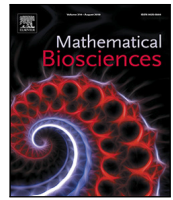




Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Original Research Article

A yeast cell cycle pulse generator model shows consistency with multiple oscillatory and checkpoint mutant datasets

Julian Fox^a, Breschine Cummins^{a,*}, Robert C. Moseley^b, Marcio Gameiro^c, Steven B. Haase^b^a Department of Mathematical Sciences, Montana State University, Bozeman, MT, USA^b Department of Biology, Duke University, Durham, NC, USA^c Department of Mathematics, Rutgers University, New Brunswick, NJ, USA

ARTICLE INFO

Dataset link: <https://github.com/julianfox8/2022-DSGRN-Phenotypes-Yeast.git>

Keywords:

Dynamical systems
Multi-level Boolean model
Identifiability
Gene regulatory networks
Cell cycle
Transcription

ABSTRACT

Modeling biological systems holds great promise for speeding up the rate of discovery in systems biology by predicting experimental outcomes and suggesting targeted interventions. However, this process is dogged by an identifiability issue, in which network models and their parameters are not sufficiently constrained by coarse and noisy data to ensure unique solutions. In this work, we evaluated the capability of a simplified yeast cell-cycle network model to reproduce multiple observed transcriptomic behaviors under genomic mutations. We matched time-series data from both cycling and checkpoint arrested cells to model predictions using an asynchronous multi-level Boolean approach. We showed that this single network model, despite its simplicity, is capable of exhibiting dynamical behavior similar to the datasets in most cases, and we demonstrated the drop in severity of the identifiability issue that results from matching multiple datasets.

1. Introduction

The promise of modeling complex molecular systems is that the rate of discovery of important mechanisms can increase through targeted experimentation informed by model predictions. However, a large barrier to the fulfillment of this promise is an unavoidable identifiability problem. That is, multiple modeling frameworks are capable of describing (generally coarse and noisy) data to a sufficient degree that the underlying system mechanisms cannot be distinguished computationally. Moreover, even if the “true” model could be identified, the high dimensionality of parameter space and sparse data generally permit many acceptable solutions to the data fitting problem, i.e., practical non-identifiability [1]. The uncertainty of system parameters means that predicted responses to perturbations have high uncertainty as well, slowing down the iterative exchange between model and experiment through loss of accuracy.

However, there is another way to view non-identifiability, and that is as a measure of robustness of a network model with regard to a particular dynamical behavior. In other words, if a model is perturbed to a different parameterization and yet maintains the observed dynamical behavior, the model is robust to that particular perturbation. A quantification of non-disruptive parameter perturbations could then serve as a measure of robustness of the network model with respect to a dynamical behavior. We adopt this perspective in this manuscript.

It would be tempting to optimize network models on this criterion of robustness and choose the highest-ranking networks as the most plausible models of a biological system. However, evolution involves a multi-objective optimization in which different objectives can be in conflict. Particularly relevant to our discussion in this paper is the trade-off between robustness and controllability. Genetic and/or protein networks that control various cellular processes may need to have different dynamical behaviors under different environmental conditions; i.e., the network’s dynamical behavior can be controlled by external factors. Therefore, it is not clear that the most robust network models are the best hypotheses for all systems, given that some level of controllability is often desirable. We claim that the expression of different dynamical behaviors induced by the controllability of a genetic network permits a meaningful reduction in the parameter space of a network model.

In this manuscript, we introduce a computational methodology via an example that leverages observations of distinct dynamical behaviors to reduce parameter space. We model a small yeast cell-cycle regulatory network along with a single control point, and match model predictions to multiple experimental datasets that represent different dynamical behaviors under different perturbations with direct or indirect effects on the activity of the control point. The example demonstrates the extreme reduction in parameter space that can be achieved by incorporating the results of a broad array of genetic perturbations at a control point.

* Correspondence to: Department of Mathematical Sciences, Montana State University, P.O. Box 172400, Bozeman, MT 59717-2400, USA.

E-mail address: breschine.cummins@montana.edu (B. Cummins).<https://doi.org/10.1016/j.mbs.2023.109102>

Received 6 May 2023; Received in revised form 13 September 2023; Accepted 27 October 2023

Available online 7 November 2023

0025-5564/© 2023 Published by Elsevier Inc.

Models for the mechanisms controlling cell-cycle progression have evolved over time. Early biochemical studies in marine invertebrates identified cyclins and cyclin dependent kinases (CDKs) as key regulators of cell-cycle oscillations along with the anaphase promoting complex (APC) (a ubiquitin ligase) [2–4]. By forming a negative feedback loop, it was hypothesized that this simple cell-cycle network was capable of producing periodic behavior [5,6]. Genetic studies in both *Saccharomyces cerevisiae* and *S. pombe* also identified cyclins, CDKs, and APC complexes as critical regulators of cell-cycle progression, indicating that machinery was highly conserved across phyla and that these components are the basic components of the cell-cycle oscillator [7–10]. Cell-cycle models became more complex over time and mounting evidence indicated that oscillations of cell cycle events could continue when the CDK/APC motif was prevented from oscillating [11]. The advent of genome technologies identified large programs of dynamic gene expression associated with the cell cycle, and new models emerged suggesting the importance of transcriptional networks on producing cell-cycle oscillations [12–15].

Although ODE models of the yeast cell-cycle that are focused primarily on biochemical interactions have been remarkably predictive of mutant phenotypes [16], there is a compelling argument that transcription factors (TFs), cyclin-dependent kinases (CDKs), and ubiquitin ligases all play key roles in regulating cell-cycle progression [17]. Multiple studies have demonstrated that temporally ordered, high-amplitude transcript dynamics were present in budding yeast with non-oscillating levels of CDK [13,17–20] indicating that the CDK/APC oscillator identified in early embryonic systems may not be the core motif driving periodic behavior during the yeast cell cycle.

In 2016, the Cross group disputed the claim that the mitotic CDK/APC may be dispensable for cell-cycle oscillations in yeast, suggesting that in cyclin mutant strains in previous experiments incompletely eliminated mitotic cyclin (Clb1) and that undetectable Clb1 oscillations were driving the observed oscillations [21]. A thorough rebuttal of claims from Rahi et al. [21] were detailed by Cho et al. [17] and will not be reproduced here. However, in a strain where the authors eliminated all mitotic cyclins (Clb2 in their strain) as well as G1 cyclins, a strikingly similar transcriptional wave moving from G1 to mitosis was observed, indicating that transcriptional dynamics can progress through the ordered phases of the cell-cycle without periodic input from CDK [21]. Thus, a substantial portion of cell-cycle transcriptional dynamics can be uncoupled from cell-cycle progression and that dynamical behavior is worth modeling.

The ability to uncouple cell-cycle progression from transcriptional progression suggested the existence of a mechanism(s) to prevent this decoupling during physiological arrests. In 2014 we demonstrated that the DNA replication checkpoint and the spindle assembly checkpoint both act directly on the transcription machinery to halt transcriptional dynamics [18]. Here we explicitly model checkpoint activation and ask whether the model output fits the data.

Regardless of this conflict, much of the cell-cycle modeling efforts have focused on biochemical interactions of cyclins, CDKs, APC, and their regulators without explicit regard for the periodic synthesis of these components (e.g. [16]). This convention likely stems from the initial biochemical assessments of oscillations and the cell-cycle oscillator in early embryonic systems where there is very little transcription, and synthesis of cell-cycle regulators comes from large maternal stores of mRNA. This framework was extrapolated to somatic cells and yeast despite the substantial cell-cycle remodeling as cells transition from very rapid embryonic divisions driven by maternal stores of mRNAs to substantially slower divisions driven by periodic transcription of cell-cycle regulators [22].

More recent work interrogated the transcriptional network model by mutating network components and assessing system-wide phenotypes [19]. The results suggest that a pulse-generating TF network containing an oscillatory mechanism was responsible for a transcriptional pulse that was thought to drive global phase-specific transcription. The pulse generator seems to operate in a start-stop manner,

where the network is first quiescent and then, after receiving a start signal, creates a wave of sequential transcription through the network which is hypothesized to be driven by the interaction of CDKs with a TF network [13,18,19]. Reactivation of the pulse generating motif is actively inhibited by the expression of transcriptional repressors until the end of the cell cycle when they are actively degraded by the APC [19]. The transcriptional pulse driving progression through the cell cycle operates consistently, meaning that the gene products express in a stereotyped order [19,23], and the timing and robustness of this periodic transcription was affected when G1-CDK activities were depleted, knocked out, or up-regulated, creating a weaker and less robust cycle [20]. It is important to note that periodic transcription, while weaker and less robust, was still present in some of the mutants. Therefore, the dynamics expressed by any hypothesized network model should exhibit oscillations under both wild-type (WT) and CDK mutant phenotypes.

Synchronous and autonomous Boolean models of yeast cell-cycle networks containing TFs and CDK have produced robust oscillatory behavior [13,24], but more sophisticated dynamical models that match observed dynamics (including transcriptional dynamics) across wild-type and mutant strains have not been reported. Moreover, these models have not addressed the observation that certain environmental perturbations (e.g. DNA damage or spindle assembly defects) can reversibly arrest the cell cycle until damage is repaired.

To jointly address these modeling gaps, we employ a type of asynchronous multi-level Boolean approach [25] called Dynamic Signatures Generated by Regulatory Networks (DSGRN) [26–28]. The DSGRN software package can exhaustively compute all of the qualitative dynamics a genetic regulatory network (GRN) can produce, allowing for a comprehensive description of potential network behaviors across high-dimensional parameter space. DSGRN is organism-agnostic and systematic, precluding the need for deep biological expertise. It has been used to model and predict genetic network behavior in biological systems similar to the yeast cell cycle, such as the epithelial-to-mesenchymal transition in cancer [29] and the Rb-E2F mechanism of the mammalian cell cycle [30]. Because of its comprehensive description of network dynamics, DSGRN is tailored to the discovery of small regions of parameter space where multiple dynamical behaviors are predicted to occur. These predictions can be matched to experimental observations, focusing attention on regions of parameter space that are hypothesized to be biologically relevant, and that can be quite small in relation to the entirety of parameter space. DSGRN overcomes the possibility that such regions will be missed, as is likely when using the sampling techniques employed in the traditional ODE modeling paradigm.

Identifying promising parameter regions using DSGRN permits the construction of parameterized Hill models from parameter sampling restricted to this (much) smaller region of parameter space. The translation of Boolean (or multi-level Boolean) models into systems of ODEs is well studied. It is known that every monotone Boolean function model of a genetic regulatory network can be written as a K-system [25], which is a collection of piecewise constant ODEs [31]. Special cases of K-systems can be written as switching ODE systems with algebraic constraints (i.e., DSGRN) [25,27], which themselves are approximations of Hill ODE models [32]. Furthermore, there is a recent simulation study [33] comparing DSGRN output to the Hill model approach of RACIPE [34]. Outside of our group, there are other approaches to building continuous models from Boolean systems, e.g. [35–37], as well as simulation studies comparing Boolean and continuous models [38].

DSGRN model validation and robustness quantification uses only coarse information from transcriptomics time series. In particular, model predictions are not required to exactly match observed trajectories, but rather large features of time series data are conserved across parameters, particularly transcriptional oscillatory behavior and checkpoint behavior. We remark that the robustness of checkpoints and the G1 cell cycle phase was previously explored in [39] for a yeast cell-cycle network model without the pulse generation module from [19]

that is hypothesized to generate transcriptional oscillations. The definition of robustness in [39] is different than in the treatment here, mostly concentrating on the size of basins of attraction of cell-cycle states, rather than on non-local parameter non-identifiability.

In Section 2, we describe the network model we explore, experimental observations of seven datasets, and the desired model output in a DSGRN frame of reference. In Section 3, we report the dramatic reduction in parameter space achieved by matching network model predictions to the seven datasets. In Section 4, we discuss the significance of our findings, and in Section 5, we present detailed methods.

2. Network modeling approach

In this section, we discuss the evidence-based network model that we check for consistency with multiple wild-type and mutant datasets. Each of these datasets exhibits a *cellular phenotype*, namely, the cell cycle is either progressing or arrested. This cellular phenotype is sometimes uncoupled from the *transcriptional phenotype*, that in some mutants can be progressing while the cellular phenotype is arrested [11, 13,18–20]. We distinguish between these cellular phenotypes and transcriptional phenotypes based on wild type and mutant microarray and RNAseq time series datasets from [13,17–19]. Transcriptional phenotypes consist of observed cycling or steady equilibrium behavior seen in the time series. We also define *dynamical phenotypes* based on DSGRN predictions of network model dynamics that allow us to determine model consistency with the data. Dynamical phenotypes are graphs produced by DSGRN that describe stable or unstable cycling behavior as well as equilibria. In order to be declared consistent with the observed data, a network model must be able to reproduce cyclic patterns in WT and mutant data and needs to support the arrest of cycling behavior during a triggered checkpoint.

2.1. Cell-cycle network models

Most mathematical models of the cell-cycle have aimed to explain cellular phenotypes related to the ordered progression of cell-cycle events such as bud emergence, DNA replication and mitosis. The goal of our modeling exercise was to address the cell-cycle transcriptional phenotypes of wild-type and mutant cells that have been largely unexplored by previous dynamical cell-cycle models. Thus, our choice of network models must accommodate these transcriptional phenotypes by explicitly modeling control of cell-cycle transcription. Gene Regulatory Network (GRN) models must represent relevant regulatory connections between transcription factors and other transcription factors as well as transcription factors and other cell-cycle regulators (e.g. cyclin genes). Although cell-cycle network models such as that described by Chen and colleagues include some of the known cell-cycle machinery, they lack many of the well-established transcriptional connections and thus are unable to capture the transcriptional behaviors we seek to explain by modeling [16]. This is also true of the textbook models of the yeast cell cycle [40]. In Fig. 1 (Left) as well as in previous publications, we have attempted to capture both relevant biochemical and transcriptional relationships in the cell-cycle network. The network simplification depicted in Fig. 1 (Right) aims to capture the relevant transcriptional connections that are appropriate for describing transcriptional phenotypes.

A network oscillator model [17] representing a collection of regulatory interactions hypothesized to be capable of exhibiting multiple cell-cycle behaviors is visualized as a cell-cycle GRN in Fig. 1 (Left). The edges in the cell-cycle GRN reflect different regulatory mechanisms. A TF regulates another TF or CDK through transcriptional control of gene products. A CDK regulates another CDK or TF only when bound in a cyclin/CDK complex (post-transcriptional control). Once assembled, the complex is able to phosphorylate the target protein, which can have either an activating or inhibiting effect depending on the target. In Fig. 1, black arrows indicate activation, red arrows

Table 1

Proxy choices for the simplified cell-cycle network.

SCC network proxy sets					
Proxies	SCC nodes				
	SBF	Nrm1/Yox1	SFF	Clb2	Swi5
Swi5-Nrm1	Swi4	Nrm1	Ndd1	Clb2	Swi5
Swi5-Yox1	Swi4	Yox1	Ndd1	Clb2	Swi5
Ace2-Nrm1	Swi4	Nrm1	Ndd1	Clb2	Ace2
Ace2-Yox1	Swi4	Yox1	Ndd1	Clb2	Ace2

indicate inhibition, solid arrows indicate transcriptional control, and dashed arrows post-transcriptional control.

While the CDKs and TFs identified in Fig. 1 (Left) are important for progression of the cell cycle, some research has focused on investigating the importance of smaller subnetworks [13,18,19]; for example, concentrating on the impact that Clb2 activity has on its targets and the progression of the transcriptional program. We continue along these lines by computationally investigating a small network derived from the cell-cycle GRN. We will refer to this network as the simplified cell-cycle GRN, or SCC network, seen in Fig. 1 (Right). We remark that although post-transcriptional interactions are modeled, the differences between transcriptional and post-transcriptional regulation are not modeled; however, see the recent advance in DSGRN [41] that incorporates the processes of ubiquitination and phosphorylation. We instead model under the assumption that the activity of Clb2 is well-correlated with its abundance, and we use transcript abundance data as a proxy for protein abundance. These data are known to be well-correlated [42].

Not all nodes in the cell-cycle GRN are associated to a unique TF or cyclin/CDK; in some cases the node refers to a complex of TFs or cyclins and CDKs. (By convention, the name of a complex of proteins has all letters capitalized, the name of an individual protein has only the first letter capitalized.) As is clear from the presence of these complexes and boxed TFs in Fig. 1, multiple choices are available for data representing each node for the SCC network. We evaluated model consistency with the data for each of the proxy sets listed in Table 1.

The SCC network model was partially chosen for computational tractability and many nodes and edges from the cell-cycle GRN are absorbed or disregarded. The choices we made have some dynamical justification based on network topology. The self-edge on SBF comes from the length three path from MBF/SBF to itself through Cln1/2 and Whi5 in Fig. 1 (Left). Since there is a double repression, this indirect self-regulation is activating. Likewise, the edge from Swi5 to SBF in the SCC network is a collapsed path through double repression in Cln3 and Whi5. The Hcm1 node in the linear transmission array is absorbed into SFF, since it has no other dynamical significance. The APC-Cdc20 off switch and Sic1 and APC-Cdh1 regulation are absorbed into the decay and activity parameters for Clb2. The removal of Cdc14 is justified since its impact on Swi5/Ace2 can be viewed as an indirect effect of SFF, which already has a parallel direct edge to Swi5/Ace2.

Lastly, the subnetwork of SCC excluding Clb2 that consists only of the four nodes SBF, SFF, Swi5/Ace2, and Nrm1/Yox1 is important in its own right for parameter reduction. We will refer to it as the transcription factor oscillator subnetwork, or TFO subnetwork.

2.2. Transcriptional phenotypes

We describe seven different transcriptional phenotypes expressed in seven datasets. The transcriptional phenotypes are called WT, Clb2 OFF, Clb2 ON, Clb2 intermediate-low (INT-L), Clb2 intermediate-high (INT-H), spindle assembly checkpoint (SAC), and DNA replication checkpoint (DRC). Their relationships to datasets and cellular phenotypes is shown in Table 2.

The wild-type (WT) microarray dataset comes from [13] where genome-wide transcription was analyzed for *S. cerevisiae*. The WT

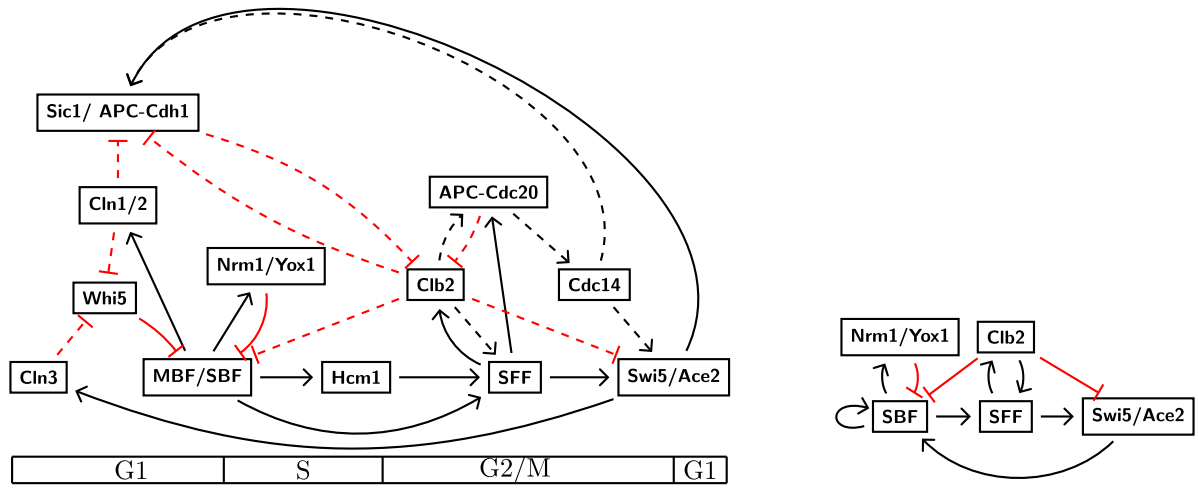


Fig. 1. (Left) A proposed cell-cycle GRN, or network oscillator, corresponding to Fig 7 A in [17]. The timing of the activity of the network nodes is given by the linear arrangement of cell cycle phases (G1, S, G2/M) at the base of the figure. (Right) The simplified cell-cycle GRN (SCC network) that attempts to capture important features of multiple datasets. We call the subnetwork without Clb2 the TF oscillator subnetwork (TFO subnetwork), consisting of nodes SBF, SFF, Swi5/Ace2, and Nrm1/Yox1. Black arrows indicate activation, red blunt arrows indicate repression, dashed lines represent post-transcriptional interactions and solid lines represent transcriptional interactions. Notice that dashed lines are not included on the right, because we do not explicitly model post-transcriptional modification in this paper.

Table 2

The mapping between dynamical and transcriptional phenotypes and their associated datasets [13,17,18].

Dataset	Transcriptional phenotype	Cellular phenotype	Dynamical phenotype
WT	WT cycling	Normal cell cycle	I: WT cycling
<i>cdc20Δ</i>	Clb2 ON with TF cycling	Metaphase-to-anaphase arrest	II: mutant cycling
<i>clbΔ</i>	Clb2 OFF with TF cycling	G1/S arrest	II: mutant cycling
<i>cdc14-3</i>	Clb2 INT-H with TF cycling	Mitotic exit arrest	II: mutant cycling
<i>cdc15-2</i>	Clb2 INT-L with TF cycling	Mitotic exit arrest	II: mutant cycling
<i>cse4</i>	SAC steady state	Metaphase-to-anaphase arrest	III: checkpoint arrest
<i>cdc8</i>	DRC steady state	G1/S arrest	III: checkpoint arrest

dataset acts as a baseline for the dynamics present within the cell cycle. Oscillations were seen in transcription factors and in cyclins both modeled in the cell-cycle GRN as well as outside it.

The Clb2 OFF mutant is a *clbΔ* mutant in which all mitotic cyclin genes *clb1-6* are disrupted [13]. The cellular phenotype corresponding to this transcriptional phenotype is arrest at the G1/S border. The Clb2 ON mutant is a *cdc20Δ* depletion mutant [18]. This knockout leads to cells arrested at the metaphase-to-anaphase border in mitosis containing high levels of non-oscillating Clb2 protein. The two mutant phenotypes with intermediate Clb2 expression are identified with the *cdc14-3* and *cdc15-2* mutants from [17]. It was observed that the temperature sensitive *cdc14-3* and *cdc15-2* mutants arrest at mitotic exit and had moderate levels of non-oscillating Clb2 activity [19,43], with *cdc14-3* cells having more Clb2 than *cdc15-2*. We accordingly name the corresponding phenotypes Clb2 INT-H and Clb2 INT-L. In all four of these mutant datasets, transcriptional oscillations were observed despite steady levels of Clb2.

We identify the SAC transcriptional phenotype from cells with a *cse4* mutation that disrupts spindle assembly and triggers the Spindle Assembly Checkpoint (SAC) [18]. The DNA replication checkpoint (DRC) transcriptional phenotype was collected from a *cdc8* mutant [18] that disrupts thymidylate kinase activity, thus inhibiting DNA synthesis [44] and causing arrest at the G1/S border. In both cases, transcriptional oscillations are silenced and the resulting steady states serve as the transcriptional phenotypes.

2.3. Dynamical phenotypes

Dynamical phenotypes can be observed in data, but are more precisely defined using modeling approaches. In the traditional ODE modeling paradigm, a mechanistic model is constructed, often using Hill functions in the GRN setting, and numerous parameters are either fit

to data or drawn from the literature. Frequently, a sensitivity analysis is performed to check the variability of model output to small perturbations in parameters. This approach, while valuable, has limitations. First, only a small fraction of parameter space can be explored, and second, the large number of parameters can lead to overfitting. We offer an alternative framework in which large regions of parameter space are excluded as unable to produce the desired dynamical behavior using a multi-level Boolean modeling approach. After identifying a reduced parameter space, parameterized Hill models can be created to replicate the desired behavior.

The modeling framework Dynamical Signatures Generated by Regulatory Networks (DSGRN) [26–28] is based on an ODE system with switching functions (Section 5.1) that leads to a fundamentally different approach. First, DSGRN provides a searchable database of coarse dynamics over the entirety of parameter space (Sections 5.2–5.4). This is possible because DSGRN divides parameter space into a finite number of regions, each called a DSGRN parameter node. The coarse dynamics are given by a collection of discrete transitions captured in a state transition graph. Any real-valued parameter set sampled from a fixed DSGRN parameter node will have the same state transition graph. Second, DSGRN uses only coarse information from a time series dataset, which avoids overfitting.

DSGRN output contains information on the number and type of fixed points (equilibria) and oscillations that the network can exhibit at a given DSGRN parameter node. The fixed points identified by DSGRN are not quantitative, they only indicate whether the associated gene product is predicted to have high, low, or intermediate concentrations. Likewise, cycling behavior is not modeled by a smooth trajectory, but rather a sequence of maxima and minima for each gene's expression level.

Model consistency in cycling behavior is determined by what we call a pattern match (Sections 5.5–5.6) between the sequence of maxima

Table 3

The fixed points representing the SAC and DRC based on the comparison between the WT dataset and the *cse4* and *cdc8* datasets, respectively (see Fig. 2).

Checkpoint	Proxies	Swi4	Nrm1/Yox1	Ndd1	Swi5/Ace2	Clb2
SAC	All proxy sets	Low	Low	High	High	Intermediate/high
DRC	Yox1 proxy sets	Low	Low	High	High	Intermediate/high
	Nrm1 Proxy sets	Low	High	High	High	Intermediate/high

and minima predicted by a network model and the observed sequence of maxima and minima for each gene in the data. Model consistency for a fixed point involves a judgment call on whether gene expression is best described as high, low, or intermediate at the time when steady behavior is observed. For both types of model consistency, the proportion of DSGRN parameter nodes that exhibit the observed behavior indicates how robustly the network model recapitulates the data. We remark that increased robustness, as measured by model consistency with the data across an enlarged number of parameter nodes, implies greater non-identifiability of the network model.

Each DSGRN parameter can be decomposed into a collection of independent DSGRN factor parameters, one per node in a GRN. In the SCC network (Fig. 1 Right), we will distinguish between the DSGRN factor parameter for Clb2 (the Clb2 parameter, Section 5.3) and the collection of remaining DSGRN factor parameters, which we call the TFO subnetwork parameter. In particular, we explore the behavior of the SCC network as the Clb2 factor parameter changes, but the TFO subnetwork parameter remains fixed, mimicking a control mechanism of Clb2 on the TFO subnetwork.

We defined three different dynamical phenotypes based on DSGRN output that are intended to represent important features of the experimental data and associated transcriptional phenotypes; see the correspondence in Table 2. Dynamical phenotype I (WT cycling) is a pattern match to the wild-type data within a stable DSGRN cycle. This phenotype is roughly analogous to ensuring similar phase relationships between expression levels that are robust to small perturbations. We identify the DSGRN parameter nodes at which consistency with the WT cycling in the data occurs.

Dynamical phenotype II captures Clb2 mutant cycling in which only the expression levels in the TFO subnetwork are stably oscillating. In addition, we seek DSGRN parameter pairs that differ only in the Clb2 parameter, where one shows WT cycling and the other shows mutant cycling. In other words, the TFO subnetwork parameter is fixed while the Clb2 parameter varies, and there is a pattern match to two different datasets. The implication of the existence of such a pair of DSGRN parameters is that the TFO subnetwork is predicted to operate independently of changes to Clb2.

In dynamical phenotype III (checkpoint arrest), we identified DSGRN fixed point (FP) surrogates of the SAC and DRC equilibria. We examined the data in [18] for the *cse4* mutant (SAC), and the *cdc8* mutant (DRC) and assigned qualitative equilibrium values, see Fig. 2 and Table 3. Qualitatively high was assigned when the endpoint of a gene transcript level in mutant time series was substantially above 50% maximum WT transcript level. Similarly, low was assigned for endpoint mutant transcript levels substantially below 50% maximum WT. Intermediate was assigned in the remaining cases, where endpoint mutant transcript levels were in the neighborhood of 50% maximum WT. We remark that all seven datasets were normalized together to permit such comparisons (see Data Availability).

The SAC FP is given in Table 3, first row. Notice that Clb2 expression levels are largely indeterminate since they have not achieved steady state by the end of the time series. For the DRC mutant in Table 3 second row, the SAC and DRC FPs are identical when considering both Yox1 proxy sets, indicating that the SCC network has an insufficient diversity of nodes to distinguish between the two checkpoints. For the Nrm1 proxy sets, the fixed points differ at the Nrm1 value, where the SAC exhibits low Nrm1 activity and the DRC exhibits high Nrm1 activity.

As in mutant cycling, we checked for the existence of DSGRN parameter node pairs where one showed WT cycling and the other a SAC/DRC FP with the same TFO subnetwork parameter. Unlike in mutant cycling, the arrest at a checkpoint in phenotype III indicates different dynamical behavior in the TFO subnetwork as compared to WT. Therefore, the existence of such a pair indicates that regulation through Clb2 alone is sufficient to control entry into a checkpoint within the SCC network model without additional external regulation at other network nodes. It is important to stress that this does not exclude the existence of other regulators in a larger network in the cell; it merely indicates that the SCC network model as constructed can replicate the mutant phenotype of interest.

3. Results

We assess the consistency of the predictions of the SCC network model with the datasets by checking for the existence of the DSGRN dynamical phenotypes at a collection of DSGRN parameters. The number of DSGRN parameter nodes for the SCC network is quite large: 275 466 240 total. We first divide DSGRN parameters into five distinct groups as described below (see Section 5.3 for technical details). This division of DSGRN parameter space is based solely on the Clb2 DSGRN factor parameter, and not on the TFO subnetwork parameter.

A full 60% of DSGRN parameter space allows for changing levels of Clb2 and therefore has the capacity for exhibiting the wild-type cycling phenotype. The remaining 40% of DSGRN parameter space is composed of parameters that have a fixed value of Clb2, with 10% each at high, low, and intermediate-high/low. These correspond to the mutant cycling dynamical phenotypes. We enforce that the SCC network model predictions can only be consistent with the Clb2 mutant datasets at the appropriately assigned parameters for Clb2 ON (fixed high), OFF (fixed low), INT-H (fixed intermediate high), and INT-L (fixed intermediate low). The checkpoint datasets do not provide any *a priori* parameter constraints and consistency is assessed over 100% of DSGRN parameter space (checkpoint arrest phenotypes). The subsets of DSGRN parameter nodes over which matches to given datasets are searched are called *phenotype-permissible* parameters. In particular, WT cycling phenotype-permissible parameters comprise 60% of parameter space, mutant cycling phenotype-permissible parameters comprise 10% of parameter space each, and checkpoint phenotype-permissible parameters are unconstrained.

Our results are reported as proportions over subsets of DSGRN parameter space. Dynamical phenotype I results are the percentages of WT cycling matches over 60% of parameter space. From this set of DSGRN parameters, the number of unique TFO subnetwork parameters was extracted and used as the normalization term for dynamical phenotypes II and III. This was justified because we sought to discover regions of parameter space where a change in Clb2 was the only impact on the network behavior; therefore, we looked for TFO parameters that showed WT cycling and matched at least one other phenotype.

3.1. Dynamical phenotype I: Wild-type cycling

The purpose of dynamical phenotype I is to show that the SCC network model can match the normal function of the cell cycle, namely the experimental WT dataset, by identifying the percent of phenotype-permissible DSGRN parameter nodes that can be pattern matched in a stable cycle. Table 4 shows that the proxy sets have WT pattern matches

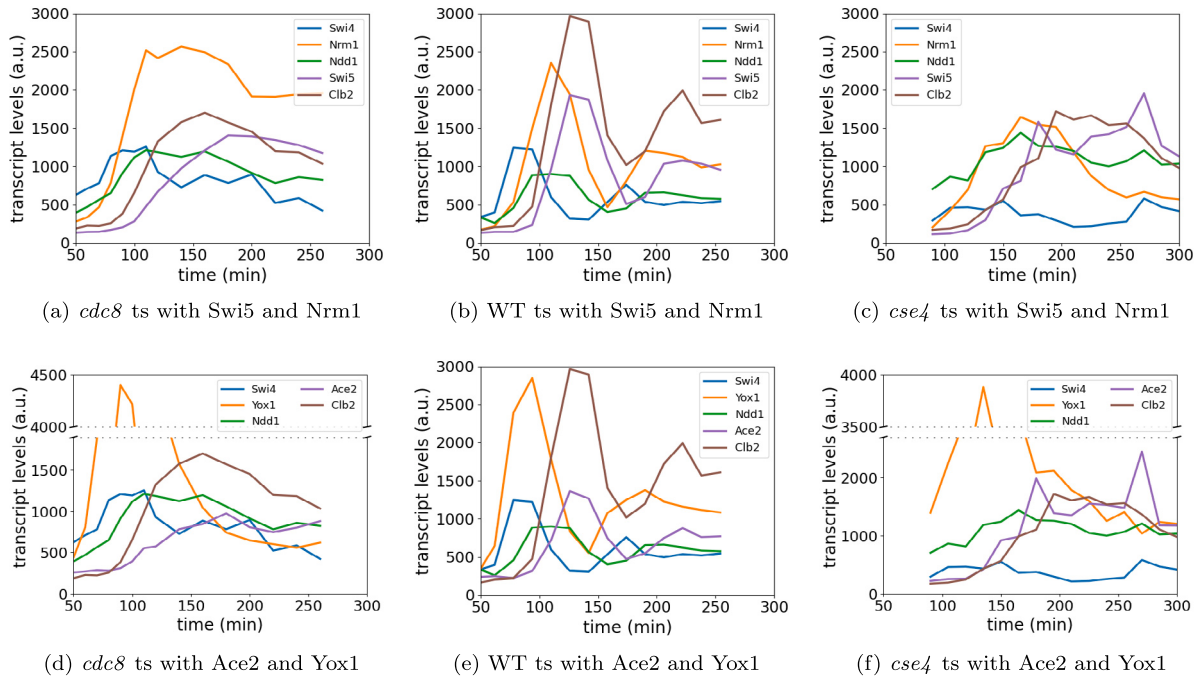


Fig. 2. (a) and (d) The transcript levels for proxy sets in the *cdc8* mutant exhibiting DRC arrest. (b) and (e) The transcript levels for proxy sets in the WT data. (c) and (f) The transcript levels for proxy sets in the *cse4* mutant exhibiting SAC arrest. The datasets have been jointly normalized and WT data is in the middle to facilitate comparison with both checkpoint datasets.

Table 4

Dynamical phenotype I: wild-type cycling results. The raw number of matches and the corresponding proportion of phenotype-permissible parameters are shown in columns 1 and 2. Column 3 shows the number of distinct TFO subnetwork parameters within the WT cycling matches. The numbers in column 3 are the normalization constants in Fig. 4.

DSGRN Phenotype I			
Proxies	matches	% of matches	TFO parameters
Swi5-Nrm1	7 073 137	4.3%	1 541 616
Ace2-Nrm1	7 073 137	4.3%	1 541 616
Swi5-Yox1	6 071 292	3.7%	1 347 895
Ace2-Yox1	6 071 292	3.7%	1 347 895

ranging from 3.7% up to 4.3% of phenotype-permissible DSGRN parameters. This demonstrates that every proxy set can recapitulate the oscillating sequence of maxima and minima seen in the WT data, and therefore the SCC network model cannot be excluded as a reasonable network model for controlling the yeast cell cycle.

In addition, matching WT cycling greatly reduced parameter space. We parameterized Hill models from this reduced region (see Section 5.7) to demonstrate the utility of DSGRN in locating useful regions of parameter space. The result of one such simulation is shown in Fig. 3(b) with the comparable WT data shown in Fig. 3(a). Both sets of curves are normalized between 0 and 1 to emphasize the order of maxima and minima in the time series, which is the behavior that DSGRN predicts. We see that in both cases, the peak of Swi4 precedes the peaks of Nrm1 and Ndd1, which themselves precede the peaks of Swi5 and Clb2. Swi5 and Clb2 have nearly indistinguishable peaks in both panels. The order of the Ndd1 and Nrm1 peaks are indistinguishable in the WT data in panel (a), but are ordered with Ndd1 first in the simulation in (b). Since Ndd1 has a faster rise in the WT data, the simulation ordering shows consistency with the WT data. We regard the recapitulation of this order of extrema as a successful model; however, we made no attempt to match either amplitude or period of the WT data.

We determined the number of unique TFO subnetwork parameters in the set of WT cycling pattern matches, shown in Table 4, column 3 for each proxy. These numbers are the normalization factors for the dynamical phenotypes II and III results shown in Fig. 4.

3.2. Dynamical phenotype II: Clb2 mutant cycling

We checked for the existence of dynamical phenotype II (mutant cycling) at Clb2 ON, Clb2 OFF, Clb2 INT-H, and Clb2 INT-L phenotype-permissible parameters; i.e., we checked for the co-existence of mutant and WT cycles at a single TFO parameter, which would indicate that a change in Clb2 alone does not disrupt TFO subnetwork oscillations in the SCC network.

The percentages of TFO parameters where this occurs are shown in Fig. 4. The nonzero results indicate that a perturbation of Clb2 allows the SCC network to transition from WT cycling to every type of mutant cycling in three of the four proxy sets.

The four Clb2 mutant datasets are shown in panels (a)–(d) of Fig. 5 normalized between 0 and 1. It can be seen that they all have approximately the same order of maxima and minima. Therefore, a single Hill model is sufficient to reproduce the behavior of all four datasets. We chose to sample TFO subnetwork parameters for the Swi5-Nrm1 proxy group that are predicted by DSGRN to be able to exhibit WT cycling as well as all four types of mutant cycling for varying Clb2 factor parameters (see Section 3.4). Fig. 5 (e)–(f) show simulations for two different TFO parameters that match the expected order of maxima and minima. We remark that Clb2 expression also oscillated in these simulations (not shown). This case is not excluded by DSGRN, since DSGRN will not predict oscillations that are too small to impact downstream targets.

It may seem inconsistent that there are Hill models that can recapitulate all four mutant cycling datasets, and yet the percent of pattern matches for the Clb2 ON, OFF, INT-L, and INT-H in Fig. 4 are not identical within each proxy set. This occurs because the DSGRN pattern matching methodology is sensitive to measurement noise that introduces spurious maxima and minima in the dataset.

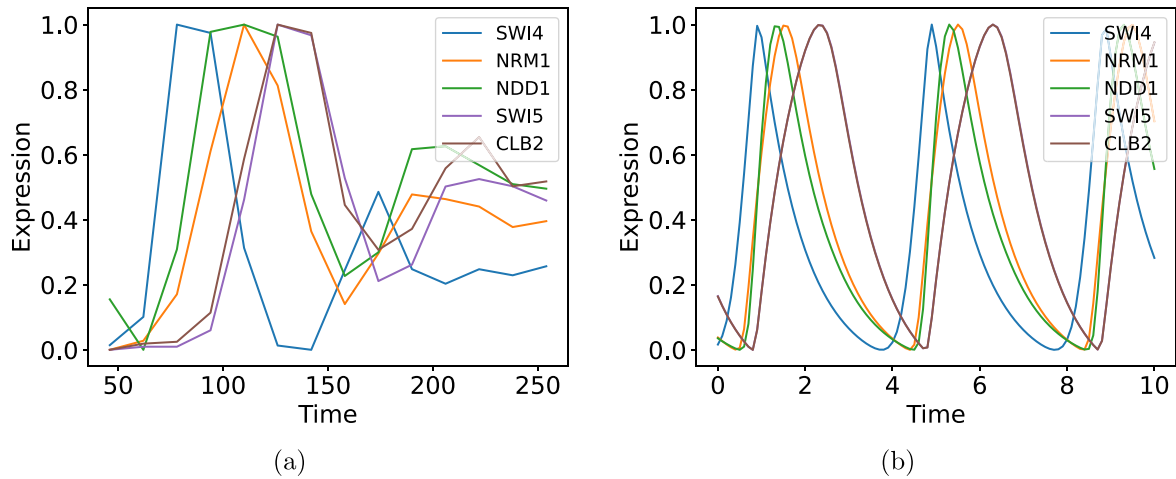


Fig. 3. Hill model simulation for wild type cells. (a) Wild type data normalized between zero and one. (b) Hill model simulation using a DSGRN TFO parameter predicted to exhibit wild type oscillations, also normalized. The curves for Swi5 and Clb2 are (nearly) coincident. No attempt was made to match amplitude or period, only the ordering of extrema.

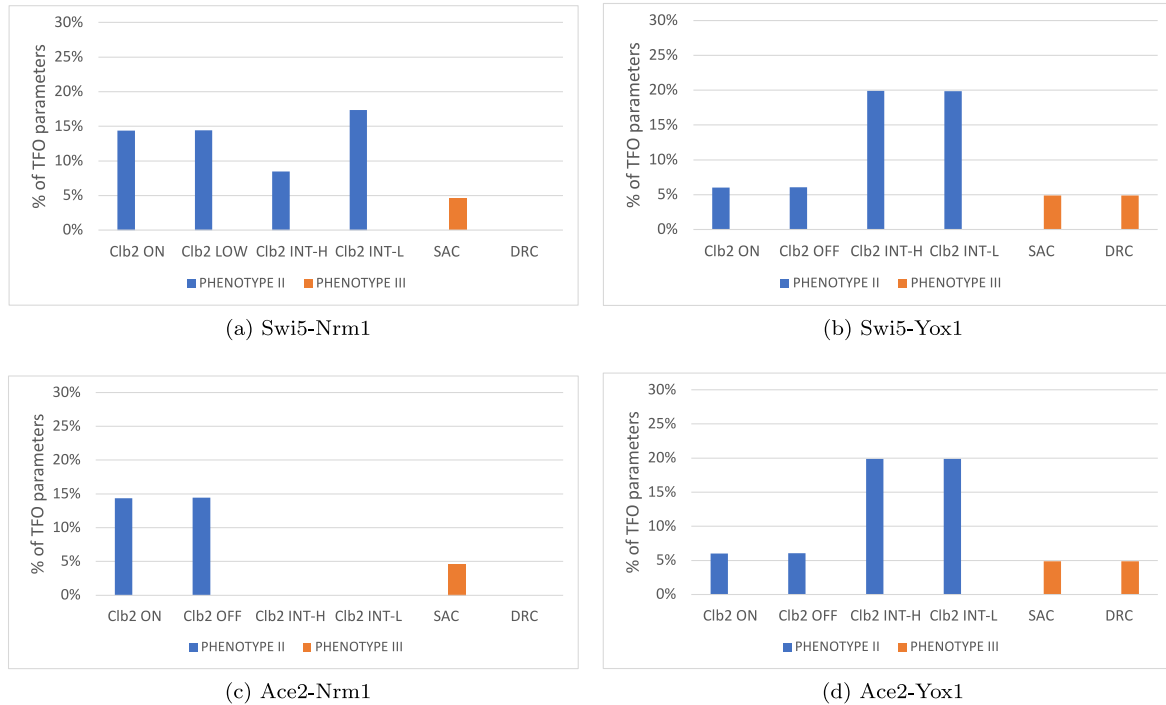


Fig. 4. The percentage of WT cycling TFO parameters that also are predicted to show behavior consistent with a dataset associated to dynamical phenotype II or III. The x-axis is labeled by the transcriptional phenotypes shown in Table 2. All percentages are nonzero except for the Clb2 INT-H and INT-L transcriptional phenotypes for the Ace2 proxy groups and the DRC transcriptional phenotype for the Nrm1 proxy groups.

3.3. Dynamical phenotype III: Checkpoint arrest

In DSGRN phenotype III, we identified TFO parameters that were predicted to exhibit a WT pattern match at one Clb2 factor parameter and a SAC or DRC FP at another. In Fig. 4, the “SAC” bars correspond to the percentages of TFO parameters exhibiting WT cycling and the SAC FP, which is identically the DRC FP for the Yox1 proxies seen by looking at the “DRC” bars. The fact that these percentages are nonzero indicates that the Yox1 proxies are consistent with both mutant datasets, which must be true given that the dynamical phenotype representation of the two datasets is identical.

On the other hand, the Nrm1 proxies are only consistent with the *cse4* mutant dataset representing the SAC phenotype. The DRC FP for the Nrm1 proxies was absent not only in TFO parameters with WT

cycling, but also absent across all of DSGRN parameter space. This finding indicates that the SCC network topology contains insufficient regulatory interactions to recapitulate the DRC phenotype when the Nrm1 time trace is considered.

An example of a simulation from a Hill model parameterized using a DSGRN TFO factor parameter predicted to show both WT cycling and a SAC FP is shown in Fig. 6 for the Swi5-Nrm1 proxy group. It is easily seen by examination that the simulation traces are consistent with the SAC FP in Table 3.

3.4. Co-existing phenotypes

We looked for TFO subnetwork parameters that could support all four mutant cycling phenotypes in addition to WT cycling at phenotype-permissible parameters. There were none such, indicating that either

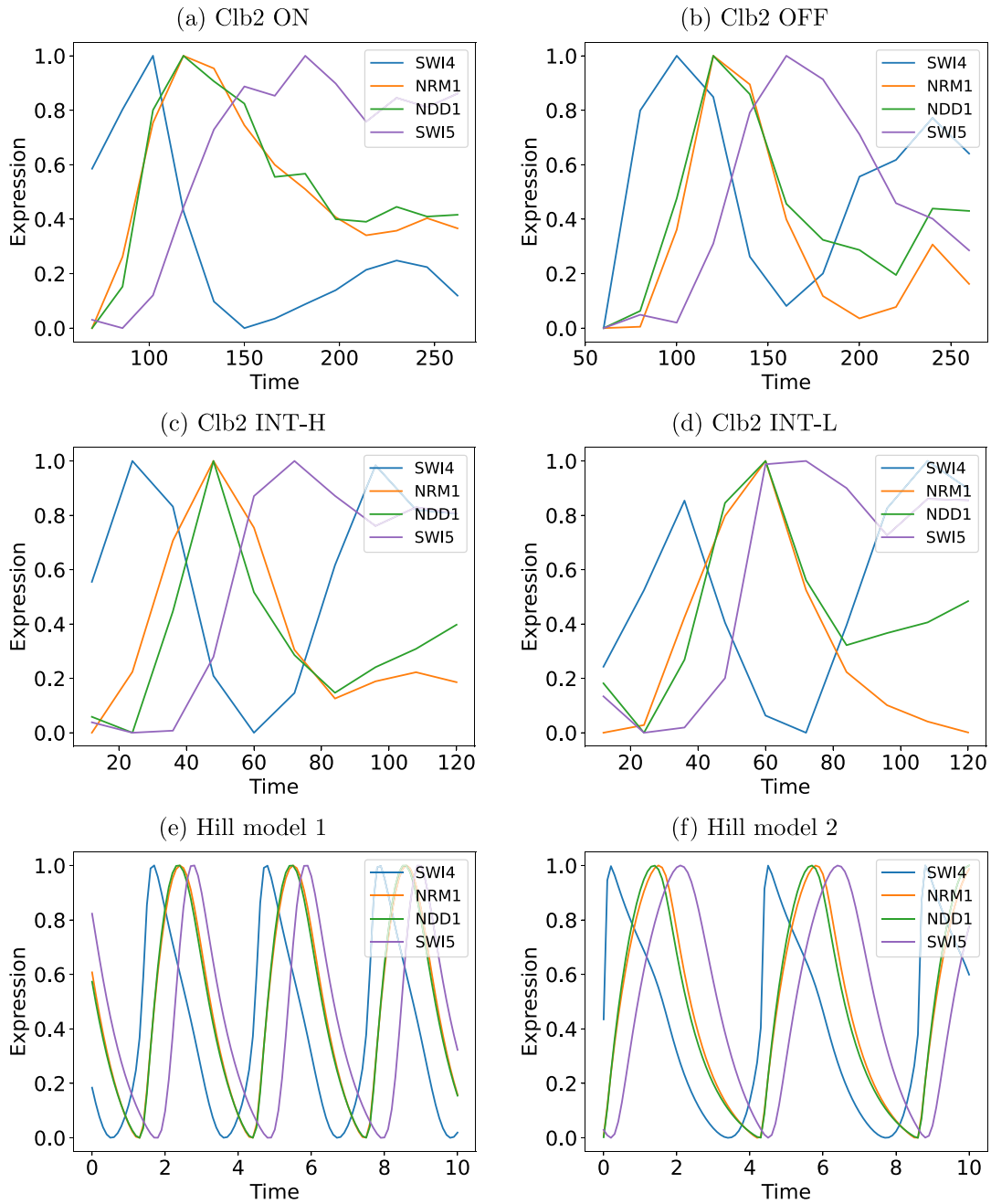


Fig. 5. Hill modeling for mutant cycling in the TFO subnetwork. (a)–(d) The four mutant datasets associated to phenotype II, normalized between 0 and 1. (e)–(f) Hill models parameterized to match the order of extrema in the datasets. Notice that all four datasets show the Swi4 peak preceding the Nrm1 and Ndd1 peaks (which are nearly coincident), followed by the Swi5 peak. Therefore, all four datasets can be roughly matched by the same parameterization of the TFO subnetwork, and two examples are given in (e)–(f) for very different TFO parameters. In both of these choices, the peak order is consistent with the datasets. As in the WT simulation, we made no attempt to match amplitude or period.

an internal change in or an external regulation of the TFO subnetwork occurs between mutant phenotypes. This result is dependent on the modeling choice of using only phenotype-permissible parameters associated to the Clb2 mutant cycling datasets. If this restriction is relaxed, the percentage of TFO subnetwork parameters at which there is both WT cycling and a single mutant cycling dataset increases (compare Fig. 7 to Fig. 4). More precisely, relaxing the phenotype-permissible restriction means that mutant cycling phenotypes are searched over all values of the Clb2 factor parameter, not just the 10% of DSGRN parameter nodes originally specified.

However, relaxing the phenotype-permissible modeling restriction still does not allow a TFO subnetwork parameter that can show matches

to every dataset in all three dynamical phenotypes, including checkpoint arrest. The closest achiever was the Swi5-Nrm1 proxy set for which the SCC network model was able to exhibit the WT cycling phenotype, the four mutant cycling phenotypes, and the SAC phenotype at 3575 TFO subnetwork parameters. This is 0.2% of DSGRN parameter nodes when compared to matching WT data alone, a substantial reduction in parameter space, although still large enough to signal an identifiability issue.

4. Discussion

We have demonstrated that a small network model (SCC network, Fig. 1 (Right)) approximating the pulse generation capacity of the

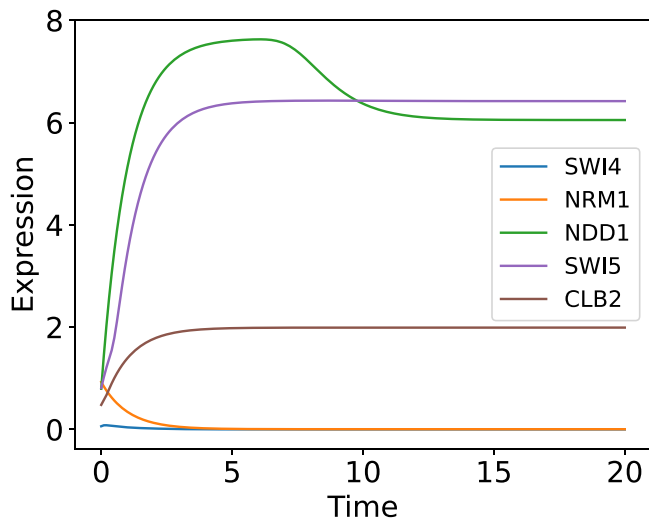


Fig. 6. Hill model of the SAC, where Swi4 and Nrm1 are low, Ndd1 and Swi5 are high, and Clb2 is not low.

yeast cell-cycle GRN (Fig. 1 (Left)) is capable of matching multiple datasets with different experimental perturbations. We devised three computational phenotypes using the software DSGRN [27,28] that associate to seven different datasets (Table 2). Dynamical phenotype I corresponds to wild-type cell cycle behavior. Dynamical phenotype II encompasses mutant cycling—oscillations of the TFO subnetwork under fixed values of the CDK Clb2 induced by various knockout experiments. In dynamical phenotype III, checkpoint behavior was evaluated by computationally seeking qualitative equilibrium values determined by each of two datasets, one representing the spindle assembly checkpoint and one representing the DNA replication checkpoint.

We interpreted the consistency of model predictions with the data in the following manner. When the SCC network model was consistent with wild-type cycling, then the model faithfully captured the behavior of the undisturbed cell cycle. When it was consistent with mutant cycling, then the model was capable of reproducing observed oscillatory behavior in the TFO subnetwork regardless of the state of Clb2. When it was consistent with checkpoint behavior, then regulation through Clb2 was sufficient to induce the checkpoint in the TFO subnetwork, even if in reality other regulators are implicated in cell cycle control. Consistency with all of these dynamical phenotypes would mean that the SCC network model cannot be rejected as a complete hypothesis for explaining the seven datasets, while the non-existence of any mutant phenotype would indicate network model incompleteness.

We found that 3.7–4.3% of potential wild-type DSGRN parameters predicted dynamical behavior consistent with the WT dataset, indicating that the network model is capable of reproducing cell-cycle behavior in the subset of proteins considered. Of these WT pattern-matched parameters, up to 20.0% were consistent with various mutant datasets, although not all phenotypes were exhibited at every proxy set. There were no TFO subnetwork parameters that could support WT cycling and all four types of mutant cycling within a single proxy group at phenotype-permissible DSGRN parameters, i.e., at fixed Clb2 levels hypothesized to be associated to particular datasets. However, if this modeling choice is not enforced, TFO subnetwork parameters were located that support WT and all four mutant cycling phenotypes, as well as (in some cases) SAC arrest. Relaxing phenotype-permissibility in the mutant cycling phenotype is akin to acknowledging that Clb2 expression in the datasets may not be sufficiently close to constant for our model assumptions to hold. If we allow for this possibility, then there exists a highly narrowed selection of TFO subnetwork parameters that can recapitulate the dynamics seen in multiple datasets. One interesting and open problem is whether this selection of parameters is most

likely to contain biologically reasonable parameterizations for the TFO subnetwork, since only Clb2 parameter modifications are necessary to induce phenotypic changes.

We observed that when using Nrm1 proxy sets, the DRC phenotype was not supported at any DSGRN parameter in the network model. This finding indicates that the model lacks important regulatory elements that are necessary for the DRC when considering Nrm1 as a proxy in the network. We hypothesize that distinguishing between SBF and MBF in the SCC network may help rectify this issue. Current models like the network seen in [17] define MBF and SBF as the same node, yet results from [45,46] indicate a mechanism for DRC arrest dependent solely on MBF activity. Given replication stress, the protein Rad53p inactivates the MBF co-repressor Nrm1. This activation of MBF induces up-regulation of G1/S genes within S-phase. As a result of this MBF pathway being activated, the DRC is initiated. We suggest that enlarging the SCC network to include MBF and Rad53p may permit consistency with the DRC phenotype.

There is another network enhancement that our work suggests. We showed that SAC arrest is supported in the network model by only modifying the parameterization of Clb2, suggesting the presence of a regulatory element controlled by the SAC mechanism that impinges solely on Clb2. One of the activities of the SAC is to inhibit APC/Cdc20 activity. Because APC/Cdc20 drives the degradation of Clb2, the outcome of SAC activation is the stabilization of Clb2 degradation [47]. In [48] it was seen that the rapid degradation of Cdc20 is necessary for SAC arrest. Using ordinary nonlinear differential equations and spatial simulations, the authors of [49] were able to identify a SAC mechanism that acts to fully sequester Cdc20 and inhibit APC activity. It was also suggested within [48] that there is potential for indirect regulation of Cdc20 through substrates of APC, such as Clb2 and Clb5. By expanding the SCC network to include feedback with APC/Cdc20, we can test the hypothesis that APC and Cdc20 interaction can initiate the Clb2 parameter change required to transition from WT cycling to SAC.

In its original implementation, DSGRN does not differentiate between transcriptional regulation vs protein modifications such as phosphorylation or ubiquitination, although these generalizations are newly available [41], as well as modeling transport across the nuclear membrane [50]. In this manuscript the original DSGRN package was used, and as a consequence transcriptional interactions and post-transcriptional modifications were not distinguished in the modeling of the SCC network in Fig. 1 (Right). This means that we used mRNA expression levels of Clb2 as a substitute for protein levels and activity. Protein and RNA levels are certainly correlated at least under some conditions [42]. An obvious next step is to expand the SCC network to model post-transcriptional modifications, as well as to incorporate other regulators. We are currently in the midst of an effort to explore the endocycling phenomenon in a larger network using the new methodology from [41].

There are also further analysis tools that are being developed in generality to leverage the computational methodology that we have presented here. We have shown that for fixed parameters of the TFO subnetwork, there can be two distinct dynamical behaviors for varying Clb2 parameters. The length of the path between these parameters can indicate how much perturbation in a control point is required to induce dynamical changes. Moreover, the path between the parameters predicts intermediate dynamics that might be revealed if the action of the control point is halted prematurely.

Another rich area for future research is the prediction of “exotic states” in a network model that may indicate novel cellular phenomena. For example, consider the 3575 TFO parameters identified in Section 3.4 that could match the Swi5-Nrm1 proxy set data for WT, the four types of mutant cycling, and the SAC phenotype, and therefore comprise the best hypothesis for biologically relevant parameters. Searching all DSGRN parameters associated to these TFO parameters reveals that monostable oscillations and fixed points dominate the dynamics. However, there are a substantial number of instances of bistable, tristable, and even quadrastable dynamics. What do these states represent? Are

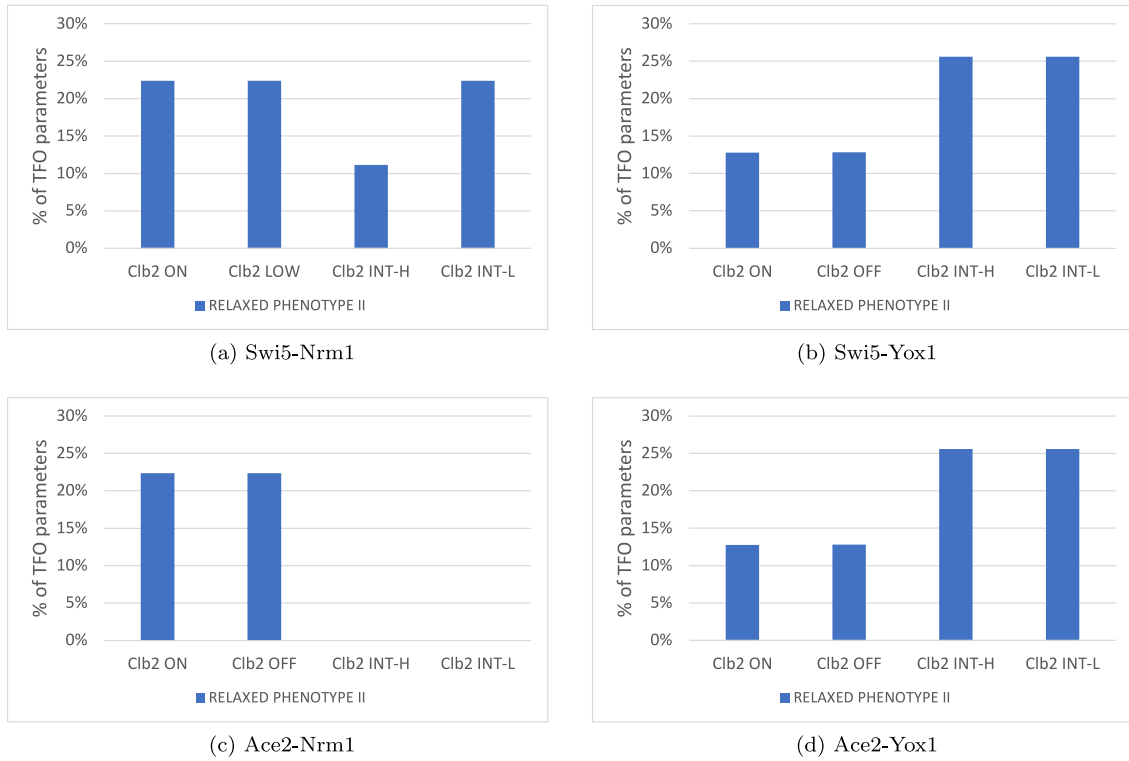


Fig. 7. The percentage of WT TFO parameters that also are predicted to show behavior consistent with a dataset associated to dynamical phenotype II after relaxing the phenotype-permissible restriction. Similar to the results from dynamical phenotype II with the phenotype-permissible restriction, all percentages are nonzero except for the Clb2 INT-H and INT-L transcriptional phenotypes for the Ace2 proxy groups.

they physiologically relevant? Can they be recovered through targeted experiments? These and other questions are raised by the predictive power of network models.

The results in this paper highlight the flexibility of dynamical behavior that a single network is capable of producing. We conclude that the SCC network is *controllable*, in the sense that a large number of dynamical phenotypes are accessible via parameter perturbation, and that the region of wild-type behavior is relatively small within parameter space. In contrast, a network might instead be *robust*, wherein wild-type behavior is prevalent across parameter space, making it difficult to push network behavior into a different dynamical regime. Even though in relative terms the simplified yeast cell-cycle GRN model is more controllable than robust, in absolute terms there remains a large region of parameter space that cannot be excluded as biologically reasonable. This idea makes intuitive sense, since the yeast cell cycle must operate under large array of environmental conditions and implies that *any* model of the cell-cycle GRN should incorporate some level of non-identifiability, so that parameters disrupted by some environmental conditions still result in the same qualitative dynamics. That is, non-identifiability may be a desirable feature of a model, and not always an obstacle to be overcome.

5. Methods

In this section, we discuss the basic properties of DSGRN [26–28,30] that are used when interpreting WT and mutant transcriptional phenotypes as DSGRN dynamical phenotypes. We discuss in detail how the dynamical phenotypes are constructed and interpreted with respect to the data.

DSGRN comprehensively computes coarse features of the dynamical behavior of a genetic regulatory network over a combinatorial representation of parameter space that is finite [27]. These coarse features include oscillatory behavior with stereotyped orders of maximum and minimum concentrations of gene products, and the number and type of

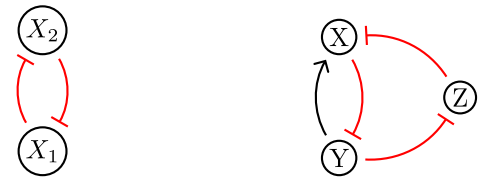


Fig. 8. Left: The two-node toggle switch where X_1 represses X_2 and X_2 represses X_1 . Right: An example three-node network, where X represses Y , Y activates X and represses Z , and Z represses X .

equilibria. DSGRN uses techniques from ordinary differential equations and graph theory to compute these behaviors.

Four kinds of graphs provide a framework for understanding DSGRN. A gene regulatory network (GRN) is the input to DSGRN and an undirected parameter graph (PG) is the basic structure of DSGRN. The output of DSGRN is a collection of directed state transition graphs (STGs) and their corresponding directed Morse graphs (MGs), one for each node in the parameter graph. As defined earlier a GRN is a system of interacting gene products that inhibit or activate one another. The GRN modeled in this paper is in Fig. 1 (Right), which is built as a directed graph. As a running example for explaining DSGRN computations in this section we will be talking about the simpler networks in Fig. 8. On the left, node X_1 inhibits node X_2 and X_2 inhibits X_1 , otherwise known as the “toggle switch” [30]. On the right, we see a three-node network that exhibits more complex dynamical behavior than the toggle switch, but remains amenable to manual construction of the various DSGRN graphs.

5.1. Switching systems and DSGRN parameters

DSGRN is, at its heart, a rigorous connection between Boolean modeling and ordinary differential equation (ODE) modeling. It can

equivalently be viewed as an asynchronous multi-level Boolean modeling approach [25] or as a dynamical systems approach [27]. In this exposition, we will introduce DSGRN using the formalism of dynamical systems.

A GRN can be modeled using a system of discontinuous ODEs called a switching system [51–55], where each node in the network is modeled by an ODE of the form:

$$\dot{x}_i = -\gamma_i x_i + \lambda_i(x), \quad i = 1, \dots, n, \quad (1)$$

where $x = (x_1, x_2, \dots, x_n)$. For every node in the network, x_i is the concentration of species i , γ_i is the decay rate of species i , and $\lambda_i(x)$ is a product of sums of step functions, one for each regulating edge j on node i . The step functions $\sigma_{i,j}^\pm(x_j)$ are given by:

$$\text{if } j \rightarrow i \text{ then } \sigma_{i,j}^+(x_j) = \begin{cases} l_{i,j} & \text{if } x_j < \theta_{i,j} \\ h_{i,j} & \text{if } x_j > \theta_{i,j} \end{cases} \quad (2)$$

$$\text{if } j \dashv i \text{ then } \sigma_{i,j}^-(x_j) = \begin{cases} l_{i,j} & \text{if } x_j > \theta_{i,j} \\ h_{i,j} & \text{if } x_j < \theta_{i,j} \end{cases} \quad (3)$$

representing activation and inhibition respectively, where $0 < l_{i,j} < h_{i,j}$ are low and high constant values. For an activating step function $\sigma_{i,j}^+(x_j)$, an increasing concentration of x_j causes an increasing rate of production of x_i as x_j crosses the threshold value $\theta_{i,j}$. Similarly, a repressing step function $\sigma_{i,j}^-(x_j)$ indicates a decreasing rate of production of x_i as x_j increases across the threshold $\theta_{i,j}$. In the case of the toggle switch we only see a single inhibiting edge per node so both nodes are defined by a single inhibiting step function, i.e. $\lambda_1(x) = \sigma^-(x_2)$ and $\lambda_2(x) = \sigma^-(x_1)$.

Suppose that node j regulates both i and k . Then we require that $\theta_{k,j} \neq \theta_{i,j}$ so that the effect of x_j on each of its downstream regulatory targets is distinct and therefore totally ordered. For example, the thresholds on the two out-edges from Y in Fig. 8 (Right) are required to be different; however there are no requirements on the relationship between any other pair of thresholds.

A DSGRN parameter is a collection of inequalities governing the relationship of the low, high, and threshold values for each node within the network. Each DSGRN parameter consists of two parts for each node in the network: a logic parameter and an order parameter [27]. A key observation is that the logic and order parameters for a node are independent of all other nodes in the network, and therefore may be chosen independently. An order parameter defines the order of the threshold values for a node. For example, the Y node in Fig. 8 (Right) has two threshold values due to its two out-edges, $\theta_{X,Y}$ and $\theta_{Z,Y}$. There are then two order parameters for Y : $\theta_{X,Y} < \theta_{Z,Y}$ and $\theta_{X,Y} > \theta_{Z,Y}$. All other nodes in our examples have a single out-edge and are trivially ordered.

The number of thresholds associated to each node determines the discretization of the corresponding gene product's expression level. Considering the node Y above, the two thresholds mean that Y has three discrete expression levels, 0, 1, and 2 (low, intermediate, and high), where the integer indicates how many thresholds the value of Y exceeds; i.e., $Y < \theta_{X,Y}, \theta_{Z,Y}$ corresponds to 0, and so forth. The consequence is that nodes in a GRN can have, and generally do have, different discretization levels. In the SCC network, Swi4 and Clb2 each have 3 out-edges and therefore 4 states (low, intermediate low, intermediate high, and high); Ndd1 has 2 out-edges and therefore 3 states (low, intermediate, and high); and Nrm1/Yox1 and Swi5/Ace2 each have 1 out-edge and therefore the Boolean states 0 and 1 (low and high).

The logic parameter for each node within the network orders the input values to the node with respect to the output thresholds of the node. For example, the X node in Fig. 8 (Right) has two in-edges, one from Y and one from Z . Therefore, it has four possible input values:

$$l_{X,Y}l_{X,Z} < \{l_{X,Y}h_{X,Z}, h_{X,Y}l_{X,Z}\} < h_{X,Y}h_{X,Z}$$

that are partially ordered due to the constraint that $0 < l < h$. In particular, notice that the input values $l_{X,Y}h_{X,Z}$ and $h_{X,Y}l_{X,Z}$ cannot have a determined order until real values are assigned to l and h . The node X has a single out-edge to node Y , giving it one threshold $\theta_{Y,X}$. A logic parameter is the insertion of this threshold into the partial order of inputs. For X , there are six possible logic parameters, since there are six possible ways to insert the threshold into the partial order:

$$\begin{aligned} l_{X,Y}l_{X,Z} &< \{l_{X,Y}h_{X,Z}, h_{X,Y}l_{X,Z}\} < h_{X,Y}h_{X,Z} < \theta_{Y,X} \\ l_{X,Y}l_{X,Z} &< \{l_{X,Y}h_{X,Z}, h_{X,Y}l_{X,Z}\} < \theta_{Y,X} < h_{X,Y}h_{X,Z} \\ l_{X,Y}l_{X,Z} &< l_{X,Y}h_{X,Z} < \theta_{Y,X} < h_{X,Y}l_{X,Z} < h_{X,Y}h_{X,Z} \\ l_{X,Y}l_{X,Z} &< h_{X,Y}l_{X,Z} < \theta_{Y,X} < l_{X,Y}h_{X,Z} < h_{X,Y}h_{X,Z} \\ l_{X,Y}l_{X,Z} &< \theta_{Y,X} < \{l_{X,Y}h_{X,Z}, h_{X,Y}l_{X,Z}\} < h_{X,Y}h_{X,Z} \\ \theta_{Y,X} &< l_{X,Y}l_{X,Z} < \{l_{X,Y}h_{X,Z}, h_{X,Y}l_{X,Z}\} < h_{X,Y}h_{X,Z} \end{aligned}$$

A DSGRN parameter is a collection of inequalities: a choice of one order parameter and one logic parameter for each node in the network. We will call the inequalities for node i a *factor parameter* for i . Since DSGRN parameters correspond to sets of inequalities, they represent connected regions within real-valued parameter space. These regions can be sampled to parameterize the switching system, or a smooth approximation such as the Hill model simulations in Figs. 3, 5, and 6.

A logic parameter for node i determines whether or not every input edge to i can have a nontrivial impact on the dynamics of i , and whether or not i can have a nontrivial impact on its downstream targets. For example, consider the logic parameter

$$l_{X,Y}l_{X,Z} < h_{X,Y}l_{X,Z} < \theta_{Y,X} < l_{X,Y}h_{X,Z} < h_{X,Y}h_{X,Z}$$

for node X in Fig. 8 (Right). This logic parameter is one at which the input Y is an ineffective activator of X . To see why, notice that when the activator Y is missing and the repressor Z is present, the abundance of node X is given by its lowest possible value, $l_{X,Y}l_{X,Z}$. If the activator Y then becomes abundant, the abundance of X is given by $h_{X,Y}l_{X,Z}$. However, from the logic parameter we have that $h_{X,Y}l_{X,Z} < \theta_{Y,X}$, so that the abundance of X is still not high enough to repress its target Y . If X is to have an effect on Y , then its repressor Z must be absent. The presence or absence of Y does not make a difference. We say that the edge $Y \rightarrow X$ is “input inessential.” Similarly, consider the logic parameter

$$l_{X,Y}l_{X,Z} < \{l_{X,Y}h_{X,Z}, h_{X,Y}l_{X,Z}\} < h_{X,Y}h_{X,Z} < \theta_{Y,X}.$$

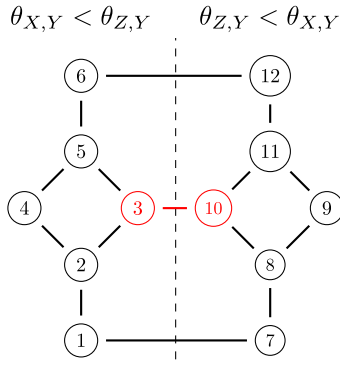
In this case, no combination of inputs can induce node X to exceed the threshold $\theta_{Y,X}$ and therefore repress Y . We say the edge $X \dashv Y$ is “output inessential.” A logic parameter that describes at least one input or output inessential edge is itself called inessential. A logic parameter is essential otherwise, and describes a situation where every input to node i is an effective regulator, and where there are combinations of inputs to node i that are above and below each one of the output thresholds of i .

We remark that the number of DSGRN parameters scales poorly with the number of edges in a network. As the number of in-edges to a node grows, the size of the partial order grows exponentially. As the number of out-edges from a node grows, the number of order parameters grows factorially. Additionally, the complexity of inserting multiple thresholds into the partial order causes a large increase in the number of logic parameters.

5.2. Parameter graph and remainder parameter

An important element of this work is the transition between WT and mutant dynamical phenotypes in parameter space. The following two sections explain the details of this transition.

The collection of all possible factor parameters for i can be represented as an undirected graph, called the factor graph for node i . These



$$\begin{aligned}
 3 : & \quad l_{Y,X} < \theta_{X,Y} < \theta_{Z,Y} < h_{Y,X} \\
 10 : & \quad l_{Y,X} < \theta_{Z,Y} < \theta_{X,Y} < h_{Y,X}
 \end{aligned}$$

Fig. 9. The factor graph for the Y node in Fig. 8 (Right), showing its two isomorphic logic graphs, one for each possible threshold ordering for Y . The threshold ordering can be seen above the corresponding logic graph of the factor graph.

factor graphs have nodes representing each factor parameter and the edges between these nodes represent a single change in an inequality.

Each factor graph for a node i is composed of M isomorphic logic graphs. The nodes of a logic graph are the collection of all logic parameters for i , and an edge exists between two logic graph nodes when there is a single change in a logic parameter inequality (for a fixed order parameter). The number of logic graphs M in the factor graph is the number of order parameters for node i . Connections between logic graphs exist whenever there is a single change in the order parameter while the logic parameter remains the same, provided that the two swapped thresholds in the order parameter have no intervening logic value. This is most easily seen by an example.

Let us now construct the factor graph for Y from the three-node network in Fig. 8 (Right). Node Y has a single in-edge and two out-edges, meaning that two thresholds are inserted between the low and high production rates of Y : $0 < l_{Y,X} < h_{Y,X}$. This results in six logic parameters for Y :

- 1 : $l_{Y,X} < h_{Y,X} < \theta_1 < \theta_2$
- 2 : $l_{Y,X} < \theta_1 < h_{Y,X} < \theta_2$
- 3 : $l_{Y,X} < \theta_1 < \theta_2 < h_{Y,X}$
- 4 : $\theta_1 < l_{Y,X} < h_{Y,X} < \theta_2$
- 5 : $\theta_1 < l_{Y,X} < \theta_2 < h_{Y,X}$
- 6 : $\theta_1 < \theta_2 < l_{Y,X} < h_{Y,X}$

where $\theta_1 < \theta_2$ is some ordering of the two thresholds of Y , $\theta_{X,Y}$ and $\theta_{Z,Y}$. These six inequalities are the nodes of the logic graph of Y . Due to the two order parameters of Y , there are two copies of this logic graph, one for each ordering, as shown in Fig. 9. The numbering of the logic parameters above is associated to the logic graph left of the dashed line, which is associated to order parameter $\theta_{X,Y} < \theta_{Z,Y}$. The isomorphic logic graph for order parameter $\theta_{Z,Y} < \theta_{X,Y}$ is on the right. Edges only exist between the two logic graphs when there is a single inequality change in the order parameter between thresholds that are adjacent in the logic parameter. An example of this type of edge can be seen in red in Fig. 9.

For a GRN N with nodes $i = 1, \dots, n$, the product of the factor graphs $FG(i)$ is the DSGRN parameter graph $PG(N)$:

$$\prod_{i=1}^n FG(i) = PG(N).$$

The parameter graph contains all possible DSGRN parameters as nodes and encodes adjacency between real-valued parameter regions as edges [27].

As an example, the DSGRN parameter graph for the three-node network in Fig. 8 (Right) is shown in Fig. 10. Each choice of a node from $FG(X)$, a node from $FG(Y)$, and a node from $FG(Z)$ is a DSGRN parameter node. Two such DSGRN parameter nodes are shown in red in the top and bottom panels in Fig. 10. By examining the factor graphs,

we see that the DSGRN parameter graph for the three-node network has a size of $6 \times 12 \times 3 = 216$.

Consider a node x_i that is expected to undergo multiple perturbations, such as knockouts or up-regulations, modeled by variability in the factor parameter of x_i . The combination of the factor parameters for all of the remaining nodes form a *remainder parameter*. We demonstrate the idea of a remainder parameter using Fig. 10 by allowing the parameter node of Z to vary while those of X and Y remain fixed. In this case we have the remainder parameter composed of the X and Y factor parameters marked in red as Z varies from its highest parameter to its lowest. The high and low Z factor parameters could represent an up-regulation or knock-out of a gene in a biological scenario. In the first case, Z is continuously expressed at its highest level while in the second, it is expressed at its lowest level, possibly zero.

5.3. Application to the SCC network

Recall from Fig. 1 (Right) that Clb2 has one in-edge and three out-edges. This means it has $3! = 6$ order parameters and ten logic parameters (see Fig. 11). Four of the ten logic parameters are taken to be representative of various Clb2 mutants (see the caption of Fig. 11). We propose that the WT phenotype is associated to one of the logic parameters in black. Notice that each of the logic parameters that denote the state of a Clb2 mutant (inequalities in color) have both the low and high values $l_{Clb2,SCF}$ and $h_{Clb2,SCF}$ together between thresholds. In other words, the model of Clb2 activity implies that even if the molecular concentration is not perfectly constant, there will never be a sufficiently large change in concentration to trigger a change in regulation at downstream genes. In contrast, the logic parameters for Clb2 WT ensures that changes in Clb2 concentration will impact at least one downstream target. We do not require regulatory activity at all downstream nodes, because the only information provided by the data is that Clb2 has noticeable oscillations. The DSGRN parameters at which we see consistency with the experimental data can give us guidance in assessing which Clb2 interactions may be important in the WT phenotype.

Dynamical phenotypes II and III involve transitions in the Clb2 logic graph from WT to mutant parameters that exhibit consistency with the corresponding mutant time series when the remainder parameter (referred to as the TFO subnetwork parameter) is constant. As an example, consider the Clb2 ON transcriptional phenotype. A match according to dynamical phenotype II is a single TFO subnetwork parameter that exhibits (1) mutant cycling consistent with the *cdc20A* dataset at the phenotype-permissible green node at the top of the Clb2 logic graph as well as (2) WT cycling at any one of the six black nodes in the Clb2 logic graph.

Notice that the Clb2 logic parameter graph includes both essential and inessential logic parameters (see the end of Section 5.1). The results in this manuscript assume that only Clb2 may assume inessential logic parameters. The TFO subnetwork parameters are always essential.

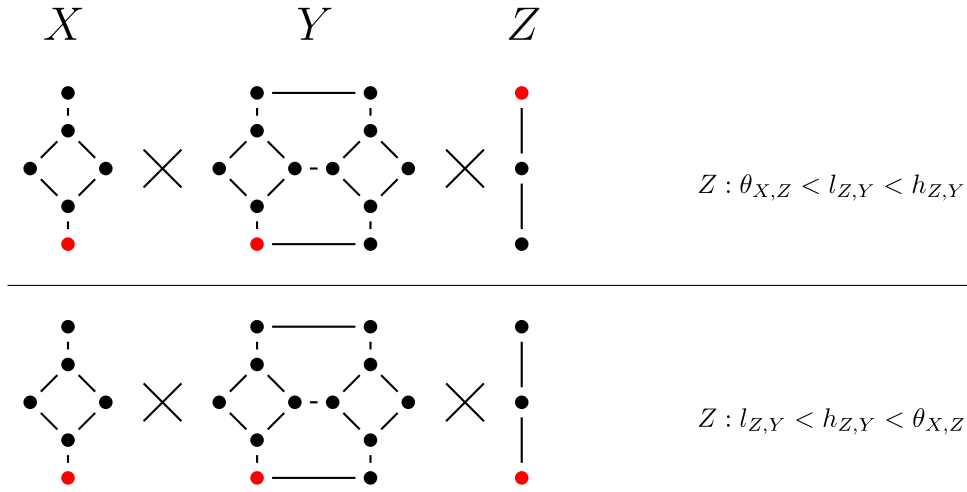


Fig. 10. Two copies of the DSGRN parameter graph PG of the three-node network in Fig. 8 (Right), each one showing a different DSGRN parameter in red. PG is the product of three factor graphs corresponding to the nodes X , Y , and Z . The example X , Y , Z triples of red factor parameters (top panel and bottom panel) correspond to two different DSGRN parameters for the network. The fixed X , Y factor parameters together form a remainder parameter that is fixed as the factor parameter of Z varies. The inequalities for the two different factor parameters for Z are shown to the left.

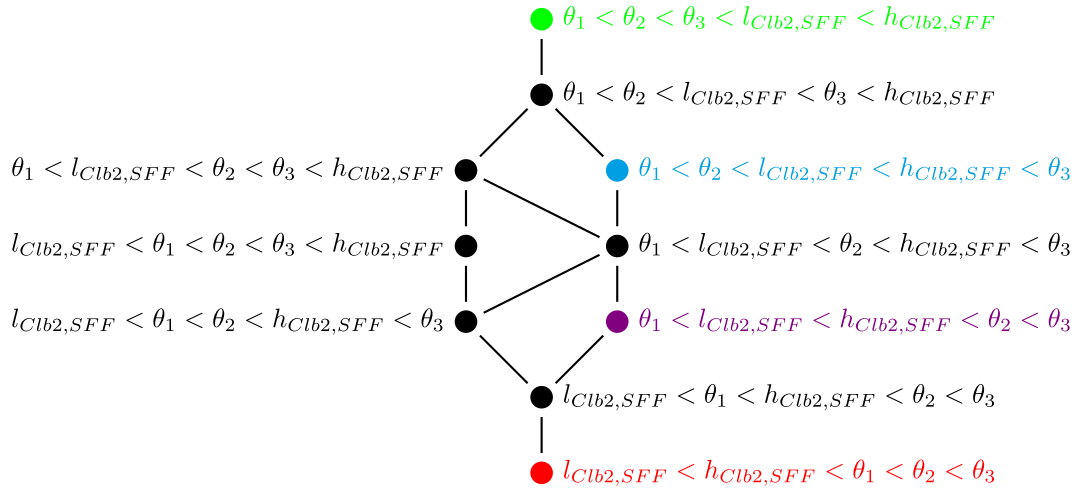


Fig. 11. One of the six logic parameter graphs corresponding to the Clb2 node from the SCC network in Fig. 1 (Right), where the thresholds $\theta_1, \theta_2, \theta_3$ are associated to the targets SBF, SFF, and Swi5 via some fixed mapping. The Clb2 ON logic parameter is represented in green at the top of the factor graph. The Clb2 OFF logic parameter is represented in red at the bottom of the factor graph. The Clb2 INT-H logic parameter in blue is two steps up from the Clb2 INT-L logic parameter in violet. The WT logic parameters are the remaining black inequalities. The checkpoint phenotypes are not restricted to any particular Clb2 logic parameter.

5.4. State transition graphs and Morse graphs

Each DSGRN parameter has a corresponding state transition graph (STG) that graphically represents the dynamics of the network at that DSGRN parameter. For ODE systems, the dynamics of a network are described by time-dependent trajectories in phase space in which all gene products associated to the network are changing concentration. Phase space is the N -dimensional real-valued and positive space where each coordinate represents a node in the GRN. DSGRN does not concern itself with these trajectories but instead looks at directional flow across thresholds and considers paths through an STG.

An example phase space for the toggle switch from Fig. 8 (Left) is shown in Fig. 12. We discretize phase space by dividing it up into rectangular boxes called domains using the collection of thresholds for the switching system; these are the dotted lines in Fig. 12, showing the division of the positive plane into four domains. Each domain corresponds to a level of X_1 and X_2 where 1 indicates above-threshold values and 0 indicates below-threshold values. As an example the domain labeled (01) corresponds to high X_2 and low X_1 . We say that (01) is the *state* corresponding to the upper left domain.

We can define flow across the boundaries of the domains by specifying a DSGRN parameter. In Fig. 12, we use

$$\text{DSGRN parameter :} \quad (4)$$

$$X_1 : l_{1,2} < \theta_{2,1} < h_{1,2}$$

$$X_2 : l_{2,1} < h_{2,1} < \theta_{1,2}$$

to determine the arrows. For example, the top arrow pointing left and the right arrow pointing down are determined by considering the starting domain, where X_1 and X_2 have high concentrations. In this case, the concentrations of X_1 and X_2 are both decreasing since the nodes are effective in repressing each other. However, note that X_2 is *always* decreasing. This happens because of the DSGRN parameter for X_2 . Regardless of the presence or absence of X_1 , X_2 will never increase across its threshold $\theta_{1,2}$ since its high concentration is still below $\theta_{1,2}$. There is also a DSGRN parameter where X_2 (or symmetrically X_1) is always increasing, indicating that one or both nodes are ineffective repressors.

The STG for the toggle switch is superposed on phase space in Fig. 12. The nodes of the STG correspond to the states of the domains

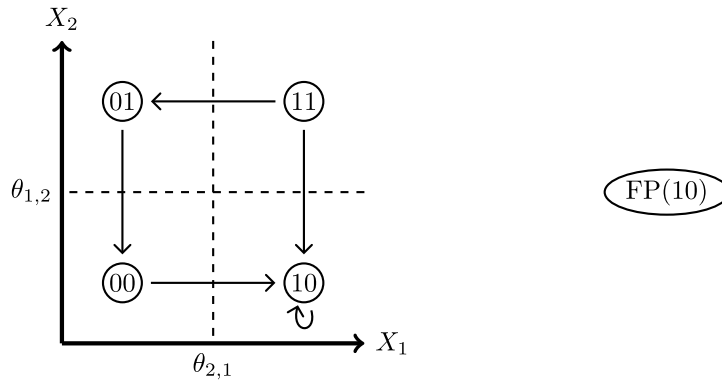


Fig. 12. (Left) The STG superposed over the rectangular domains dividing phase space for the toggle switch with the DSGRN parameter in (4). (Right) The corresponding MG.

Table 5

The order of the nodes for the checkpoint phenotype FPs is Swi4, which has activity levels in the range 0–3, Nrm1/Yox1, which has levels 0–1, Ndd1, which has levels 0–2, Clb2, which has levels 0–3, and Swi5/Ace2, which has levels 0–1, dependent on the number of out-edges from each node in the SCC network shown in Fig. 1 (Right) (see Section 5.1 for details).

Checkpoint FPs: FP(Swi4, Nrm1/Yox1, Ndd1, Swi5/Ace2, Clb2)		
SAC FPs:	All proxies:	
	FP(0, 0, 2, 1, 1)	
	FP(0, 0, 2, 1, 2)	
	FP(0, 0, 2, 1, 3)	
DRC FPs:	Yox1 proxies:	Nrm1 Proxies:
	FP(0, 0, 2, 1, 1)	FP(0, 1, 2, 1, 1)
	FP(0, 0, 2, 1, 2)	FP(0, 1, 2, 1, 2)
	FP(0, 0, 2, 1, 3)	FP(0, 1, 2, 1, 3)

and the directed edges between states generally correspond to the flow across the intervening boundary. The one exception is that self-edges are added whenever it is not possible to leave that domain because all of the flow points inward; see the lower right domain in Fig. 12.

The size of the STG grows with the size of the GRN, since the addition of a node to the GRN adds another dimension to phase space and the addition of an edge, or threshold value, adds another set of domains to an existing dimension of phase space. Both of these mechanisms increase the number of nodes in the STG, which rapidly becomes large and difficult to interpret. It is therefore useful for clarity to examine a summary of the STG called a Morse graph (MG) [27]. We build the MG from the recurrent components of the STG. A recurrent component R , or Morse set, is a maximal set of nodes in the STG that induces a subgraph containing a path from any node u to any other node v for all $u, v \in R$. In particular, a recurrent component is a strongly connected component with at least one edge. Each of these Morse sets are represented in the MG as Morse nodes, where an edge between Morse nodes indicates that there is a path between some node u_1 in Morse set 1 to some node u_2 in Morse set 2.

Each Morse node in the MG has an annotation indicating the dynamics in the associated Morse set. The annotations consist of the labels full cycle (FC), partial cycle (PC), and fixed point (FP). The FC annotation indicates a Morse set in which there is a recurrent component in the STG such that every node x_i changes its integer state at least once. For example, if a Morse node for some parameter of the SCC network was annotated FC, then all nodes SBF, SFF, Nrm1/Yox1, Swi5/Ace2, and Clb2 would have to change state from, e.g., 1 to 0, etc. The PC annotation indicates a recurrent component in which only a subset of the x_i change state. The FP annotation indicates a Morse set consisting of a single state with a self-edge. The Morse graph for the toggle switch is the FP with state coordinates (10), denoted FP(10); see Fig. 12. The SAC and DRC FPs for the SCC network that are shown qualitatively in Table 3 are represented by the state coordinates shown in Table 5.

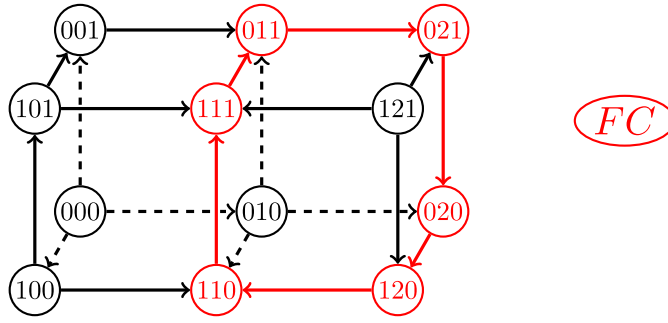
The Morse graph also encodes the *stability* of a Morse set. Stability in the sense of dynamical systems roughly means that trajectories close to a stable manifold in phase space will approach that manifold asymptotically over time. Stability of a Morse node in the Morse graph, whether a fixed point or a cycle, is identified with having no possible exit from a Morse set once a path enters the Morse set; i.e., a Morse node is stable if and only if it has no out-edges in the MG. Morse nodes with out-edges are unstable, since there are potential exit paths from the associated Morse set. In the SCC network model, wild-type cycling is evaluated within stable FCs, mutant cycling is evaluated in stable PCs where Clb2 is in a fixed state and the TFO subnetwork nodes are oscillating, and checkpoints are evaluated as FPs, which are always stable in this formulation.

In order to understand the implications of the different dynamic phenomena described above, we will analyze by hand the example 3-node network seen on the right in Fig. 8, which exhibits all possible annotations. We will be looking at three different DSGRN parameters that give rise to three different Morse graphs. Since this network has three nodes, phase space will have three dimensions where the states are represented in the order (XYZ) ; for example, (010), (110), etc. The X and Z dimensions are divided into two domains each since they have only one out-edge apiece. The Y dimension contains three domains since there are two out-edges. This means that instead of two states, 0 and 1, Y has three states, 0, 1, and 2, representing the three possible positions with respect to two distinct thresholds.

The first example DSGRN parameter we will investigate gives rise to an MG containing a stable FC. The corresponding STG and DSGRN parameter can be seen in Fig. 13. The nodes have been arranged in the same spatial order as the rectangular domains in phase space. The red arrows and nodes indicate the path that contains the FC. It can be seen through inspection of the STG that starting anywhere within the STG leads to this stable cycle, in which there are state changes in all three directions.

The second example DSGRN parameter contains both a stable and an unstable PC in the variables X and Y . The STG and DSGRN parameter inequalities are shown in Fig. 14. Again the stable cycle is represented by the red arrows and nodes while the unstable cycle is represented in the blue arrows and nodes. Through inspection of the STG it is easily seen that the blue cycle can be exited and the red cycle cannot. Unlike the FC parameter we only see oscillations in a subset of gene product concentrations in the GRN, seen because the cycle only moves through two of the three dimensions. In particular, the stable cycle exists when Z has state 0 and the unstable cycle exists when Z has state 1.

The last DSGRN parameter we will look at in this network exhibits an FP. The corresponding STG and inequalities can be seen in Fig. 15. The FP for this DSGRN parameter can be seen as the single red node with a self-edge within the STG, FP(100). Similar to the last example, there exists an unstable PC $\{X, Y\}$ that is represented with blue nodes



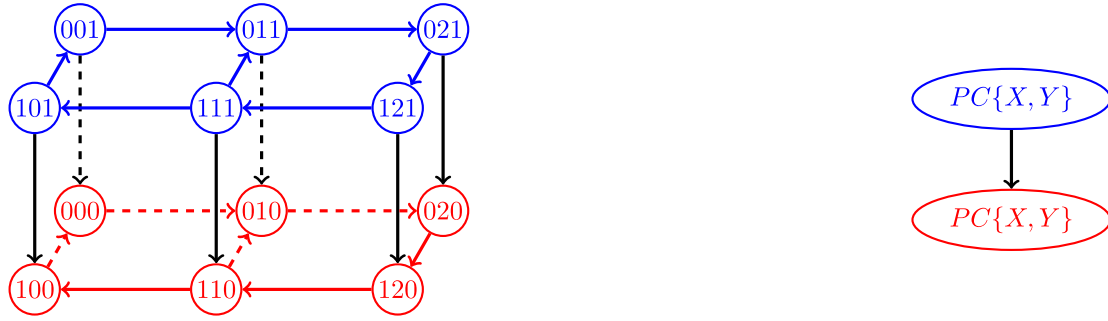
DSGRN parameter FC :

$$X : l_{X,Y}l_{X,Z} < h_{X,Y}l_{X,Z} < \theta_{Y,X} < l_{X,Y}h_{X,Z} < h_{X,Y}h_{X,Z}$$

$$Y : \theta_{X,Y} < l_{Y,X} < \theta_{Z,Y} < h_{Y,X}$$

$$Z : l_{Z,Y} < \theta_{X,Z} < h_{Z,Y}$$

Fig. 13. An example STG (Left) and MG (Right) computed from the given DSGRN parameter (Bottom) for the 3-node network in Fig. 8. The STG exhibits a stable full cycle shown by the red nodes and edges. The dashed lines are for visual effect only. They represent edges on the “back” of the cubes.



DSGRN parameter PC :

$$X : l_{X,Y}l_{X,Z} < l_{X,Y}h_{X,Z} < \theta_{Y,X} < h_{X,Y}l_{X,Z} < h_{X,Y}h_{X,Z}$$

$$Y : l_{Y,X} < \theta_{Z,Y} < \theta_{X,Y} < h_{Y,X}$$

$$Z : l_{Z,Y} < h_{Z,Y} < \theta_{X,Z}$$

Fig. 14. An example STG (Left) and MG (Right) computed from the given DSGRN parameter (Bottom) for the 3-node network in Fig. 8. The STG exhibits a stable partial cycle in the variables X and Y (red) and an unstable partial cycle in X and Y (blue).

and arrows. Through inspection it can be seen that the (100) domain has no out-edges and therefore must be an FP. Also notice that once the unstable cycle is exited one can never return to that cycle and will eventually end up at FP(100).

5.5. Time series discretization

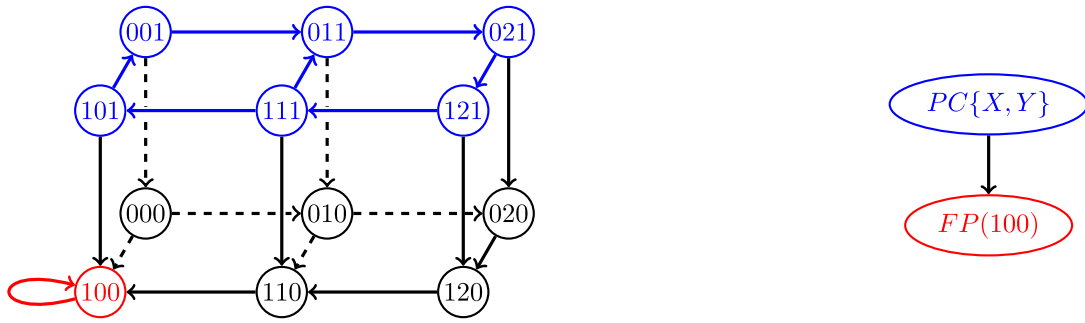
To evaluate DSGRN network model consistency with a dataset exhibiting oscillations, we need to extract the sequence of maxima and minima (together called extrema) from a time series dataset. We briefly discuss the methodology from [56], in which for each extremum, a time interval is assigned representing experimental uncertainty in the timing of extrema. The idea is that at a specified noise level, an extremum can occur anywhere within the assigned time interval, perhaps due to sparse temporal sampling and/or measurement error. Since the time intervals assigned to different extrema may overlap, the ordering of some extrema with respect to others is indeterminate. However, within any single time series the order of extrema is known.

Fig. 16 schematically shows the process for assigning a time interval for an extremum. The example shown is the gene expression level for Ndd1 in the WT dataset. The blue line is the collected data (with linear

interpolation) and the orange and green lines are the original curve $\pm 10\%$ of the difference between the global maximum and global minimum. This is called a 10% noise level of the Ndd1 data. The purpose of choosing a noise level is to smooth out small, spurious extrema and also account for imprecision in the timing of large extrema.

The example computation in Fig. 16 assigns a time interval to the second minimum in the time series which occurs at 158 min, called an extremal interval, by locating a region of the graph around the minimum that is below the + 10% curve and above the -10% curve. It was proven in [57] that any perturbation of the original data that is bounded within $\pm\epsilon$ curves for some noise level ϵ is guaranteed to have a minimum located within the extremal interval. That is, the existence of the minimum is robust within the extremal interval. The extremal interval for a maximum is recovered in a similar manner.

This procedure is repeated for every local extremum in every time series, where an extremum is identified as either an endpoint of the time series, or a point such that the nearest neighbors on either side are both higher or both lower than the point in question. The end result is a collection of time intervals, and the intervals across time series are compared to see if they are overlapping. When two intervals overlap, it is possible for the extrema associated to the intervals to occur



DSGRN parameter FP :

$$X : l_{X,Y}l_{X,Z} < \theta_{Y,X} < h_{X,Y}l_{X,Z}, l_{X,Y}h_{X,Z} < h_{X,Y}h_{X,Z}$$

$$Y : l_{Y,X} < \theta_{Z,Y} < \theta_{X,Y} < h_{Y,X}$$

$$Z : l_{Z,Y} < h_{Z,Y} < \theta_{X,Z}$$

Fig. 15. An example STG (Left) and MG (Right) computed from the given DSGRN parameter (Bottom) for the 3-node network in Fig. 8. The STG exhibits an unstable partial cycle and a fixed point.

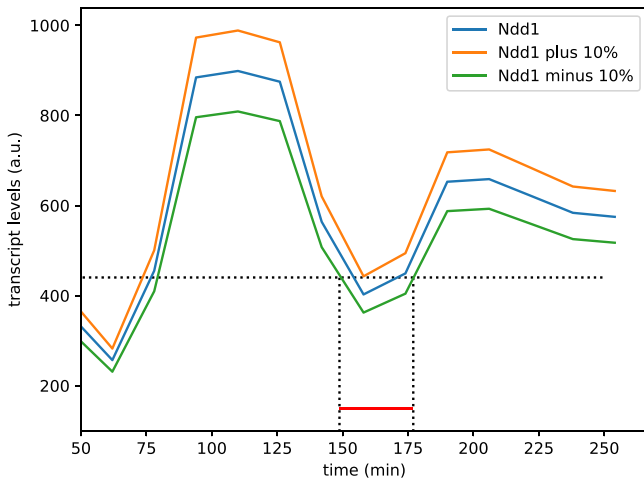


Fig. 16. The time interval corresponding to a 10% noise level at the second minimum within the wild-type Ndd1 time series. The red line indicates the size of the time interval.

in either order in time. In other words, the ordering of the extrema is indeterminate and the extrema are called incomparable. However, when intervals are non-overlapping, then the order between the two extrema is known, and they are called comparable.

A *pattern diagram* is a discrete, graphical representation of all the comparability relationships between extremal intervals. This is known mathematically as the Hasse diagram of a partially ordered set. In the most straightforward case, all extrema of a dataset are comparable and form a linear sequence of events in time. This is true for any single time series. However, for more than one time series this never happens, since the start of every time series co-occurs and the concentrations for each node are either at a maximum or a minimum.

Fig. 17 shows the pattern diagram for the WT data (Fig. 2b) for the nodes in the SCC network. The time intervals for each extremum were computed at 10% noise, as were all time series discretizations in this manuscript. Each node in the pattern diagram is a type of extremum, with nodes at the top occurring at the beginning of the time series. A directed arrow between a source node and target node means that the source extremum is known to occur earlier in time than the target extremum. If there is no path between two nodes, then the associated extrema are incomparable.

5.6. Pattern matching

Pattern diagrams are the basis for testing DSGRN network model consistency with oscillating time series data. While the technical details differ [57], a straightforward demonstration of pattern matching can be done by labeling the edges of a DSGRN state transition graph with the extrema labels that occur in the nodes of the pattern diagram. We begin our example of pattern matching by discussing how to use the pattern diagram, and then proceed to the labeling of the STG.

The key idea is that of a *linear extension* of the pattern diagram. A linear extension is any sequence of all the nodes in a pattern diagram that does not contradict any arrows in the pattern diagram. For example, in Fig. 17, the top two nodes Clb2 min and Swi5 max may occur in either order in a linear extension because they are incomparable. However, the arrow from Clb2 min to Clb2 max enforces that in any linear extension, Clb2 min occurs before Clb2 max. We claim that any of the linear extensions arising from a pattern diagram might be the “true” sequence of extrema in the biological system given the measurement error and sampling density of the data.

Continuing our example of the three-node network in Fig. 8, we have constructed two example pattern diagrams that could have arisen from two different datasets of nodes X and Y (top row of Fig. 18), with the idea of checking if a pattern match exists between either of these datasets and the unstable partial cycle in Fig. 14 (blue nodes). Via inspection of the two example pattern diagrams it can be seen which extrema are comparable and incomparable. In the first pattern diagram on the left, the node X_{min} is comparable with any other node in the pattern diagram, but X_{max} and Y_{max} are not comparable with each other. For the pattern diagram on the right, we see a shift in the ordering of the extrema where X_{max} and Y_{min} are incomparable. Each pattern diagram has two linear extensions (bottom row of Fig. 18). To construct the linear extensions for the left pattern diagram a decision is made that defines the ordering of the two incomparable extrema X_{max} and Y_{max} . Similarly for the poset on the right, the decision is between the order of Y_{min} and X_{max} .

The nodes of an STG can be labeled with information about whether a node concentration is increasing (I), decreasing (D), or both ($*$) within the corresponding domain in phase space. This information is readily extracted from the arrows in the state transition graph (and the knowledge that the dissipativity of the switching system means that the long term dynamics lies in a compact region of phase space). For example, in Fig. 15, the leftward-going arrows on the front face of the cube indicate that Y is decreasing in the domains corresponding to 111 and 110. If the Y direction contained arrows that are both increasing

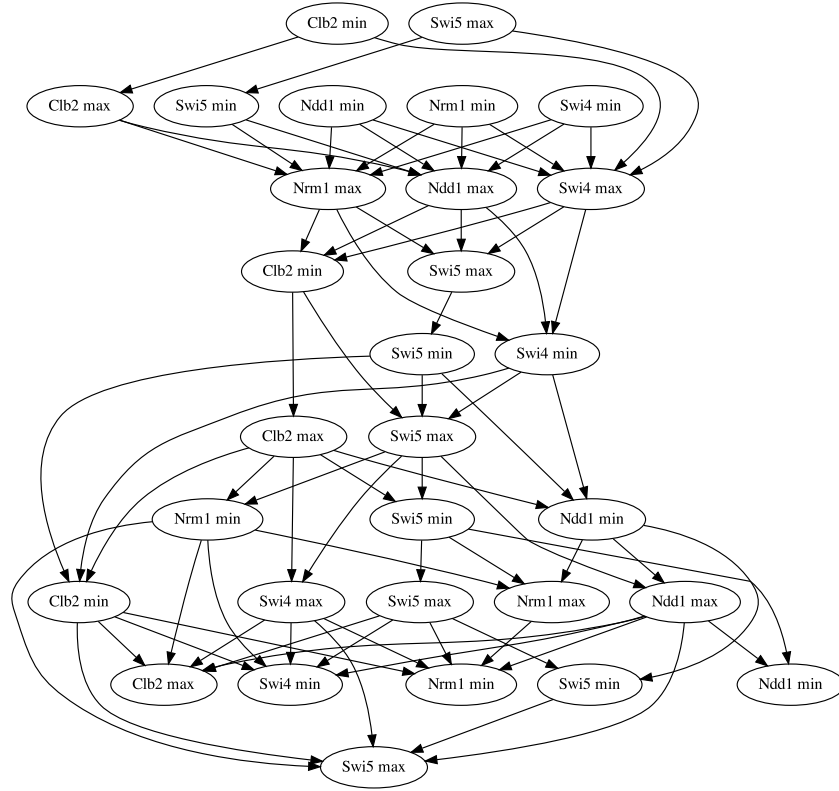


Fig. 17. The pattern diagram for the WT dataset in Fig. 2(b) at a noise level of 10%.

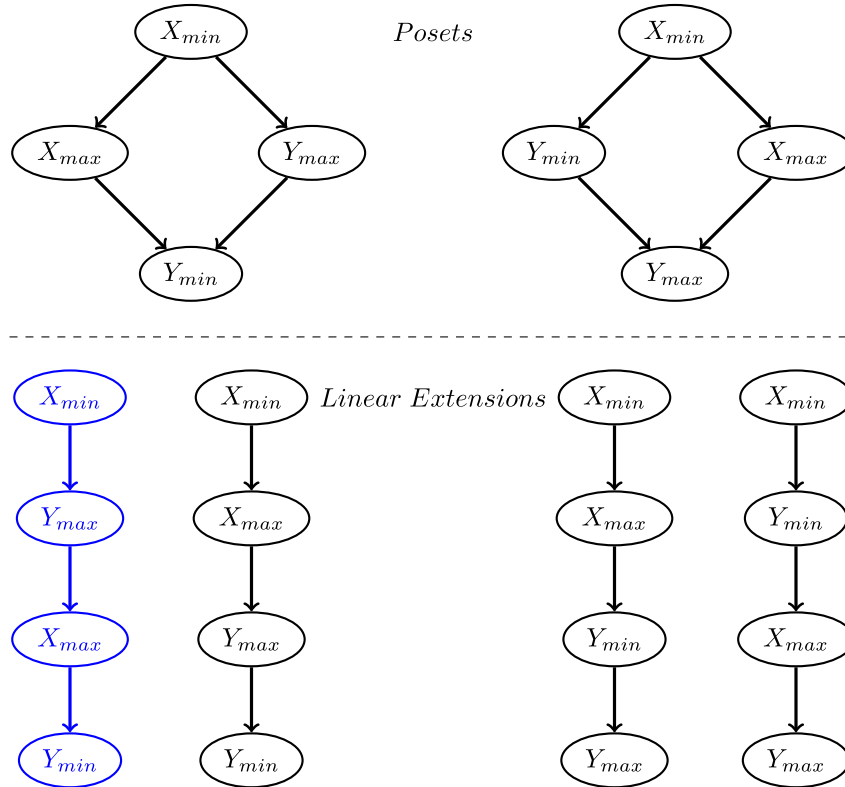


Fig. 18. Two example pattern diagrams corresponding to two hypothetical datasets (top row) and their corresponding collections of linear extensions (bottom row). The linear extension in blue is a pattern match to the unstable partial cycle in Fig. 14, as seen by the labeling of the STG in Fig. 19.

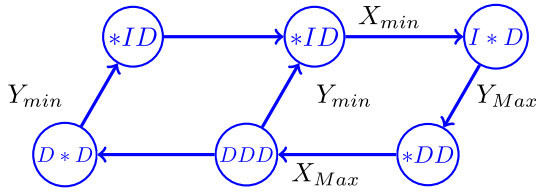


Fig. 19. The part of the STG in Fig. 14 corresponding to the unstable partial cycle. Node labels indicate whether node concentrations are increasing or decreasing within the corresponding domains in phase space. Edge labels indicate when an extremum between domains can be inferred. Notice that an extremum does not necessarily need to occur between two domains in phase space.

and decreasing at a domain, then the label on the node of the STG would be *. This latter case is possible when Y is self-regulating, or when the corresponding domain in phase space is unbounded. For a detailed description of the labeling procedure, see [57]. From this information, edge labels in the STG can be deduced that describe which extremum could occur between any two neighboring domains.

To continue our example, Fig. 19 contains the unstable PC from Fig. 14 along with the I , D , and $*$ labels on each node in the order (XYZ) . Consider the edge going from the node labeled (DDD) to the node labeled $(*ID)$. In the first domain, Y is decreasing and in the second domain, it is increasing. Therefore Y had to undergo a minimum on the edge between the two. Continuing to the node labeled $(I*D)$, we note that we are moving along the Y axis. Y does not regulate itself, and therefore we know that starting from node $(*ID)$ means Y must continue to increase. Therefore, no extremum may occur in Y on the arrow from $(*ID)$ to $(I*D)$. However, on the next edge to $(*DD)$, Y transitions from increasing to decreasing, therefore must have achieved a maximum. Similar arguments allow the assignment of the other edge labels.

To continue with pattern matching, we compare the sequence of extrema in the STG to the linear extensions in Fig. 18. It can be seen that within the STG X_{min} is followed by Y_{max} then X_{max} and finally Y_{min} . This is exactly the left-most linear extension in Fig. 18, highlighted in blue. On the other hand, looking at the rest of the linear extensions in Fig. 18, it can be seen that none of them match the path through the STG. We conclude that the three-node network model in Fig. 8 is consistent with the hypothetical dataset that generated the pattern diagram on the left, but is not consistent with the pattern diagram on the right.

This pattern matching procedure is conducted analogously for the WT and mutant datasets against the SCC network model in Section 3.

5.7. Hill models

After locating a region of interest in the TFO subnetwork parameters, we parameterized Hill models through real-valued sampling of DSGRN parameter nodes to recapitulate the experimental data. To model an activating regulation $x \rightarrow y$, we used a Hill function of the form:

$$H_{y,x}^+(x) = (h_{y,x} - l_{y,x}) \frac{x^n}{\theta_{y,x}^n + x^n} + l_{y,x}$$

where l , h , and θ are the switching system parameters introduced in Section 5.1. Similarly, a repressing regulation $x \dashv y$ is given by:

$$H_{y,x}^-(x) = (h_{y,x} - l_{y,x}) \frac{\theta_{y,x}^n}{\theta_{y,x}^n + x^n} + l_{y,x}$$

We created a Hill model of the SCC network in Fig. 1 (Right) of the following form:

$$\begin{aligned} \dot{S} &= -S + H^-(N)H^-(C)H^+(S)H^+(W) \\ \dot{N} &= -N + H^+(S) \end{aligned}$$

$$\begin{aligned} \dot{D} &= -D + H^+(S)H^+(C) \\ \dot{W} &= -W + H^-(C)H^+(D) \\ \dot{C} &= -C + H^+(N) \end{aligned}$$

where S (Swi4), N (Nrm1), D (Ndd1), W (Swi5), and C (Clb2) represent expression levels of the corresponding mRNA. The subscripts on the Hill functions have been suppressed for clarity.

The parameters for the simulations in Figs. 3, 5, and 6 are provided in a public code repository (see Data Availability). The Hill exponent n was taken to be 10 in all simulations. Time series were constrained to have a minimum absolute fold change of 2 between extrema for all oscillating variables.

CRediT authorship contribution statement

Julian Fox: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Breschne Cummins:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Robert C. Moseley:** Software, Data curation. **Marcio Gameiro:** Software, Writing – review & editing, Funding acquisition. **Steven B. Haase:** Conceptualization, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare no conflicts of interest.

Data availability

The code to generate the figures and results in this paper is publicly available at <https://github.com/julianfox8/2022-DSGRN-Phenotypes-Yeast.git>. This repository includes the jointly normalized data used in this manuscript, but does not include the original data. That is available by request from the Haase lab at Duke University.

Acknowledgments

JF was partially supported by a grant from the Undergraduate Scholars Program at Montana State University, United States. BC and SBD were partially supported by NSF TRIPODS-X, United States grant DMS-1839299 and NIH 5R01GM126555-01. RCM, United States was partially supported by NIH 5R01GM126555-01. MG was partially supported by the National Science Foundation, United States under awards DMS-1839294 and HDR TRIPODS, United States award CCF-1934924, DARPA, United States contract HR0011-16-2-0033, National Institutes of Health, United States award 5R01GM126555-01, and Air Force Office of Scientific Research, United States under award numbers FA9550-23-1-0011 and AWD00010853-MOD002, FAPESP, Brazil grant 2019/06249-7, and by CNPq, Brazil grant 309073/2019-7. The funding agencies had no role in the study design, execution, analysis, interpretation of results, or decision to submit.

References

- [1] F.-G. Wieland, A.L. Hauber, M. Rosenblatt, C. Tönsing, J. Timmer, On structural and practical identifiability, *Curr. Opin. Syst. Biol.* 25 (2021) 60–69, <http://dx.doi.org/10.1016/j.coisb.2021.03.005>.
- [2] A. Hershko, D. Ganoth, V. Sudakin, A. Dahan, L.H. Cohen, F.C. Luca, J.V. Ruderman, E. Eytan, Components of a system that ligates cyclin to ubiquitin and their regulation by the protein kinase cdc2, *J. Biol. Chem.* 269 (7) (1994) 4940–4946.
- [3] T. Hunt, Maturation promoting factor, cyclin and the control of M-phase, *Curr. Opin. Cell Biol.* 1 (1989) 268–274.
- [4] T. Hunt, Cell cycle gets more cyclins, *Nature* 350 (1991) 462–463.
- [5] A. Goldbeter, A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase, *Proc. Natl. Acad. Sci. USA* 88 (20) (1991) 9107–9111.

- [6] J. Minshull, J. Pines, N. Standart, L. Stewart, S. Mackie, A. Colman, J. Blow, M. Wu, J. Ruderman, T. Hunt, Protein synthesis, proteolysis, and the control of cell division in early embryos: Does the synthesis and destruction of cyclin comprise the cytoplasmic oscillator? in: D. Beach, C. Basilio, J. Newport (Eds.), *Cell Cycle Control in Eukaryotes*, Cold Spring Harbor Laboratory, N.Y., 1988.
- [7] L.H. Hartwell, J. Culotti, J.R. Pringle, B.J. Reid, Genetic control of the cell division cycle in yeast, *Science* 183 (1974) 46–51.
- [8] L.H. Hartwell, R.K. Mortimer, J. Culotti, M. Culotti, Genetic control of the cell division cycle in yeast: V. Genetic analysis of cdc mutants, *Genetics* 74 (1973) 267–286.
- [9] P. Nurse, Genetic control of cell size at cell division in yeast, *Nature* 256 (1975) 547–551.
- [10] P. Nurse, P. Thuriaux, K. Nasmyth, Genetic control of the cell division cycle in fission yeast *schizosaccharomyces pombe*, *Molec. Gen. Genet.* 146 (1976) 167–178.
- [11] S.B. Haase, S.I. Reed, Evidence that a free-running oscillator drives G1 events in the budding yeast cell cycle, *Nature* 401 (6751) (1999) 394–397, <http://dx.doi.org/10.1038/43927>.
- [12] T.I. Lee, N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, E. Volkert, D.K. Gifford, R.A. Young, Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science* 298 (5594) (2002) 799–804.
- [13] D.A. Orlando, C.Y. Lin, A. Bernard, J.Y. Wang, J.E. Socolar, E.S. Iversen, A.J. Hartemink, S.B. Haase, Global control of cell-cycle transcription by coupled CDK and network oscillators, *Nature* 453 (7197) (2008) 944–947.
- [14] T. Pramila, W. Wu, S. Miles, W.S. Noble, L.L. Breeden, The forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle, *Genes. Dev.* 20 (16) (2006) 2266–2278.
- [15] I. Simon, J. Barnett, N. Hannett, C.T. Harbison, N.J. Rinaldi, T.L. Volkert, J.J. Wyrick, J. Zeitlinger, D.K. Gifford, T.S. Jaakkola, R.A. Young, Serial regulation of transcriptional regulators in the yeast cell cycle, *Cell* 106 (6) (2001) 697–708.
- [16] K.C. Chen, A. Csikasz-Nagy, B. Gyorfy, J. Val, B. Novak, J.J. Tyson, Kinetic analysis of a molecular model of the budding yeast cell cycle, *Mol. Biol. Cell* 11 (1) (2000) 369–391.
- [17] C.Y. Cho, F.C. Motta, C.M. Kelliher, A. Deckard, S.B. Haase, Reconciling conflicting models for global control of cell-cycle transcription, *Cell Cycle* 16 (20) (2017) 1965–1978, <http://dx.doi.org/10.1080/15384101.2017.1367073>.
- [18] S.L. Bristow, A.R. Leman, L.A. Simmons Kovacs, A. Deckard, J. Harer, S.B. Haase, Checkpoints couple transcription network oscillator dynamics to cell-cycle progression, *Genome. Biol.* 15 (9) (2014) 446, <http://dx.doi.org/10.1186/s13059-014-0446-7>.
- [19] C.Y. Cho, C.M. Kelliher, S.B. Haase, The cell-cycle transcriptional network generates and transmits a pulse of transcription once each cell cycle, *Cell Cycle* 18 (4) (2019) 363–378, <http://dx.doi.org/10.1080/15384101.2019.1570655>.
- [20] L.A. Simmons Kovacs, M.B. Mayhew, D.A. Orlando, Y. Jin, Q. Li, C. Huang, S.I. Reed, S. Mukherjee, S.B. Haase, Cyclin-dependent kinases are regulators and effectors of oscillations driven by a transcription factor network, *Mol. Cell* 45 (5) (2012) 669–679.
- [21] S.J. Rahi, K. Pecani, A. Ondracka, C. Oikonomou, F.R. Cross, The CDK-APC/C oscillator predominantly entrains periodic cell-cycle transcription, *Cell* 165 (2) (2016) 475–487, <http://dx.doi.org/10.1016/j.cell.2016.02.060>.
- [22] S.E. Brantley, S. Di Talia, Cell cycle control during early embryogenesis, *Development* 148 (13) (2021) <http://dx.doi.org/10.1242/dev.193128>.
- [23] B.L. Powers, M.C. Hall, Re-examining the role of Cdc14 phosphatase in reversal of Cdk phosphorylation during mitotic exit, *J. Cell Sci.* 130 (16) (2017) 2673–2681.
- [24] V. Sevim, X. Gong, J.E. Socolar, Reliability of transcriptional cycles and the yeast cell-cycle oscillator, *PLoS Comput. Biol.* 6 (7) (2010) e1000842, <http://dx.doi.org/10.1371/journal.pcbi.1000842>.
- [25] P. Crawford-Kahrl, B. Cummins, T. Gedeon, Joint realizability of monotone boolean functions, *Theoret. Comput. Sci.* 922 (2022) 447–474, <http://dx.doi.org/10.1016/j.tcs.2022.04.045>.
- [26] B. Cummins, T. Gedeon, S. Harker, K. Mischaikow, DSGRN: Examining the dynamics of families of logical models, *Front. Physiol.* 9 (2018) 549.
- [27] B. Cummins, T. Gedeon, S. Harker, K. Mischaikow, K. Mok, Combinatorial representation of parameter space for switching networks, *SIAM J. Appl. Dyn. Syst.* 15 (4) (2016) 2176–2212.
- [28] M. Gameiro, Dynamic signatures generated by regulatory networks, 2018, GitHub repository, <https://github.com/marciogameiro/DSGRN>.
- [29] Y. Xin, B. Cummins, T. Gedeon, Multistability in the epithelial-mesenchymal transition network, *BMC Bioinformatics* 21 (1) (2020) 71.
- [30] T. Gedeon, B. Cummins, S. Harker, K. Mischaikow, Identifying robust hysteresis in networks, *PLoS Comput. Biol.* 14 (4) (2018) e1006121.
- [31] G. Bernot, F. Cassez, J.-P. Comet, F. Delaplace, C. Müller, O. Roux, Semantics of biological regulatory networks, *Electron. Notes Theor. Comput. Sci.* 180 (3) (2007) 3–14, <http://dx.doi.org/10.1016/j.entcs.2004.01.038>, Proceedings of the First Workshop on Concurrent Models in Molecular Biology (BioConcur 2003).
- [32] T. Gedeon, S. Harker, H. Kokubu, K. Mischaikow, H. Oka, Global dynamics for steep nonlinearities in two dimensions, *Physica D* 339 (2017) 18–38, <http://dx.doi.org/10.1016/j.physd.2016.08.006>.
- [33] K. Hari, W. Duncan, M.A. Ibrahim, M.K. Jolly, B. Cummins, T. Gedeon, Assessing biological network dynamics: Comparing numerical simulations with analytical decomposition of parameter space, *npj Syst. Biol. Appl.* 9 (1) (2023) 29, <http://dx.doi.org/10.1038/s41540-023-00289-2>.
- [34] B. Huang, D. Jia, J. Feng, H. Levine, J.N. Onuchic, M. Lu, RACIPE: A computational tool for modeling gene regulatory circuits using randomization, *BMC Syst. Biol.* 12 (1) (2018) 1–12, <http://dx.doi.org/10.1186/s12918-018-0594-6>.
- [35] J. Krumsiek, S. Pölsterl, D.M. Wittmann, F.J. Theis, Odey-from discrete to continuous models, *BMC Bioinform.* 11 (2010) 1–10, <http://dx.doi.org/10.1186/1471-2105-11-233>.
- [36] D.M. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D.A. Lauffenburger, S. Klamt, F.J. Theis, Transforming Boolean models to continuous models: Methodology and application to T-cell receptor signaling, *BMC Syst. Biol.* 3 (1) (2009) 1–21.
- [37] M. Chaves, M. Preto, Hierarchy of models: From qualitative to quantitative analysis of circadian rhythms in cyanobacteria, *Chaos* 23 (2) (2013) <http://dx.doi.org/10.1063/1.4810922>.
- [38] A. Saadatpour, R. Albert, A comparative study of qualitative and quantitative dynamic models of biological regulatory networks, *EPJ Nonlinear Biomed. Phys.* 4 (2016) 1–13.
- [39] F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, The yeast cell-cycle network is robustly designed, *Proc. Natl. Acad. Sci.* 101 (14) (2004) 4781–4786, <http://dx.doi.org/10.1073/pnas.0305937101>.
- [40] D.O. Morgan, *The Cell Cycle: Principles of Control*, in: (Primers in Biology), New Science Press Ltd, London, 2007, p. 297.
- [41] B. Cummins, M. Gameiro, T. Gedeon, S. Kopley, K. Mischaikow, L. Zhang, Extending combinatorial regulatory network modeling to include activity control and decay modulation, *SIAM J. Appl. Dyn. Syst.* 21 (3) (2022) 2096–2125.
- [42] C.M. Kelliher, M.W. Foster, F.C. Motta, A. Deckard, E.J. Soderblom, M.A. Moseley, S.B. Haase, Layers of regulation of cell-cycle gene expression in the budding yeast *saccharomyces cerevisiae*, *Mol. Biol. Cell* 29 (22) (2018) 2644–2655, <http://dx.doi.org/10.1091/mbc.E18-04-0255>.
- [43] M. Bäumer, G. Braus, S. Irniger, Two different modes of cyclin Clb2 proteolysis during mitosis in *saccharomyces cerevisiae*, *FEBS Lett.* 468 (2–3) (2000) 142–148.
- [44] R.A. Sclafani, W.L. Fangman, Yeast gene CDC8 encodes thymidylate kinase and is complemented by herpes thymidine kinase gene TK, *Proc. Natl. Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.81.18.5821>.
- [45] B. de Oliveira, M. Francisco, M. Harris, P. Brazauskas, R. de Bruin, M. Smolka, Linking DNA replication checkpoint to MBF cell-cycle transcription reveals a distinct class of G1/S genes, *EMBO J.* 31 (7) (2012) 1798–1810, <http://dx.doi.org/10.1038/emboj.2012.27>.
- [46] A. Travesa, D. Kuo, R.A.M. de Bruin, T.I. Kalashnikova, M. Guaderrama, K. Thai, A. Aslanian, M.B. Smolka, J.R. Yates III, T. Ideker, C. Wittenberg, DNA replication stress differentially regulates G1/S genes via rad53-dependent inactivation of Nrm1, *EMBO J.* 31 (7) (2012) 1811–1822, <http://dx.doi.org/10.1038/emboj.2012.28>.
- [47] J. Nilsson, M. Yekezare, J. Minshull, J. Pines, The APC/C maintains the spindle assembly checkpoint by targeting Cdc20 for destruction, *Nat. Cell Biol.* 10 (2008) 1411–1420.
- [48] R. Wang, J.L. Burton, M.J. Solomon, Transcriptional and post-transcriptional regulation of Cdc20 during the spindle assembly checkpoint in *S. cerevisiae*, *Cell. Signal.* 33 (2017) 41–48.
- [49] B. Ibrahim, Spindle assembly checkpoint is sufficient for complete Cdc20 sequestering in mitotic control, *Comput. Struct. Biotechnol. J.* 13 (2015) 320–328.
- [50] E. Fox, B. Cummins, W. Duncan, T. Gedeon, Modeling transport regulation in gene regulatory networks, *Bull. Math. Biol.* 84 (8) (2022) 1–42, <http://dx.doi.org/10.1007/s11538-022-01035-1>.
- [51] R. Thomas, Regulatory networks seen as asynchronous automata: A logical description, *J. Theoret. Biol.* 153 (1991) 1–23, [http://dx.doi.org/10.1016/S0022-5193\(05\)80350-9](http://dx.doi.org/10.1016/S0022-5193(05)80350-9).
- [52] S.R. Veflingstad, E. Plahte, Analysis of gene regulatory network models with graded and binary transcriptional responses, *Biosystems* 90 (2) (2007) 323–339.
- [53] H. de Jong, Modeling and simulation of genetic regulatory systems: A literature review, *J. Comput. Biol.* 9 (2002) 67–103, <http://dx.doi.org/10.1089/1066527025283208>.
- [54] R. Edwards, Chaos in neural and gene networks with hard switching, *Diff. Eq. Dyn. Sys.* (9) (2001) 187–220.
- [55] L. Ironi, L. Panzeri, E. Plahte, V. Simoncini, Dynamics of actively regulated gene networks, *Physica D* 240 (8) (2011) 779–794, <http://dx.doi.org/10.1016/j.physd.2010.12.010>.
- [56] E. Berry, B. Cummins, R.R. Nerem, L.M. Smith, S.B. Haase, T. Gedeon, Using extremal events to characterize noisy time series, *J. Math. Biol.* 80 (5) (2020) 1523–1557.
- [57] B. Cummins, T. Gedeon, S. Harker, K. Mischaikow, Model rejection and parameter reduction via time series, *SIAM J. Appl. Dyn. Syst.* 17 (2) (2018) 1589–1616.