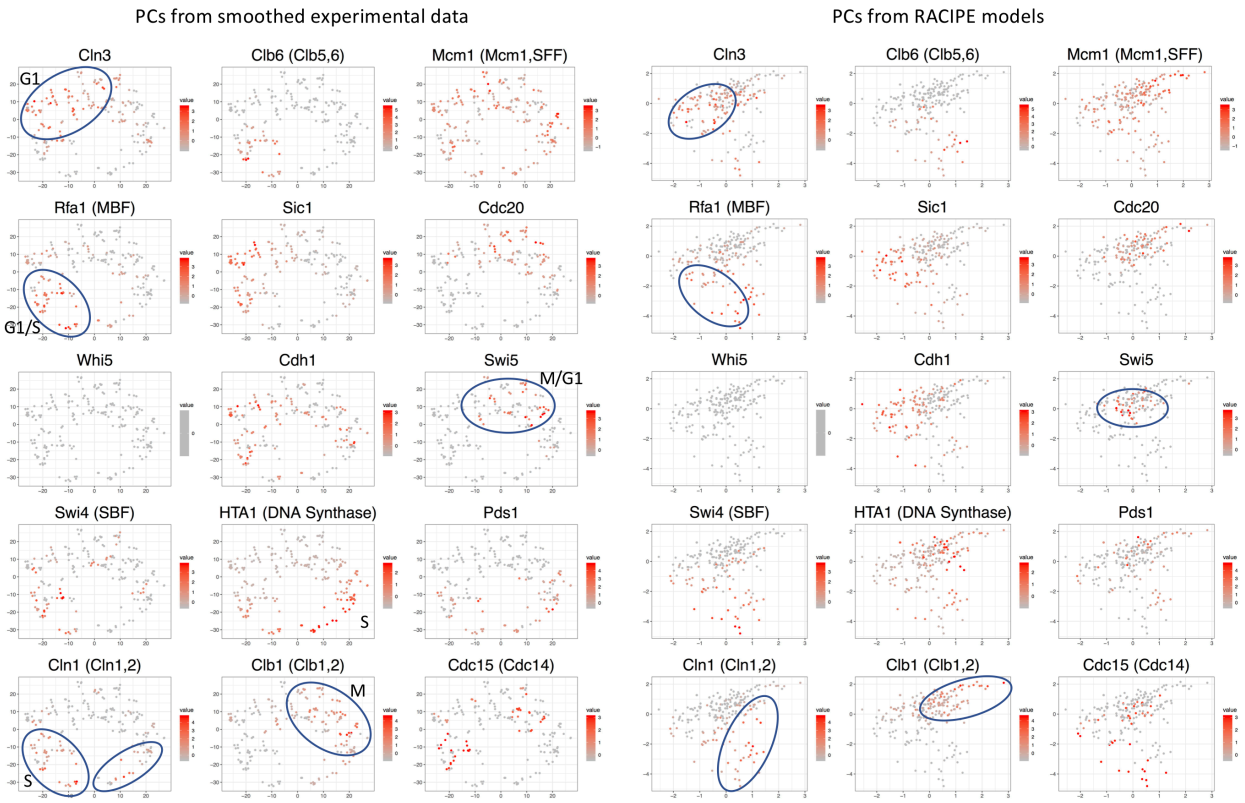


Figure S5. Single cell gene expressions (after smoothing) of the selected network genes projected onto the first two principal components (PCs), related to Figures 1E, 1F**.** Left panel - the PCs obtained from the experimental data (smoothed). Right panel - the PCs from the RACIPE models.

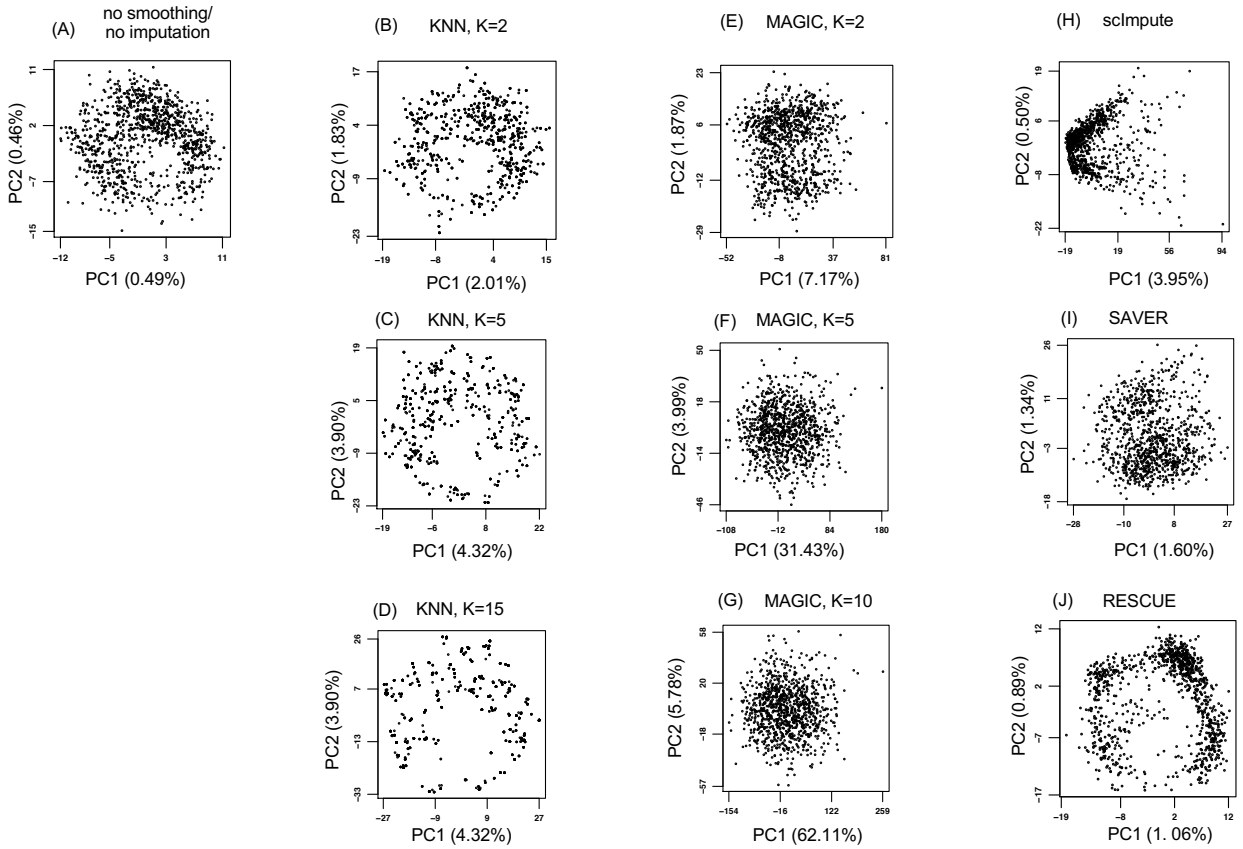


Figure S6. Single cell gene expressions of 5624 genes across 976 WT cells projected onto the first two principal components (PC1 and PC2), related to Figures 1E, 1F**. (A)** no smoothing/imputation, and **(B-J)** data smoothed/imputed by specific methods: **(B-D)** K-nearest neighbor (KNN) smoothing for K=2, 5, and 15, respectively, **(E-G)** MAGIC imputation for K=2, 5, 10, respectively, **(H)** scImpute imputation, **(I)** SAVER imputation, and **(J)** RESCUE imputation. The PCs on each panel are obtained from the respective dataset – **(A)** data before smoothing/imputation and **(B-J)** data smoothed/imputed by specific methods and parameter values.

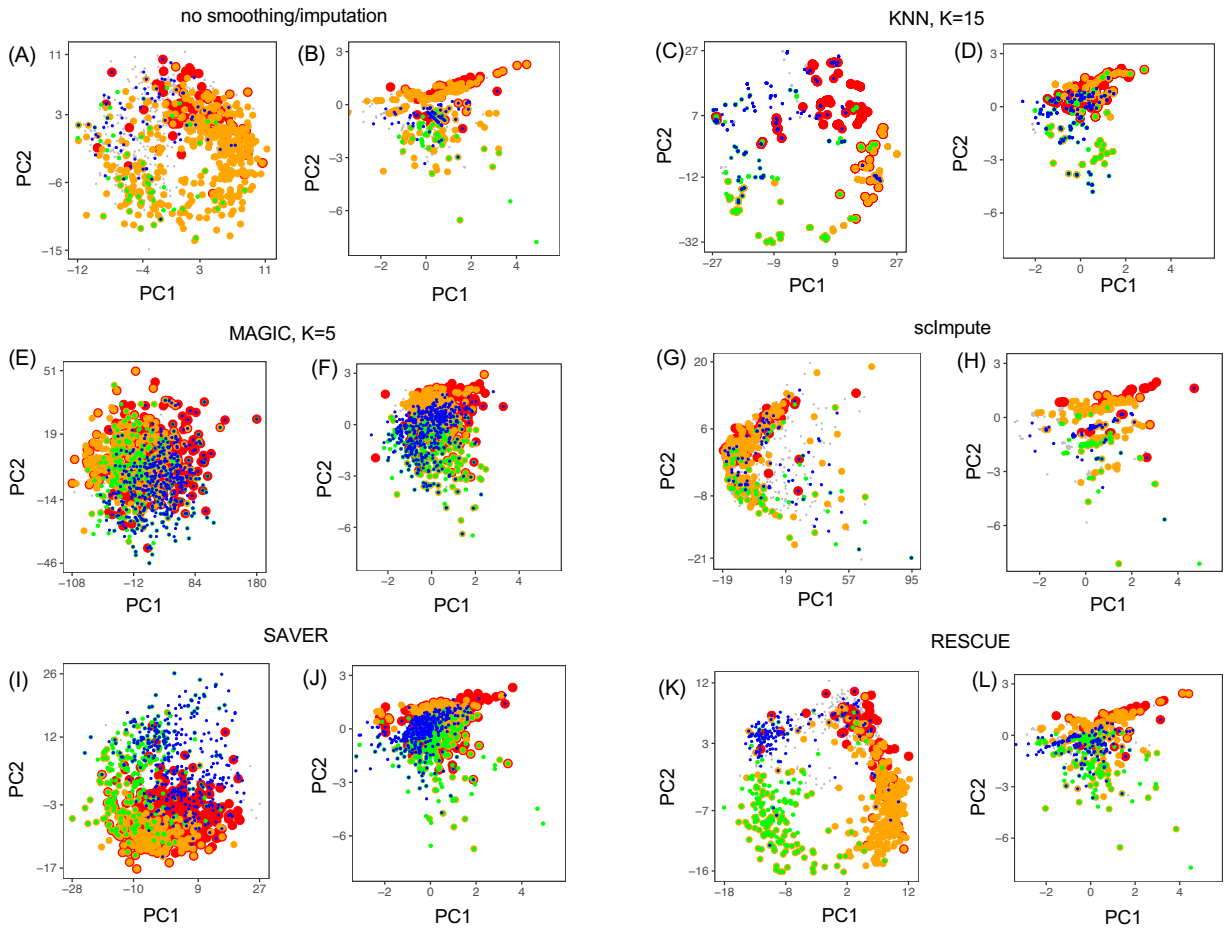


Figure S7. Annotation of four selected network genes (Cln3, Rfa1, Hta1, and Clb1) on projected single cell gene expression profiles, related to Figures 1E, 1F**.** The four marker genes correspond to specific cell cycle phases - Cln3:G1 (blue), Rfa1: G1/S (green), Hta1:S (orange), and Clb1:M (red). For the left of each panel pair, which are A, C, E, G, I, and K, the PCs are obtained from the genome-wide gene expression profiles. For the right of each panel pair, which are B, D, F, H, J, and L, the PCs are obtained from the RACIPE models, and the projection was performed using the gene expression of the corresponding network genes. The gene expressions are used either before smoothing/imputation or after: **(A, B)** data before smoothing/imputation and **(C-L)** data smoothed/imputed by specific methods: **(C, D)** K-nearest neighbor (KNN) smoothing with K=15, (E, F) MAGIC imputation for K=5, **(G, H)** scImpute imputation, **(I, J)** SAVER imputation, and **(K, L)** RESCUE imputation.

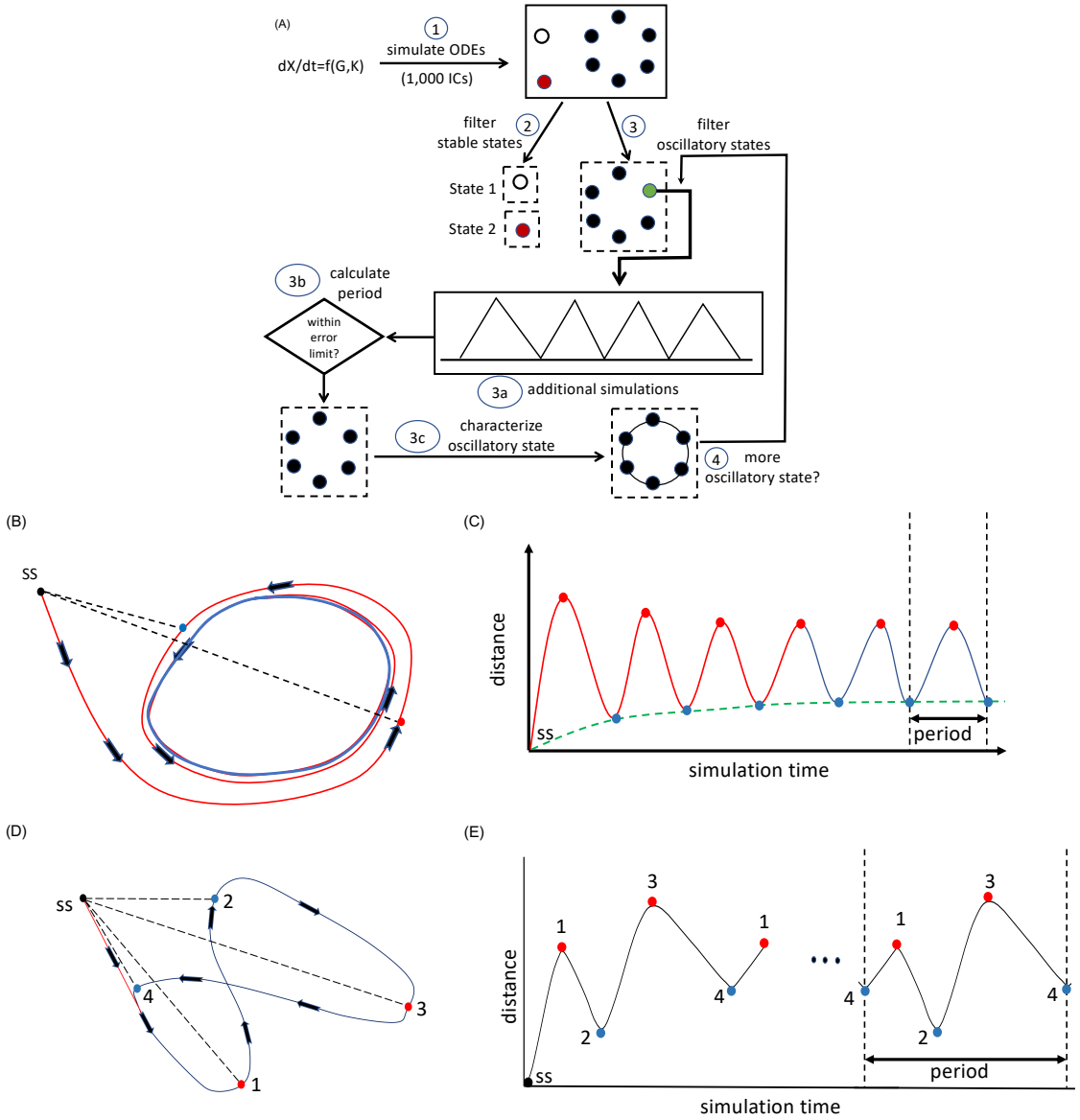


Figure S8. Algorithm for detecting oscillatory states, related to Figure 1A. **(A)** Flowchart for detecting the coexistence of steady states and oscillatory states in the generalized RACIPE. **(B-E)** Depicted are two relaxation scenarios generating oscillatory states. (B and C: Scenario 1). **(B)** Schematic of relaxation simulation. Plot shows a simulated trajectory from the starting nonsteady state, SS, and moving towards (red curve) a stable oscillatory state (blue curve). The distance from SS to each state along the simulation was calculated. This distance starts from zero and then repeatedly goes up and down. The red point on the projected trajectory marks where the distance reaches a high point and the blue point marks where it reaches a low point. **(C)** The distance function (as defined on panel a) as a function of simulation time. The local maxima and minima are marked with red and blue dots, respectively. (D and E: Scenario 2). **(D)** Another relaxation simulation leading to a more complicated oscillatory state. The two states, colored blue and labeled 2 and 4, are two local minima. The two states, colored red and labeled 1 and 3, are two local maxima. **(E)** The distance function shows that the period can be calculated between two consecutive local minima at the same distance.

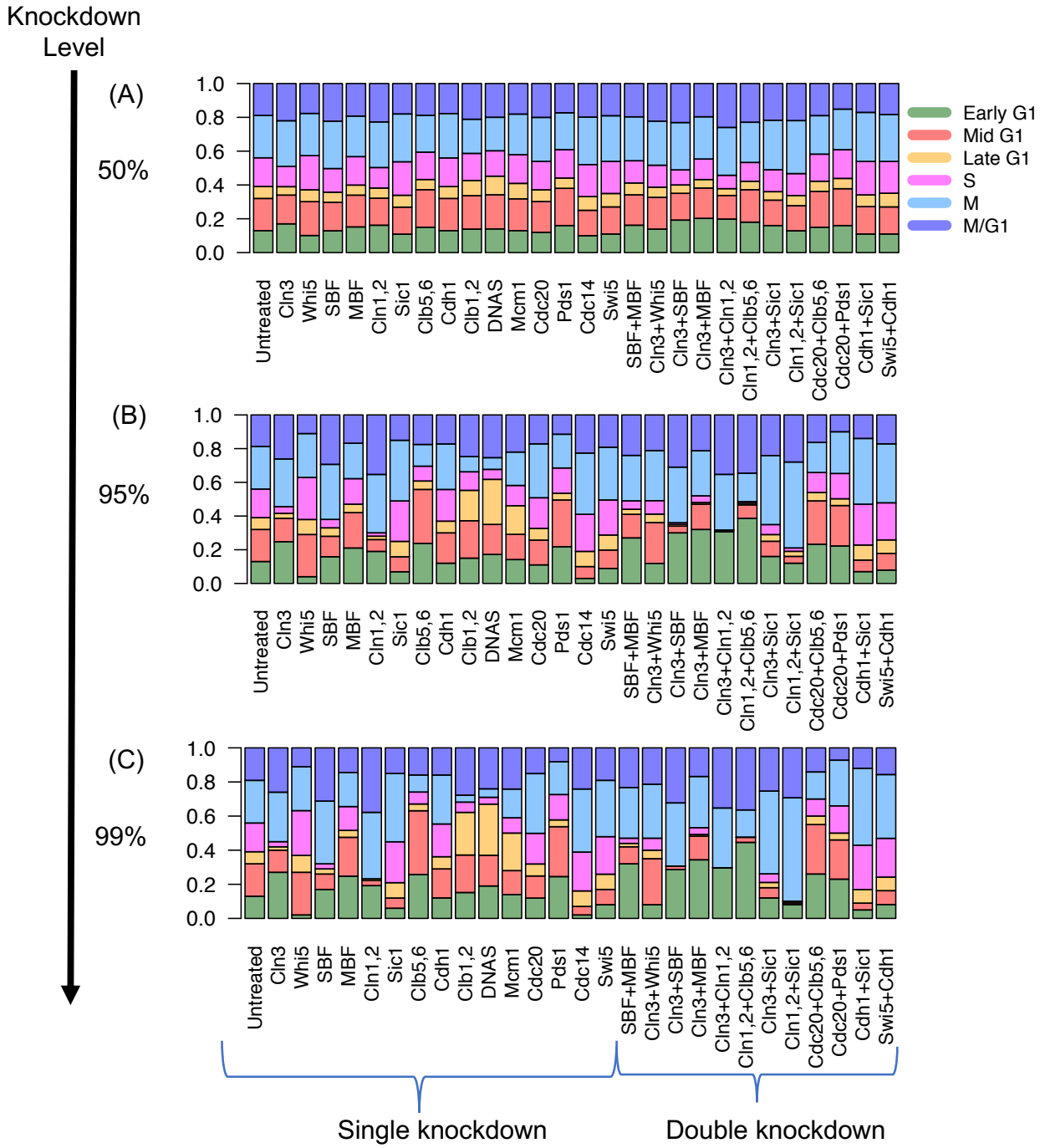


Figure S9. Distribution of the fraction of models in each cell cycle phase from the simulations of untreated and knockdown conditions at three knockdown levels, related to Figure 2**.** For the three knockdown levels, the maximum production rate for a gene was reduced by **(A)** 50%, **(B)** 95%, and **(C)** 99%. The fraction of models in each cell cycle phase for the 95% reduction is very close to 99% reduction.

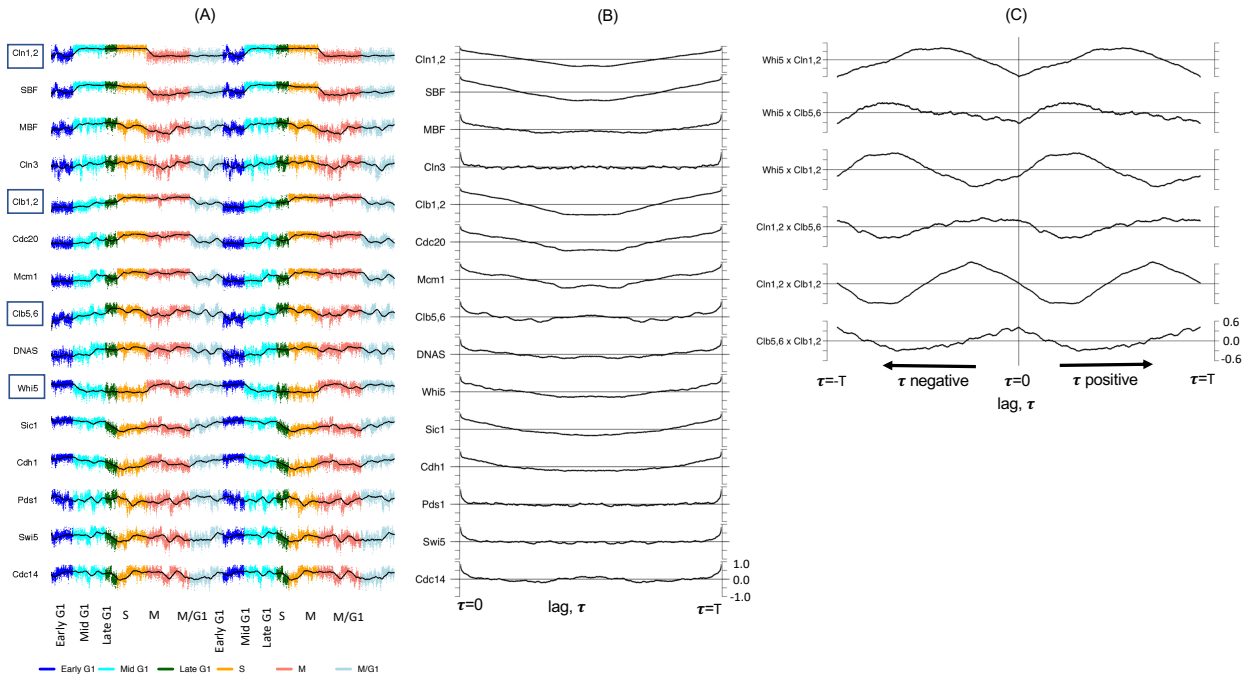


Figure S10. Cell cycle time dynamics reconstructed using hierarchical clustering analysis, related to Figure 3A**. (A)** Cell cycle time dynamics reconstructed from the RACIPE models (2 periods). Each panel shows the simulated activity of every gene in the circuit model along the pseudotime (see Section 9). The dynamical gene activity profiles illustrate the transitions from early G1 to mid G1 to late G1 to S to M to M/G1 (denoted by different colors). The change of the four marked genes (Whi5, Cln1,2, Clb5,6, and Clb1,2) along the pseudotime and how they are related to the cell cycle progression are described in the main text (Figure 3A). **(B and C)** Auto and cross correlations. We calculate cross correlation, f = corr (X(t), Y(t+τ)) where corr is the Pearson's correlation between the activity of gene X and the activity of gene Y with time lag τ. When X and Y denote the same gene, the function gives autocorrelation. The horizontal axis in both panels shows the time lag τ, annotated by the time course of cell cycle phases. **(B)** Autocorrelation of gene activities. The plot illustrates that most genes show clear periodicity, except for Cdc14, Swi5, Pds1, and Cln3. **(C)** Cross-correlation of the four selected genes (Whi5, Cln1,2, Clb5,6, and Clb1,2) illustrates their order of activity. As an example, in case of Whi5 x Clb1,2 (third row), function f peaks at positive τ implying that Whi5 activity is ahead of Clb1,2 activity. In case of Whi5 x Clb1,2 (top row), we observe the activities of the two genes have opposite phases.

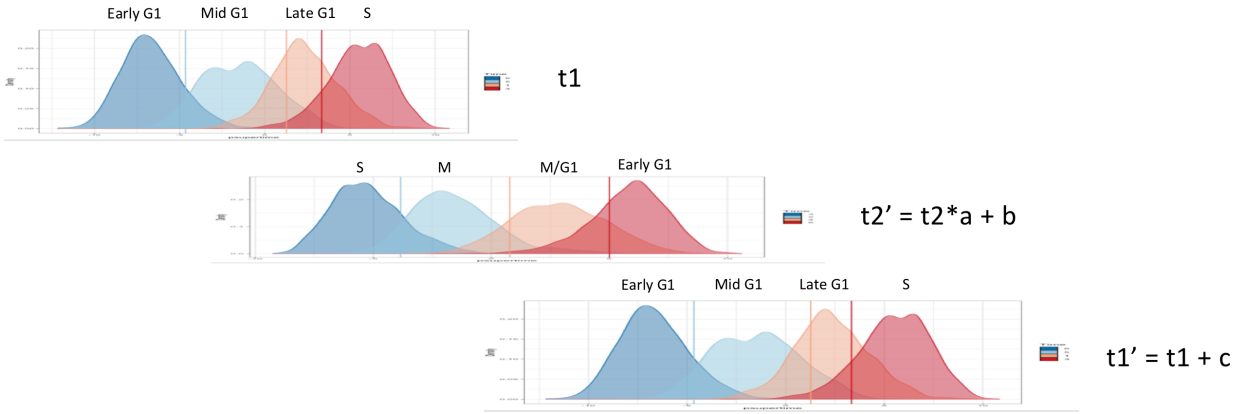


Figure S11. Illustration of psupertime adaptation for pseudotime inference, related to Figure 3A**.** We applied psupertime twice – the first time with data from Early G1 to S phases (first line), and the second time with data from S to Early G1 (second line). The inferred pseudotime indices were annotated as t1 and t2, respectively. We then scaled t2 to t2′ so that the distributions of t1 and t2′ for the S phase match, and the same for the early G1 phase.

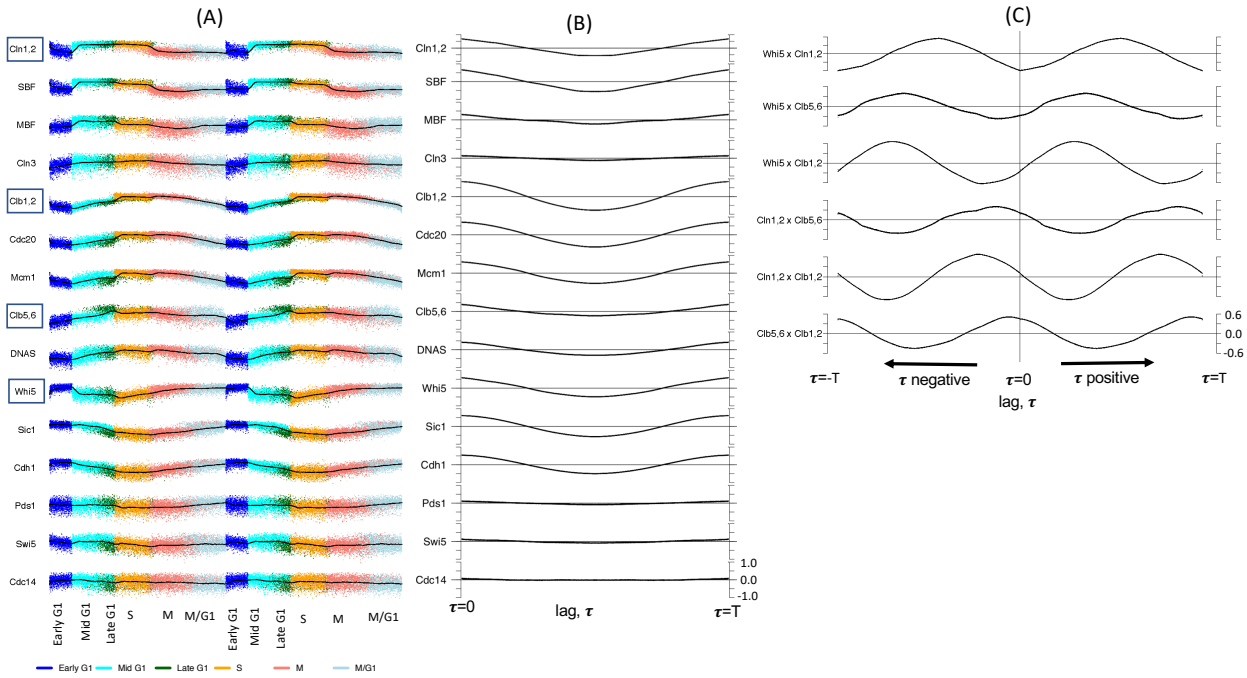


Figure S12. Cell cycle time dynamics reconstructed using adapted psupertime method, related to Figure 3A**. (A)** Cell cycle time dynamics reconstructed from the RACIPE models (2 periods). Each panel shows the simulated activity of every gene in the circuit model along the pseudotime (see Section 9). The dynamical gene activity profiles illustrate the transitions from early G1 to mid G1 to late G1 to S to M to M/G1 (denoted by different colors). The change of the four marked genes (Whi5, Cln1,2, Clb5,6, and Clb1,2) along the pseudotime and how they are related to the cell cycle progression are described in the main text (**Figure 3A**). **(B and C) Auto and cross correlations**. We calculate cross correlation, f = corr (X(t), Y(t+τ)) where corr is the Pearson's correlation between the activity of gene X and the activity of gene Y with time lag τ. When X and Y denote the same gene, the function gives autocorrelation. The horizontal axis in both panels shows the time lag τ, annotated by the time course of cell cycle phases. **(B) Autocorrelation of gene activities.** The plot illustrates that most genes show clear periodicity, except for Cdc14, Swi5, Pds1, and Cln3. **(C) Cross-correlation of the four selected genes (Whi5, Cln1,2, Clb5,6, and Clb1,2) illustrates their order of activity.** As an example, in case of Whi5 x Clb1,2 (third row), **f** peaks at positive τ implying that Whi5 activity is ahead of Clb1,2 activity. In case of Whi5 x Clb1,2 (top row), we observe the activities of the two genes have opposite phases.

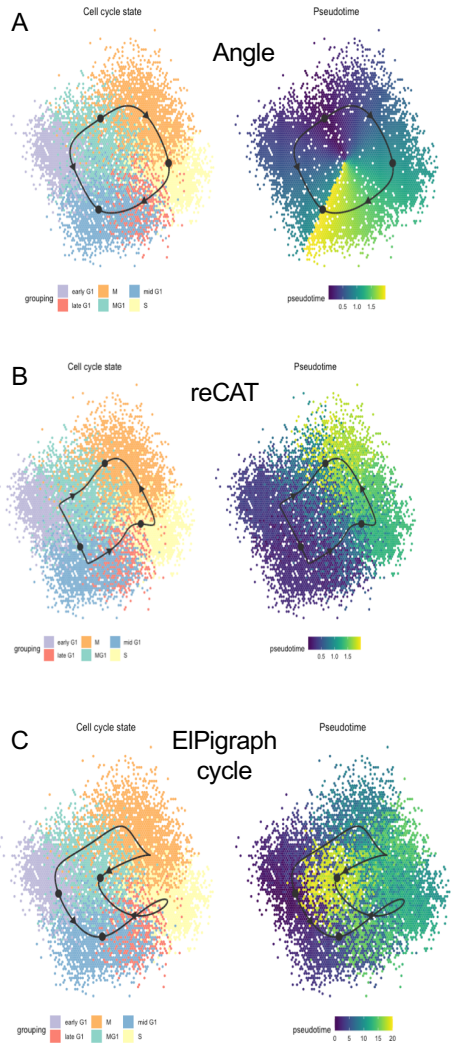


Figure S13. Pseudotime results using Angle, reCAT, ElPigraph, related to Figure 3A**.** Trajectories (left panels) and pseudotimes (right panels) inferred by three trajectory inference methods, **(A)** Angle, **(B)** reCAT, **(C)** ElPigraph cycle, projected on the first two principal components of the simulated gene activities. Each method inferred pseudotimes with a bifurcating structure, instead of a circular structure.

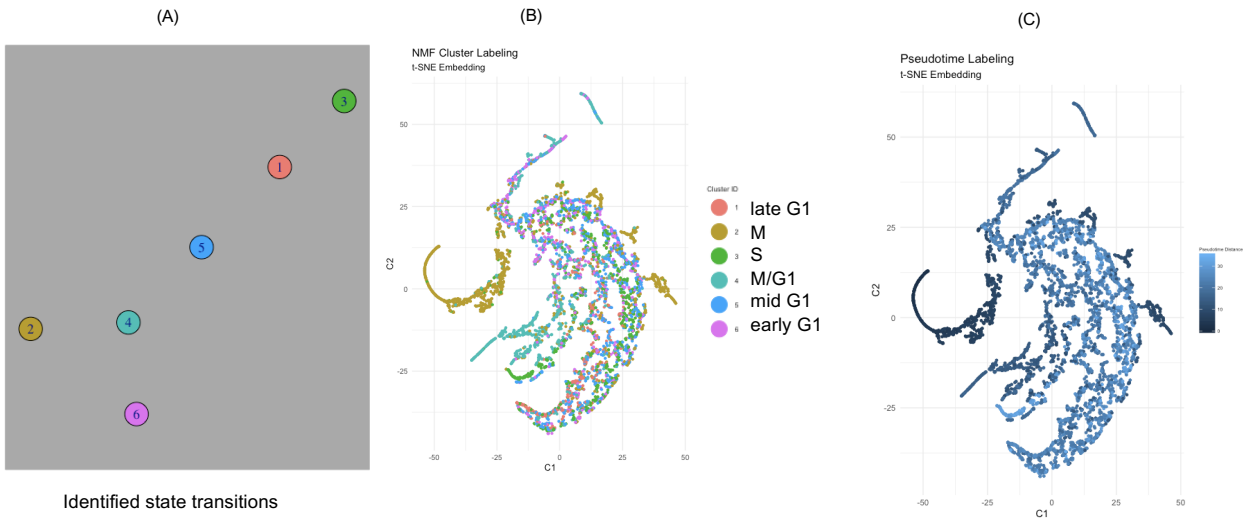
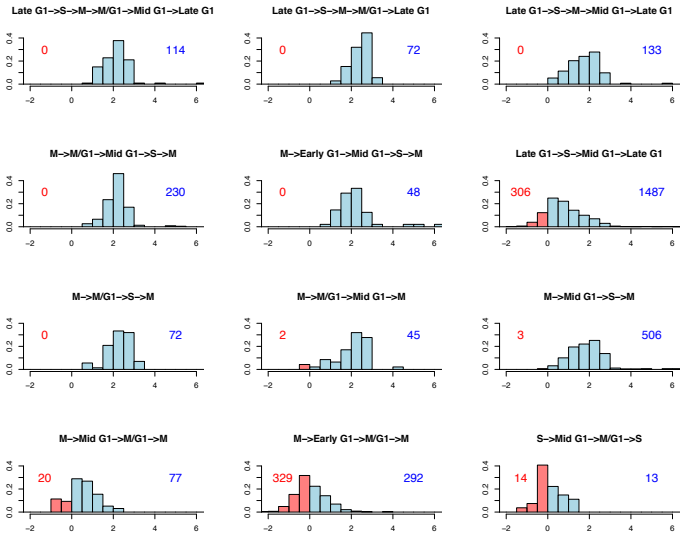


Figure S14. Pseudotime results using SoptSC, related to Figure 3A**.** (A) Minimum spanning tree (MST) on the cluster-cluster graph. The method coul-d not find a complete MST and only identified a subset of transitions as shown at the bottom, (B) models plotted on the first two t-SNE components and colored by their cell cycle state, (C) models colored by the computed pseudotime distance.



spanned clusters	patterns	counts	patterns	counts
6	1324651	3		
5	132451	114	246532	11
	132651	13		
4	24532	230	26532	48
	13251	133	12451	11
	13241	72		
3	1351	1793	2452	47
	2642	621	3543	27
	2532	509	1341	17
	2542	97	2652	17
	2432	72	1451	14

Figure S15. Polarity distributions for oscillatory states with various state transition patterns, related to Figure 4C. Polarity distribution for 12 highest frequency patterns (left) and their counts (right). Six of the patterns show positive polarity (light blue) only, whereas six show both negative (red) and positive (light blue) polarity. Numeric code used for cell cycle phases: 1 – early G1, 2 – mid G1, 3 – late G1, 4 – S, 5 – M, and 6 – M/G1.

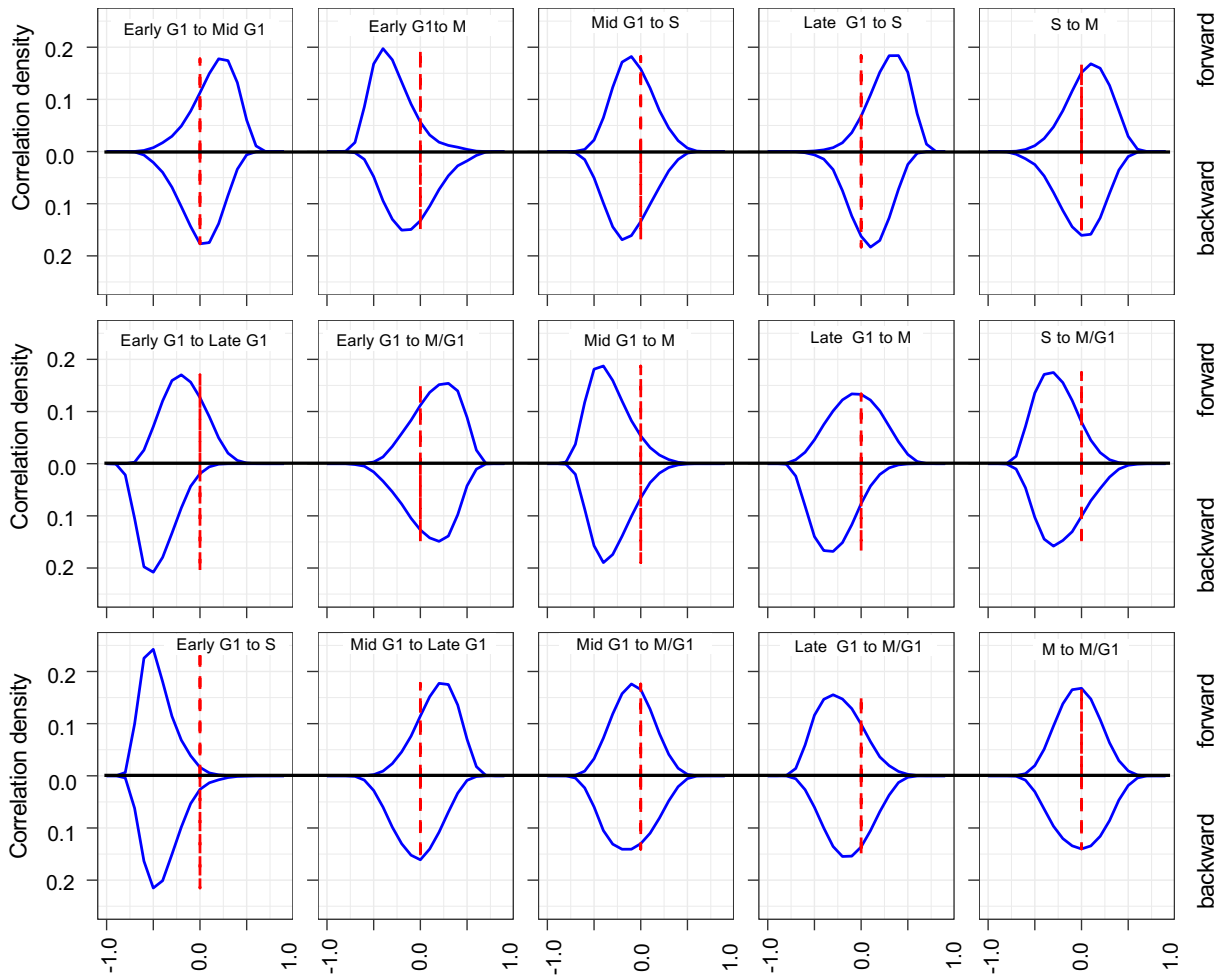


Figure S16. Probability density of the forward and backward delayed correlations between two cell cycle phases, related to Figure 5A**.** Density of delayed correlations on each panel corresponds to any two cell cycle phases – on each panel, (top): the forward correlation and (bottom): the backward correlation. Vertical dotted lines show the zero correlation values.

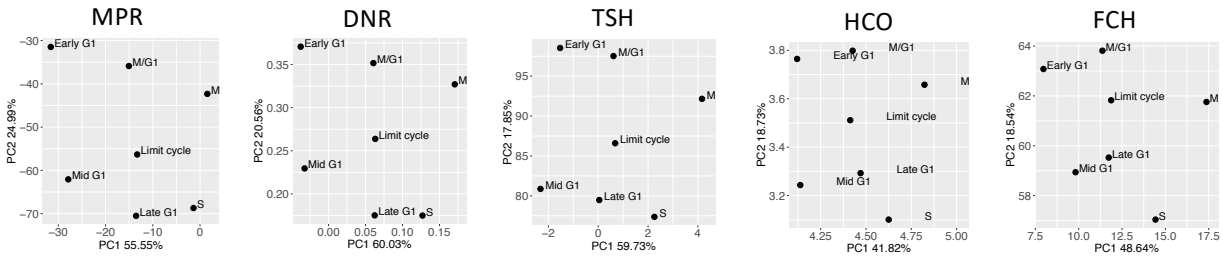


Figure S17. Average parameters (individual) projected on PC space, related to Figure 5D**.** MPR –production rate, DNR – degradation rate, TSH – threshold, HCO – Hill coefficient, FCH – fold change.

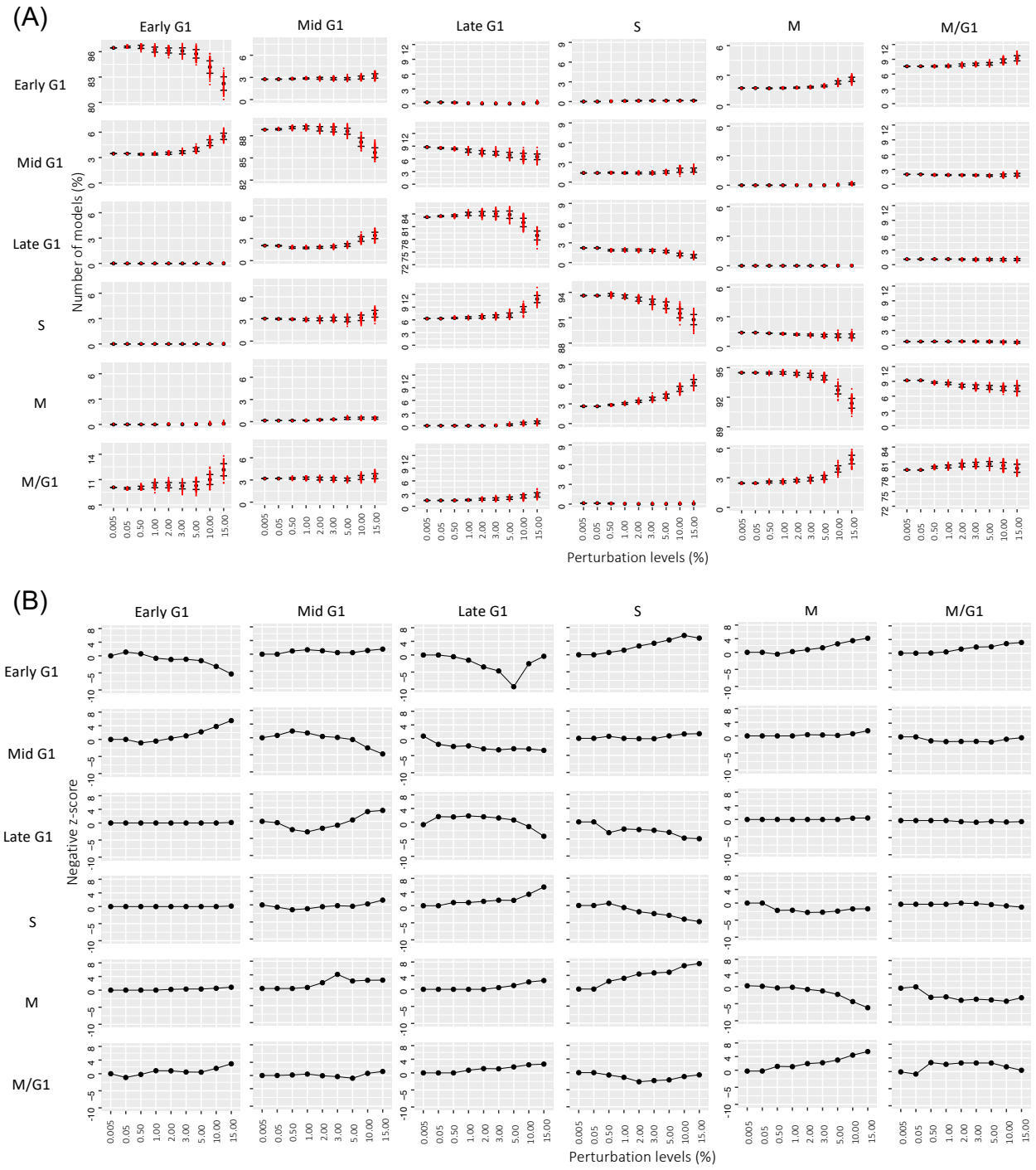


Figure S18. Simulation of RACIPE models after relaxing parameters toward global parameter average, related to Figure 5E**. (A)** Mean and standard deviation (SD) of percent models at all nine perturbation levels show the transitions of the perturbed models away from the original states. Models from any cell cycle phase mainly move to the subsequent cell cycle phase, shown as mean ± SD. **(B)** Distribution of the negative z-scores at all nine perturbation levels. At larger perturbation level, the negative z-score is also larger. This is consistent with the outcomes from panel (A).

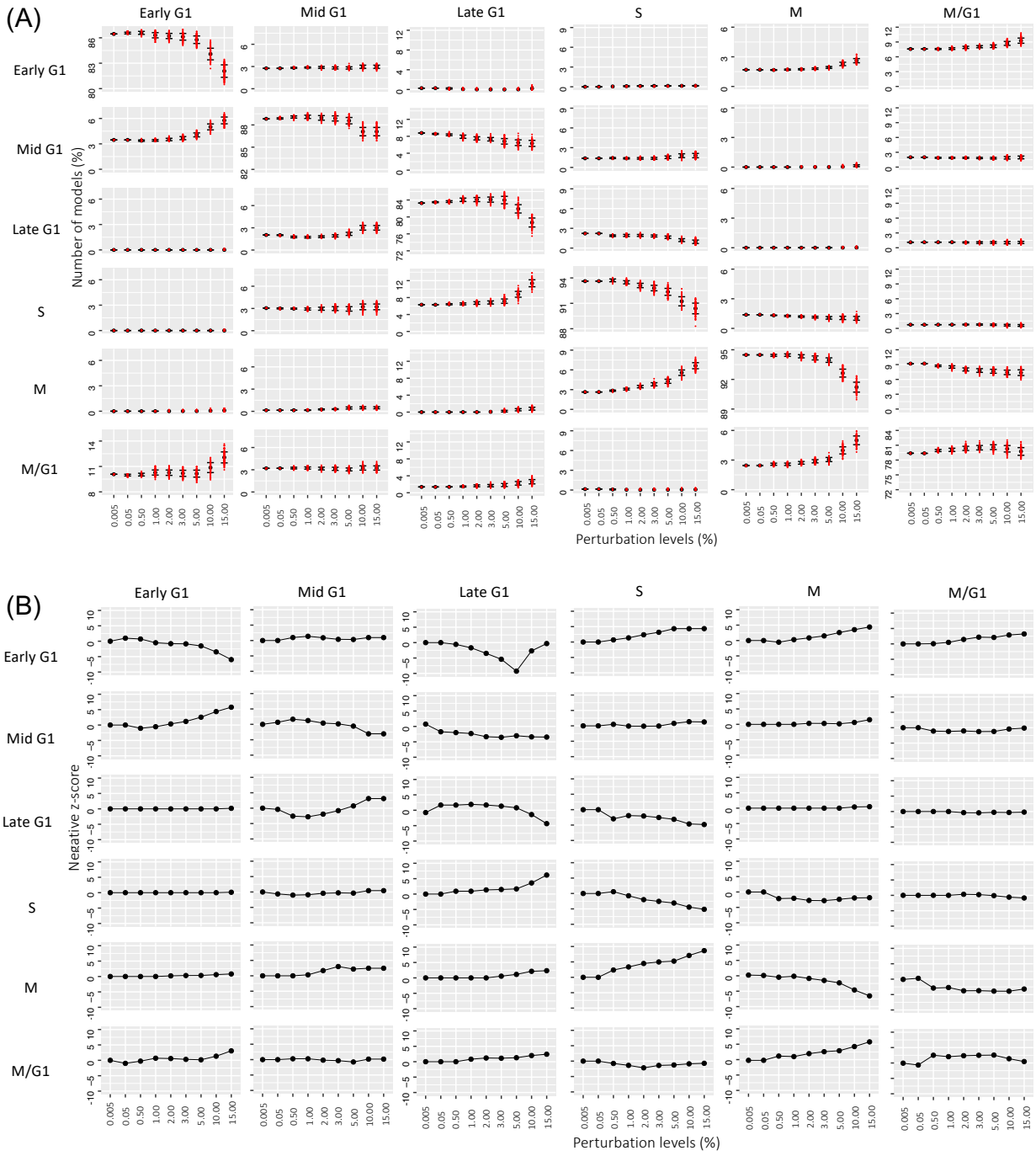


Figure S19. Simulation of RACIPE models after relaxing parameters toward the average parameters of all limit cycle models, related to Figure 5E**. (A)** Mean and SD of percent of models at different perturbation levels, shown as mean ± SD. **(B)** Distribution of negative z-scores for all perturbation levels. See Figure S18 for details**.**

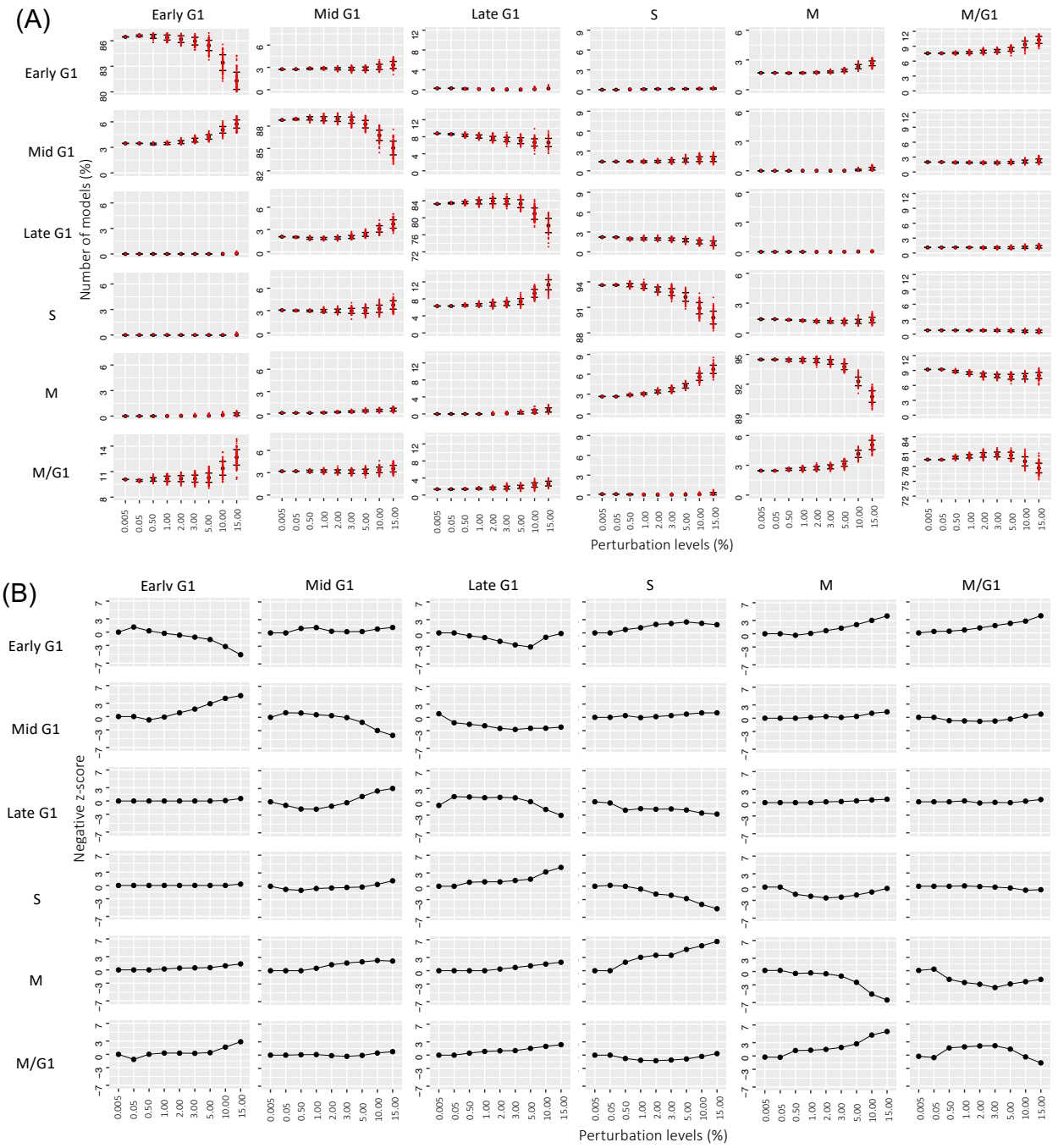


Figure S20. Simulation of RACIPE models after relaxing parameters toward the parameters of any randomly selected limit cycle, related to Figure 5E**.** (A) Mean and SD of percent of models at different perturbation levels, shown as mean ± SD. (B) Distribution of negative z-scores for all perturbation levels. See Figure S18 for details.

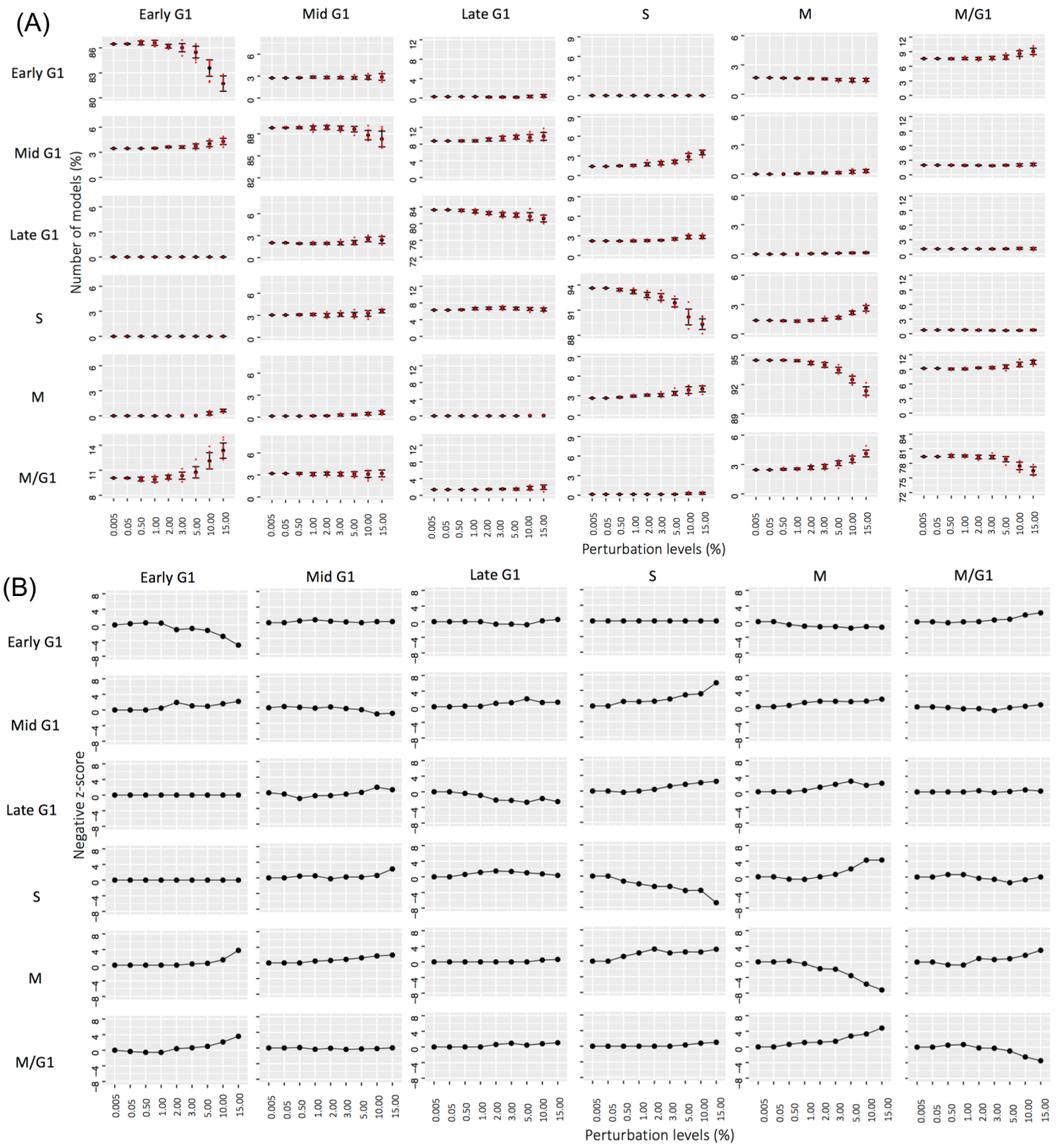
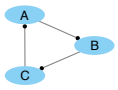
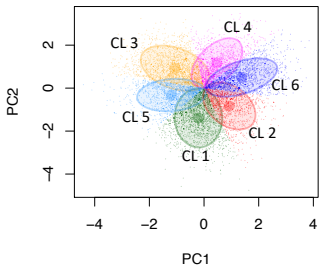


Figure S21. Simulation of RACIPE models after random perturbation of the parameters, related to Figure 5E**.** (A) Mean and SD of percent of models at different perturbation levels, shown as mean ± SD. (B) Distribution of negative z-scores for all perturbation levels. See Figure S18 for details**.**

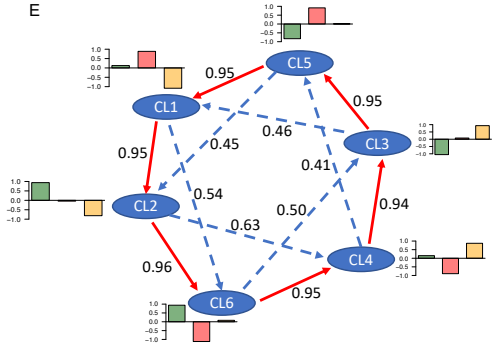
A



C



E



B

CL #	% models
1	21
2	14
3	21
4	14
5	13
6	17

D

	CL1	CL2	CL3	CL4	CL5	CL6
CL1	NA	**0.95**	-0.95	-0.30	-0.59	**0.54**
CL2	-0.46	NA	-0.50	**0.63**	-0.95	**0.96**
CL3	**0.46**	-0.50	NA	-0.64	**0.95**	-0.96
CL4	-0.67	-0.95	**0.94**	NA	**0.41**	-0.34
CL5	**0.95**	**0.45**	-0.45	-0.96	NA	-0.56
CL6	-0.96	-0.51	**0.50**	**0.95**	-0.44	NA

F

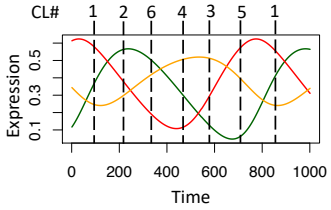


Figure S22. Modeling a repressilator circuit, related to Figure 6C**. (A)** Illustration of the circuit topology. **(B)** Percentage of models in each of the six steady state clusters obtained hierarchical clustering of the RACIPE simulated models. The slight differences in the numbers (e.g. for CL# 1, 3, and 6) are because of the random sampling in RACIPE and the clustering algorithm. But the modeling statistics for state transitions, as described below, are robust. **(C)** The steady state gene expression profiles, colored by the identity of clusters, projected onto the first two principal components. **(D)** Median of the delayed correlations computed between any two clusters – dominant correlations (>0.90) are marked as red bold. **(E)** A model for state transitions based on the delayed correlations from panel d. The thick red line formed by the dominant state transitions shows the path for the major state transitions. The bar plots show the average expression of the three genes for different clusters. **(F)** A continuous simulation of the repressilator circuit shows the dynamic change of the expression of the three genes. Vertical dotted lines show snapshots of gene expression patterns that correspond to the six clusters. The order of the states along the time trajectory matches the inferred results from the delayed correlations.

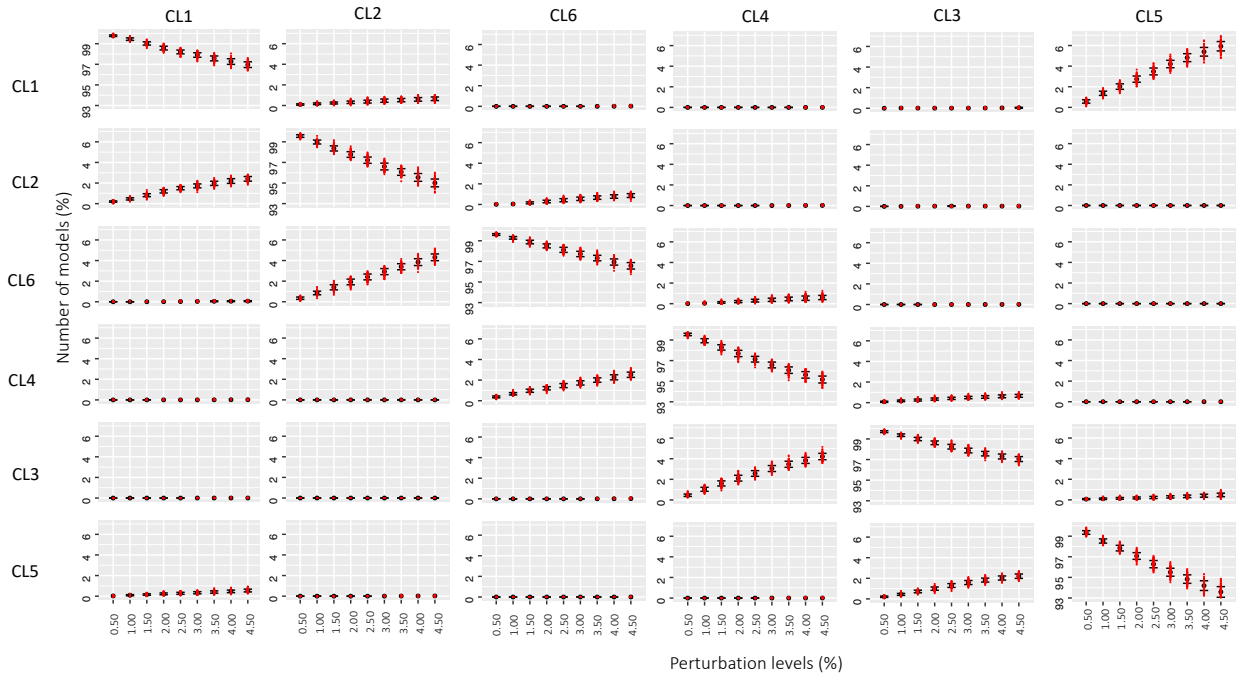


Figure S23. Repressilator - state transitions from simulation of models with perturbed parameters, related to Figure 6C**.** Each panel corresponds to simulation of the models belonging to the cluster number labeled at the top. The simulations were repeated 100 times at each of the nine perturbation levels (0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, and 4.5%, respectively). Mean and SD of the percent of models at each panel (shown as mean ± SD) reveal how the model population in the corresponding cluster changes and how that change affects the other clusters. For example, from the panels on the first column, we observe that the percent of models in CL1 decreases more and more as the perturbation level increases, whereas that in CL2 increases consistently and no major changes in the other clusters, implying a state transition CL1 → CL2. According to the data from all of these panels, we can conclude the following state transition CL1 → CL2 → CL6 → CL4 → CL3 → CL5 → CL1, which is also consistent with Figure S18E.

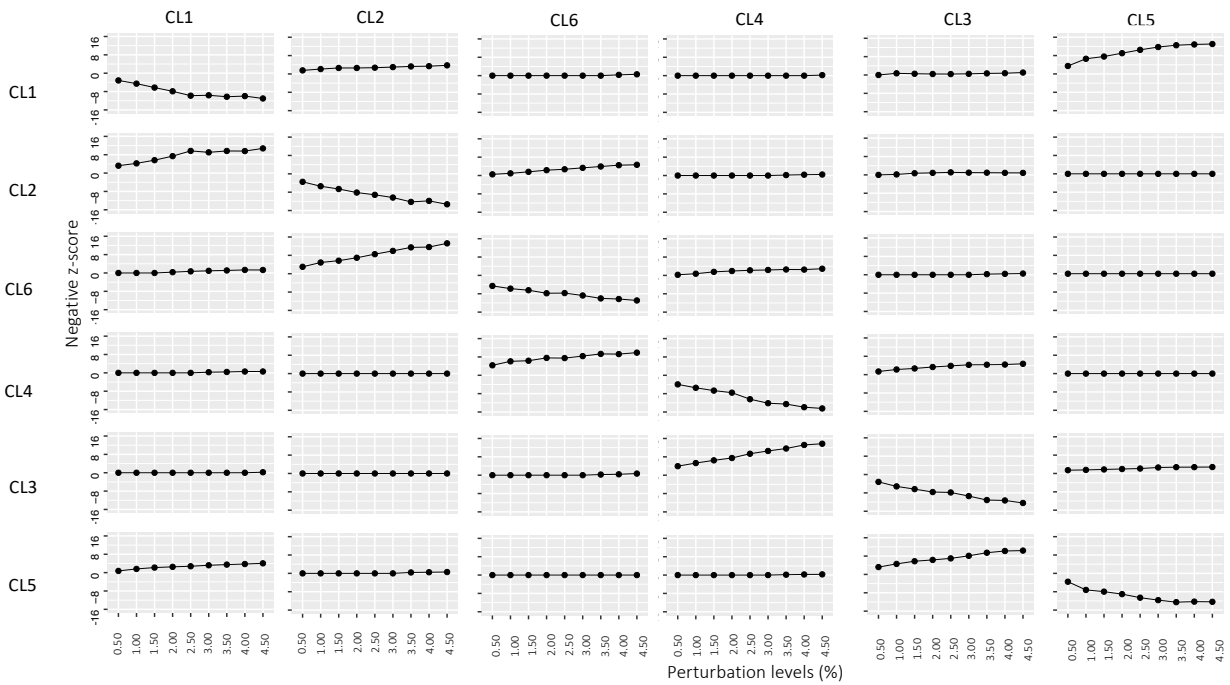


Figure S24. Repressilator - distribution of the negative z-scores of the percent models in each cluster after each perturbation level, related to Figure 6C (see also Figure S23). These negative z-scores depict the proportion of models in a cluster leaving that cluster after systematically perturbating parameters at a certain level.

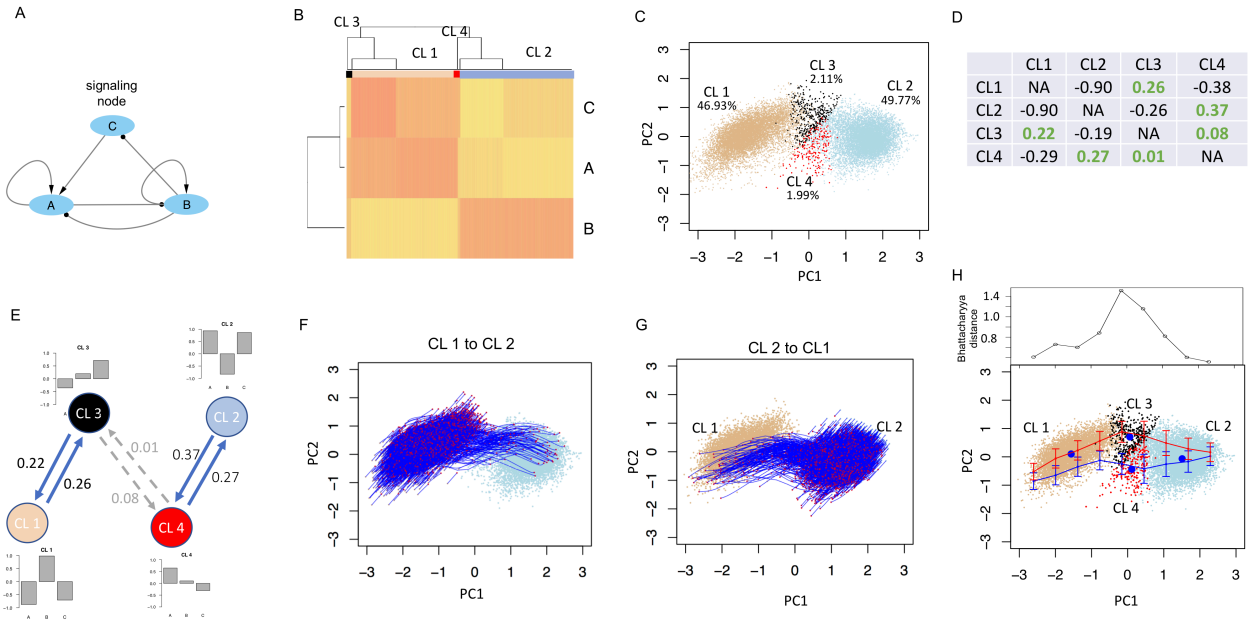


Figure S25. Modeling a toggle switch circuit with an input node, related to Figure 6D**. (A)** Illustration of the circuit topology. **(B)** Four distinct clusters are identified from the hierarchical clustering analysis (r function hlcust with method: complete and dist: 1-Pearson cor) of the stable steady states of 10,000 RACIPE models. Colors for the clusters – CL1: beige, CL2: light blue, CL3: black, CL4: red. **(C)** The steady state gene expression profiles, colored by the identity of clusters, projected onto the first two principal components. There are two major clusters (CL1 with 46.93% states, CL2 with 49.77% states) and two minor clusters (CL3 with 2.11% states and CL4 with 1.99% states). **(D)** Medians of the forward and backward delayed correlations for any two clusters. **(E)** A model of state transitions between these four clusters from the delayed correlations. We predict the major transitions to occur between cluster pairs 1 and 3 and between cluster pairs 2 and 4. The bar plots show the average expression of the genes for different clusters. **(F)** Further RACIPE simulation of the bistable models (with steady states from both CL1 and CL2) support this state transition model. When we simulated these models with CL1 states as the initial conditions and the increased production rate for the input gene C, we found the trajectories moved towards CL3 before they reached CL2. **(G)** On the other hand, simulating the models with CL2 states as the initial conditions and the decreased production rate for gene C, we found the trajectories to move towards CL4 before they reach CL1. **(H) (bottom)** Mean and standard deviation of the distribution of the binned PC2 coordinates along PC1 for the forward and backward trajectories, shown as mean ± SD. (**top**) Bhattacharyya distance between the forward and backward trajectories becomes maximum as they pass through CL3 and CL4.

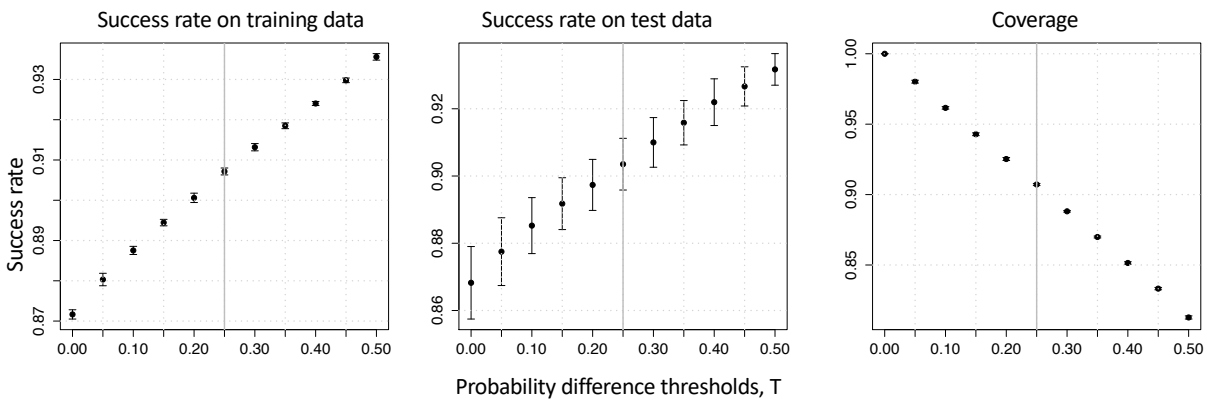


Figure S26. Performance of the neural net model in assigning cell cycle phases from simulated activity profiles, related to Figure 2**.** Along an oscillatory trajectory, the assignment of cell cycle phases was only applied to the gene activity profiles when the differences between the first two highest probabilities are larger than a threshold value T. Here, a 10-fold cross-validation was performed for different probability difference thresholds T. Leftmost and middle panels show the success rate for the traning and testing data. The rightmost panel shows the coverage, defined as the percentage of gene activity profiles used for the assignment of cell cycle phases due to the filtering step. On each panel, data are shown as mean ± SD.

Supplemental Tables

Table S1. The assignment of the levels of gene activity in different clusters - H: high level, M: medium level, L: low level, related to Figure 1C.

	Expression levels from RACIPE					
	C1	C5	C4	C2	C3	C6
Cln3	L	H	L	M	H	L
MBF	M	H	L	L	M	M
Whi5	L	L	M	H	L	H
SBF	H	H	L	L	H	L
Clb5,6	H	M	L	H	H	L
Cln1,2	H	H	L	L	H	L
Sic1	L	H	M	L	L	H
Cdh1	L	H	H	L	L	H
Clb1,2	L	M	L	H	H	L
DNAS	L	L	L	H	H	L
Mcm1,SFF	H	H	L	H	H	L
Cdc20	L	L	L	H	H	L
Swi5	L	H	L	L	M	H
Pds1	H	M	H	L	L	L
Cdc14	L	H	L	M	H	H

Table S2. Activity levels of cell cycle genes at different cell cycle phases from experiments, related to Figures 1B,1C.

#	Gene name	G1	S	G2	M	M/G1
1	Cln3	H Fig.1. (Spellman et al., 1998)	H Fig.1. (Spellman et al., 1998)	H→L Fig.1. (Spellman et al., 1998)	L - Fig.1. (Spellman et al., 1998) H - Fig.7. (Spellman et al., 1998)	L Fig.1. (Spellman et al., 1998)
2	MBF	H Tab. 4. (Spellman et al., 1998)				
3	Whi5					
4	SBF	H Tab. 4. (Spellman et al., 1998)				
5	Clb5	H Fig.3&7. (Spellman et al., 1998)	M Fig. 3. (Spellman et al., 1998)	M Fig. 3. (Spellman et al., 1998)	M Fig. 3. (Spellman et al., 1998)	L → H Fig. 3. (Spellman et al., 1998)
	Clb6	H Fig. 3. (Spellman et al., 1998)	H-(Jackson et al., 2006) M - Fig. 3. (Spellman et al., 1998)	L Fig. 3. (Spellman et al., 1998)	L Fig. 3. (Spellman et al., 1998)	L → H Fig. 3. (Spellman et al., 1998)
	Clb5,6					
6	Cln1	H Fig. 7. (Spellman et al., 1998)				
	Cln2	H Fig. 3&7. (Spellman et al., 1998)	H→M Fig. 3&7. (Spellman et al., 1998)	M Fig. 3&7. (Spellman et al., 1998)	L Fig. 3&7. (Spellman et al., 1998)	L Fig. 3&7. (Spellman et al., 1998)
	Cln1,2					
7	Sic1		L (Jackson et al., 2006)			H Fig. 7. (Spellman et al., 1998)
8	Cdh1					
9	Clb1				H Fig. 7. (Spellman et al., 1998)	
	Clb2	L Fig. 3. (Spellman et al., 1998)	L Fig. 3. (Spellman et al., 1998)	L→H Fig. 3. (Spellman et al., 1998)	H Fig. 3&7. (Spellman et al., 1998)	H→L Fig. 3. (Spellman et al., 1998)
	Clb1,2					
10	DNAS					
	Mcm1				H (Ferrezuelo et al., 2010)	H (Ferrezuelo et al., 2010)
11	SFF					
	Mcm1, SFF					
12	Cdc20				H Fig. 7. (Spellman et al., 1998)	
13	Swi5		H			H (Ferrezuelo et al., 2010)
14	Pds1	H Fig. 7. (Spellman et al., 1998)				
15	Cdc14					

Table S3. Node names of the cell cycle circuit (shown in Figure 1A) marked according to the active role they play in specific cell cycle phases and the related genes for extracting expression from single cell RNAseq data, related to Figures 1A,1E.

Cell cycle phase	Node name	Gene used from single cell RNA seq data
G1	Cln3	Cln3
G1	Whi5	Whi5; We also examined Rnr1, which is regulated by SBF and MBF(Costanzo et al., 2004); and its homology, Rnr3. However, both genes and Whi5 have zero expressions in most cells.
G1	SBF	Swi4, which is a component of SBF (SBF = Swi4+Swi6)
G1	MBF	RFA, which is target of MBF complex(Ferrezuelo et al., 2010)
G1/S	Cln1,2	Cln1
G1/S	Clb5,6	Cln6
S	DNAS	HTA1 expression from data is used for hypothetical node DNA Synthesis (DNAS). Because, HTA1 expression peaks during S phase (Figure 7(Spellman et al., 1998)).
M	Clb1,2	Clb1
M	Cdh1	Cdh1
M	Mcm1,SFF	Mcm1
M/G1 (MEN)	Sic1	Sic1
M/G1 (MEN)	Pds1	Pds1
M/G1 (MEN)	Cdc20	Cdc20
M/G1 (MEN)	Swi5	Swi5
M/G1 (MEN)	Cdc14	Cdc15 is used instead of Cdc14 as Cdc14 activates Cdc15 to promote mitotic exit(Jaspersen and Morgan, 2000) and Cdc15 expression is found to be more detectable in the data.

Table S4. List of knockdowns (KD) that generate significant change in phase stability, related to Figure 2. This table shows an interpretation of KD in Figure 2 and Figure S9.

Phase name	Perturbation changes stability	
	increases	**decreases**
early G1	$Cln3^{KD}MBF^{KD}$ $Cln1^{KD}Clb5^{KD}$	$Cdc14^{KD}$
mid G1	$Clb5^{KD}$ $Pds1^{KD}$	$Cln3^{KD}Cln1^{KD}$
late G1	$Clb1^{KD}$ $DNAS^{KD}$ $Mcm1^{KD}$	
S	$Whi5^{KD}$ $Sic1^{KD}$ $Cdh1^{KD}Sic1^{KD}$	
M	$Cln1^{KD}Sic1^{KD}$	$Clb1^{KD}$ $DNAS^{KD}$
M/G1	$Cln1^{KD}$ $Cln3^{KD}Cln1^{KD}$ $Cln1^{KD}Clb5^{KD}$	$Cdc20^{KD}Pds1^{KD}$

Table S5. Single and double mutants with the observed phenotypes, related to Figure 2.

Mutants	Phenotype	References
Cln3	Extended G1 phase	(Dirick et al., 1995; Hartwell, 1973; Stuart and Wittenberg, 1995; Tyers et al., 1993; Wijnen et al., 2002)
	Large cell size	(Dirick et al., 1995) – Fig 3
Whi5	Smaller than wild type	(de Bruin et al., 2004; Jorgensen et al., 2002)
SBF	Viable	(Futcher, 2002; Nasmyth, 1996)
MBF	Viable	
Cln1		(Dirick et al., 1995) – Fig 3
Sic1		Short G1 - (Schneider, 1996)
Clb5		
Cdh1		Grows slowly but viable – (Jorgensen et al., 2002; Wäsch and Cross, 2002)
Clb1 DNAS Mcm1		
Cdc20		Metaphase arrested – (Lim et al., 1998; Shirayama et al., 1998)
Pds1 Cdc14 Swi5		
Cln3, Whi5	Smaller than wild type and as small as ΔWhi5	(Costanzo et al., 2004)
Cln3, SBF Cln3, MBF		
Cln3, Sic1	G1 short, size 1.5 x WT, smaller than ΔCln1	Table 3 (Chen et al., 2000)
SBF, MBF	Inviable. Cell cycle arrest before START	(Koch et al., 1993)
Cln3, Cln1	Arrest in G1	(Wittenberg et al., 1990)
Cln1, Sic1	Smaller cell size	Fig 4 – Size between ΔCln1 and WT (Dirick et al., 1995)
Cln1, Clb5	G1 arrest	Fig. 5 (Schwob and Nasmyth, 1993)
Cdh1, Sic1 Swi5, Cdh1		
Cdc20, Clb5	Metaphase arrested	(Shirayama et al., 1999)
Cdc20, Pds1	Telophase arrested	(Shirayama et al., 1999)

Table S6. Mapping between yeast and human cell cycle genes, related to Figure 1A.

Budding Yeast	**Mammalian Cells**
Cln3/Cdc28	CycD/Cdk4,6
Cln1,2/Cdc28	CycE/Cdk2
Clb5,6/Cdc28	CycA/Cdk1,2
Clb1,2/Cdc28	CycB/Cdc2
Sic1	$p27^{Kip1}$
Cdc20	$p55^{Cdc}$
Cdc14	hCdc14
Mcm1	Mcm

Table S7. Distinct state transition patterns (a total of 47) of the identified oscillatory trajectories from 160,000 RACIPE models, related to Figure 4C. Numeric code used for cell cycle phases: 1 – early G1, 2 – mid G1, 3 – late G1, 4 – S, 5 – M, 6 – M/G1.

Patterns	Counts	Patterns	Counts
1351	1793	26542	7
2642	621	3653	7
2532	509	24652	5
24532	230	26432	4
13251	133	4564	4
132451	114	4654	4
2542	97	1324651	3
13241	72	1321	3
2432	72	2632	3
26532	48	35643	2
2452	47	2462	2
3543	27	2562	2
2652	17	2352	2
1341	17	1326451	1
1451	14	143251	1
132651	13	124651	1
246532	11	134651	1
12451	11	265432	1
3453	10	256432	1
13451	9	13541	1
25432	8	14351	1
1251	8	14251	1
1241	8	3563	1
25642	7		

Supplemental References

Albergante, L., Mirkes, E., Bac, J., Chen, H., Martin, A., Faure, L., Barillot, E., Pinello, L., Gorban, A., and Zinovyev, A. (2020). Robust and scalable learning of complex intrinsic dataset geometry via ElPiGraph. Entropy 22, 296.

Chen, K.C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., Tyson, J.J., and Solomon, M.J. (2000). Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol. Biol. Cell 11, 369–391.

Costanzo, M., Nishikawa, J.L., Tang, X., Millman, J.S., Schub, O., Breitkreuz, K., Dewar, D., Rupes, I., Andrews, B., and Tyers, M. (2004). CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. Cell 117, 899–913.

de Bruin, R.A.M., McDonald, W.H., Kalashnikova, T.I., Yates III, J., and Wittenberg, C. (2004). Cln3 Activates G1-Specific Transcription via Phosphorylation of the SBF Bound Repressor Whi5. Cell 117, 887–898.

Dirick, L., Böhm, T., and Nasmyth, K. (1995). Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of Saccharomyces cerevisiae. EMBO J. 14, 4803–4813.

Duda, R., Hart, P., and Stork, D. (2000). Pattern Classification, 2nd Edition | Wiley, 2nd ed. Wiley.

Fasy, B.T., Kim, J., Lecci, F., and Maria, C. (2015). Introduction to the R package TDA. ArXiv: 1411.1830 Cs.MS.

Ferrezuelo, F., Colomina, N., Futcher, B., and Aldea, M. (2010). The transcriptional network activated by Cln3 cyclin at the G1-to-S transition of the yeast cell cycle. Genome Biol. 11, R67.

Futcher, B. (2002). Transcriptional regulatory networks and the yeast cell cycle. Curr. Opin. Cell Biol. 14, 676–683.

Hartwell, L.H. (1973). Three Additional Genes Required for Deoxyribonucleic Acid Synthesis in Saccharomyces cerevisiae. J. Bacteriol. 115, 966–974.

Huang, B., Jia, D., Feng, J., Levine, H., Onuchic, J.N., and Lu, M. (2018). RACIPE: a computational tool for modeling gene regulatory circuits using randomization. BMC Syst. Biol. 12, 74.

Huang, B., Lu, M., Jia, D., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2017). Interrogating the topological robustness of gene regulatory circuits by randomization. PLOS Comput. Biol. 13, e1005456.

Huang, C.Y. and Ferrell, J.E. (1996). Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc. Natl. Acad. Sci. 93, 10078–10083.

Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J.I., Raj, A., Li, M., and Zhang, N.R. (2018). SAVER: gene expression recovery for single-cell RNA sequencing. Nat. Methods 15, 539–542.

Jackson, C.A., Castro, D.M., Saldi, G.-A., Bonneau, R., and Gresham, D. (2020). Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. eLife 9.

Jackson, L.P., Reed, S.I., and Haase, S.B. (2006). Distinct mechanisms control the stability of the related S-phase cyclins Clb5 and Clb6. Mol. Cell. Biol. 26, 2456–2466.

Jaspersen, S.L. and Morgan, D.O. (2000). Cdc14 activates Cdc15 to promote mitotic exit in budding yeast. Curr. Biol. 10, 615–618.

Jorgensen, P., Nishikawa, J.L., Breitkreutz, B.-J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. Science 297, 395–400.

Koch, C., Moll, T., Neuberg, M., Ahorn, H., and Nasmyth, K. (1993). A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. Science 261, 1551–1557.

Kraeutler, M.J., Soltis, A.R., and Saucerman, J.J. (2010). Modeling cardiac β-adrenergic signaling with normalized-Hill differential equations: comparison with a biochemical model. BMC Syst. Biol. 4, 157.

Li, W.V., Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat. Commun. 9, 1–9.

Lim, H.H., Goh, P.-Y., and Surana, U. (1998). Cdc20 is essential for the cyclosome-mediated proteolysis of both Pds1 and Clb2 during M phase in budding yeast. Curr. Biol. 8, 231–237.

Liu, Z., Lou, H., Xie, K., Wang, H., Chen, N., Aparicio, O.M., Zhang, M.Q., Jiang, R., and Chen, T. (2017). Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat. Commun. 8, 22.

Lu, M., Onuchic, J., and Ben-Jacob, E. (2014). Construction of an effective landscape for multistate genetic switches. Phys. Rev. Lett. 113, 078102.

Macnair, W. and Claassen, M. (2019). psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels. bioRxiv: 10.1101/622001.

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics btw777.

Nasmyth, K. (1996). At the heart of the budding yeast cell cycle. Trends Genet. 12, 405–412.
Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. 37, 547.
Schneider, B.L. (1996). Linkage of Replication to Start by the Cdk inhibitor Sic1 272, 4.
Schwob, E. and Nasmyth, K. (1993). CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in Saccharomyces cerevisiae. Genes Dev. 7, 1160–1175.
Shirayama, M., Tóth, A., Gálová, M., and Nasmyth, K. (1999). APCCdc20 promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. Nature 402, 203.
Shirayama, M., Zachariae, W., Ciosk, R., and Nasmyth, K. (1998). The Polo-like kinase Cdc5p and the WD-repeat protein Cdc20p/fizzy are regulators and substrates of the anaphase promoting complex in Saccharomyces cerevisiae. EMBO J. 17, 1336–1349.
Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell 9, 3273–3297.
Steinway, S.N., Zañudo, J.G.T., Ding, W., Rountree, C.B., Feith, D.J., Loughran, T.P., and Albert, R. (2014). Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic Hedgehog and Wnt pathway activation. Cancer Res. 74, 5963–5977.
Stuart, D. and Wittenberg, C. (1995). CLN3, not positive feedback, determines the timing of CLN2 transcription in cycling cells. Genes Dev. 9, 2780–2794.
Tracy, S., Yuan, G.-C., and Dries, R. (2019). RESCUE: imputing dropout events in single-cell RNA-sequencing data. BMC Bioinformatics 20, 388.
Tyers, M., Tokiwa, G., and Futcher, B. (1993). Comparison of the Saccharomyces cerevisiae G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. EMBO J. 12, 1955–1968.
van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., Bierie, B., Mazutis, L., Wolf, G., Krishnaswamy, S., and Pe'er, D. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell 174, 716-729.e27.
Venables, W.N. and Ripley, B.D. (2002). Modern Applied Statistics with S, 4th ed. Springer, New York, NY, USA.
Wagner, F., Yan, Y., Yanai, I. (2018). K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. bioRxiv: 10.1101/217737
Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Res. 47, e66–e66.
Wäsch, R. and Cross, F.R. (2002). APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit. Nature 418, 556–562.
Wijnen, H., Landman, A., and Futcher, B. (2002). The G1 cyclin Cln3 promotes cell cycle entry via the transcription factor Swi6. Mol. Cell. Biol. 22, 4402–4418.
Wittenberg, C., Sugimoto, K., and Reed, S.I. (1990). G1-specific cyclins of S. cerevisiae: Cell cycle periodicity, regulation by mating pheromone, and association with the p34CDC28 protein kinase. Cell 62, 225–237.