Cell cycle time series gene expression data encoded as cyclic attractors in Hopfield systems

Anthony Szedlak[1], Spencer Sims[1], Nicholas Smith[2], Giovanni Paternostro[3], Carlo Piermarocchi[1,*],

1 Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA
2 Salgomed Inc., Del Mar, CA, USA
3 Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA
* E-mail: Corresponding carlo@pa.msu.edu

Abstract

Modern time series gene expression and other omics data sets have enabled unprecedented resolution of the dynamics of cellular processes such as cell cycle and response to pharmaceutical compounds. In anticipation of the proliferation of time series data sets in the near future, we use the Hopfield model, a recurrent neural network based on spin glasses, to model the dynamics of cell cycle in HeLa (human cervical cancer) and *S. cerevisiae* cells. We study some of the rich dynamical properties of these cyclic Hopfield systems, including the ability of populations of simulated cells to recreate experimental expression data and the effects of noise on the dynamics. Next, we use a genetic algorithm to identify sets of genes which, when selectively inhibited by local external fields representing gene silencing compounds such as kinase inhibitors, disrupt the encoded cell cycle. We find, for example, that inhibiting the set of four kinases *BRD4*, *MAPK1*, *NEK7*, and *YES1* in HeLa cells causes simulated cells to accumulate in the M phase. Finally, we suggest possible improvements and extensions to our model.

Author Summary

Cell cycle – the process in which a parent cell replicates its DNA and divides into two daughter cells – is an upregulated process in many forms of cancer. Identifying gene inhibition targets to regulate cell cycle is important to the development of effective therapies. Although modern high throughput techniques offer unprecedented resolution of the molecular details of biological processes like cell cycle, analyzing the vast quantities of the resulting experimental data and extracting actionable information remains a formidable task. Here, we create a dynamical model of the process of cell cycle using the Hopfield model (a type of recurrent neural network) and gene expression data from human cervical cancer cells and yeast cells. We find that the model recreates the oscillations observed in experimental data. Tuning the level of noise (representing the inherent randomness in gene expression and regulation) to the "edge of chaos" is crucial for the proper behavior of the system. We then use this model to identify potential gene targets for disrupting the process of cell cycle. This method could be applied to other time series data sets and used to predict the effects of untested targeted perturbations.

Introduction

Originally proposed by Conrad Waddington in the 1950s [1] and Stuart Kauffman in the 1970s [2], analysis of biological processes such as cellular differentiation and cancer development using attractor models – dynamical systems whose configurations tend to

evolve toward particular sets of states – has gained significant traction over the past decade [3–12]. One such attractor model, the Hopfield model [13], is a type of recurrent artificial neural network based on spin glasses. It was designed with the ability to recall a host of memorized patterns from noisy or partial input information by mapping data directly to attractor states. A great deal of analytical and numerical work has been devoted to understanding the statistical properties of the Hopfield model, including its storage capacity [14], correlated patterns [15], spurious attractors [16], asymmetric connections [17], embedded cycles [18], and complex transition landscapes [19]. Due to its prescriptive, data-driven design, the Hopfield model has been applied in a variety of fields including image recognition [20, 21] and the clustering of gene expression data [22]. It has also been used to directly model the dynamics of cellular differentiation and stem cell reprogramming [23], as well as targeted inhibition of genes in cancer gene regulatory networks [24].

Techniques for measuring large scale omics data, particularly transcriptomic data from microarrays and RNA sequencing (RNA-seq), have become standard, indispensable tools for measuring the states of complex biological systems [25–27]. However, analysis of the sheer variety and vast quantities of data these techniques produce requires the development of new mathematical tools. Inference and topological analysis of gene regulatory networks has garnered much attention as a method for distilling meaningful information from large datasets [28–34]. But because life is a non-equilibrium phenomenon that can only be truly understood at the dynamical level, there is a growing need to develop new methods for analyzing time series data. As experimental methods continue to improve, more and more high-resolution time series omics and even multi-omics [35] data sets will inevitably become available. Here, we demonstrate that time series omics data (in this case, transcriptomic data) representing cyclic biological processes can be encoded in Hopfield systems, providing a new model for analyzing the dynamics of, and exploring effects of perturbations to, such systems.

The dynamics of cell cycle (CC) – the process in which a parent cell replicates its DNA and divides into two daughter cells – is both scientifically interesting and therapeutically important. Even relatively simple simulated systems such as an isolated, positively self-regulating gene subject to noise can exhibit rich dynamical behavior [36]; but like many biological processes, the proper functioning of CC requires the decentralized, coordinated action of hundreds of genes. CC thus provides researchers with a convenient case study of self-organization in a noisy environment. CC is also an upregulated process in many forms of cancer [37–40], and control of CC using pharmaceutical compounds such as kinase inhibitors is a critical goal in cancer research. The combinatorics of selectively inhibiting sets of genes makes exhaustive experimental searches difficult or impractical [41]. However, network-based mathematical models such as the one presented here enable researchers to examine the effects of perturbations to complex systems [42, 43] by testing potential inhibition targets *in silico*. The efficacy of these predictions can then be experimentally validated or invalidated, providing new information and insights to further refine models.

The remainder of this article is structured as follows. In the Models section we first discuss how periodic genes were identified in the time series gene expression data sets, and how Boolean attractors were extracted from the continuous data (explained in greater detail in the Methods section). We then introduce the Hopfield model and discuss the specific form of the coupling matrix used in this application. We discuss how to interpret the results of Hopfield simulations in the context of gene expression and cells. We also explain the objective function used by the genetic algorithm to identify potential inhibition targets, designed with the intention of disrupting CC. In the Results and Discussion section, we show that this model qualitatively recreates experimental gene expression data, and we demonstrate and analyze some dynamical properties of the
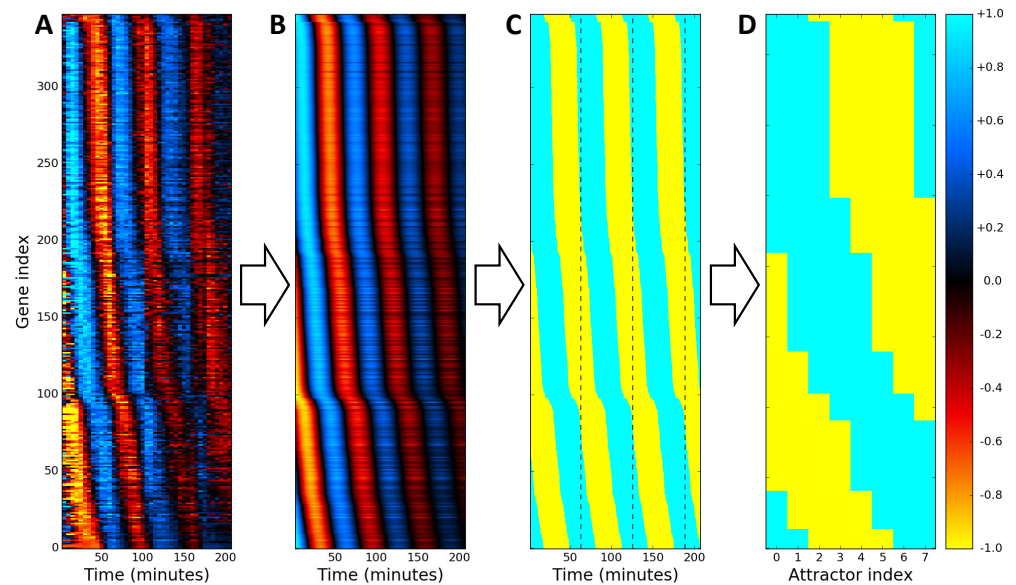
Figure 1. Obtaining attractors from expression data. (A) Heat map of the expression of all detected periodic genes from [44] sorted by their fitted phases. (B) Fitted gene expression. (C) Boolean form of fitted expression, separated into periods by dashed black lines. (D) Final set of $p = 8$ attractors taken from one period.

delayed Hopfield model, including the role played by noise. We include supplementary videos to emphasize the dynamical nature of this model. Optimal control fields for both unconstrained searches (in which any gene may be inhibited) and searches constrained to kinases in HeLa cells are discussed. Finally, we recap our results and suggest possible improvements and generalizations to our methods in the Conclusions section.

Models

Periodic gene selection

Microarray and RNA-seq time series data sets were obtained from Eser et al. (*S. cerevisiae*) [44] and Dominguez et al. (HeLa, human cervical cancer) [45]. For consistency and due to its higher resolution, the *S. cerevisiae* data set was chosen to produce all images and movies in this article, but both data sets were analyzed. In order to encode these CC data sets into the Hopfield model, periodic genes needed to be identified, their frequencies and phases computed, and their expression converted from continuous to Boolean form. As detailed in the Methods section, decaying sinusoids were fitted to the trajectory of each gene i, and genes with sufficiently high quality fits were kept. This resulted in 379 periodic genes in *S. cerevisiae* and 519 periodic genes in HeLa cells. Figure 1A shows a heat map of the expression of all periodic genes detected in the Eser data set sorted by their fitted phases, and Figure 1B shows the same genes with the fitted expression curves. These fitted curves were converted from continuous values $x_i(t) \geq 0$ to Boolean values $\xi_i(t) = \pm 1$ (over/underexpressed) as shown in Figure 1C. Finally, one CC period was divided into eight uniformly spaced states $\{\xi_i^\mu\} = \{\xi_i^0, \xi_i^1, \ldots, \xi_i^7\}$ with $\xi_i^\mu = \pm 1$. These states, shown in Figure 1D, were used as the eight attractor patterns in the Hopfield model.

The Hopfield model

The Hopfield model [13] is an Ising model whose configuration is defined by N spins $\sigma_i(t)$ at integer time t. The state of each node (gene) takes one of two values, $\sigma_i(t) = \pm 1$ (over/underexpressed). The coupling matrix J_{ij} defines the strength and sign of the signal sent from node j to node i, and its construction is discussed in the following subsection. The total field at node i at time t is given by

$$h_i(t) = \sum_j J_{ij}\sigma_j(t) + h_i^{\text{ext}} \ , \tag{1}$$

where $\sum_j J_{ij}\sigma_j(t)$ is the internal field at node i due to its coupling with all nodes j and h_i^{ext} is an optional external field applied to node i representing the action of therapeutic compounds, e.g. kinase inhibitors. The dynamical update rule is given by

$$\sigma_i(t+1) = \begin{cases} +1 & \text{with probability } (1 + e^{-2h_i(t)/T})^{-1} \\ -1 & \text{otherwise} \end{cases} \ , \tag{2}$$

where the factor of 2 in the exponent is conventional and T is an effective temperature representing the level of noise (not a physical temperature). Biologically, this noise represents the effects of all kinds of biochemical fluctuations present in cells. Note that for $h_i(t) \to \pm\infty$, $\sigma_i(t+1) = \pm 1$; for $T \to \infty$, $\sigma_i(t+1) = \pm 1$ with equal probability; and for $T \to 0$, $\sigma_i(t+1) = \text{sign}(h_i(t))$.

The update rule from Eq. 2 may be implemented in various ways. The synchronous scheme updates the state of all nodes in the system at every time step, but this is sensible only if the simulated system has a central pacemaker coordinating the activity of all nodes. A more appropriate choice for decentralized systems like gene regulatory networks is the asynchronous scheme in which the state of a randomly chosen subset of nodes is updated at each time step. Here, we use the asynchronous scheme with update probability 0.2 for each node.

Coupling matrix

In the canonical Hopfield model, the coupling matrix is constructed to store a set of p linearly independent (i.e. distinct) Boolean patterns $\xi_i^\mu = \pm 1$ as point attractors, where $i = 0, 1, \ldots, N-1$ is the node index and $\mu = 0, 1, \ldots, p-1$ is the pattern index. The point attractor coupling matrix J_{ij}' is given by

$$J_{ij}' = \frac{1}{N} \sum_{\mu\nu} \xi_i^\mu \left(Q^{-1}\right)_{\mu\nu} \xi_j^\nu \ , \tag{3}$$

where [15, 46]

$$Q_{\mu\nu} = \frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu \ . \tag{4}$$

With this coupling matrix and $T = 0$, if at some time t the configuration is given by $\sigma_i(t) = \xi_i^\mu + \delta_i$ for a small perturbation δ_i, then $\lim_{t\to\infty} \sigma_i(t) = \xi_i^\mu$. Note that this formulation means $\pm\xi_i^\mu$ are both attractors of the system.

A simple modification [19] to Eq. 3 produces a cyclic attractor coupling matrix $\widetilde{J}_{ij}$, constructed according to

$$\widetilde{J}_{ij} = \frac{1}{N} \sum_{\mu\nu} \xi_i^{\text{mod}_p(\mu+1)} \left(Q^{-1}\right)_{\mu\nu} \xi_j^\nu \ . \tag{5}$$

At $T = 0$, this coupling matrix cyclically maps through the sequence of p patterns

$$\xi_i^0 \to \xi_i^1 \to \ldots \to \xi_i^{p-2} \to \xi_i^{p-1} \to \xi_i^0 \to \xi_i^1 \to \ldots \tag{6}$$

or their negatives. For the remainder of this article, all attractor indexing is understood to be modulo p.

A delayed cyclic Hopfield model may be constructed by combining the point and cyclic attractor matrices into one coupling matrix,

$$J_{ij}(\lambda) = (1 - \lambda)J'_{ij} + \lambda \widetilde{J}_{ij} , \tag{7}$$

for an adjustable transition strength parameter λ with $0 \leq \lambda \leq 1$. If $\sigma(t) = \xi_i^\mu$, $\lambda \ll 1$, and $T = 0$, the point attractor term dominates and $\sigma_i(t) = \sigma(t + \Delta t)$ for all $\Delta t = 1, 2, \ldots$. If $T > 0$, however, stochastic fluctuations eventually push the configuration out of the basin of attraction of the μth attractor and into the $(\mu + 1)^{\text{th}}$ basin, then eventually to the $(\mu + 2)^{\text{th}}$ basin, and so on. The dynamics of the delayed cyclic Hopfield model are thus governed by noise-induced transitions.

Due to the sinusoidal nature of the gene expression in these CC data sets, however, the attractors are structured such that $\xi_i^\mu = -\xi_i^{\mu+4}$, making $Q_{\mu\nu}$ rank deficient and thus noninvertible. Because the definition of J_{ij} automatically guarantees that if any sequence $\{+\xi_i^\mu\}$ is an attractor, then $\{-\xi_i^\mu\}$ is also an attractor, encoding the sequence

$$\xi_i^0 \to \xi_i^1 \to \xi_i^2 \to \xi_i^3 \to \xi_i^4 \left(= -\xi_i^0\right) \tag{8}$$

automatically encodes the sequence

$$\xi_i^4 \to \xi_i^5 \to \xi_i^6 \to \xi_i^7 \to \xi_i^0 \left(= -\xi_i^4\right) . \tag{9}$$

In this special case of sinusoidal trajectories, the limits of summation in Eqs. 3-5 need only run over the first four indices, $\mu = 0, 1, 2, 3$.

Finally, to reflect the fact that real gene regulatory networks are sparse, weak edges were removed by setting all elements of the coupling matrix with $|J_{ij}| < \text{median}(|J|)$ to zero, where $|J|$ is element-wise absolute value.

Biological interpretation of the dynamics

Extracting biological meaning from this model requires defining some convenient coarse-grained quantities. The overlap of the state vector $\sigma_i(t)$ with the μth pattern is given by

$$m^\mu(t) = \frac{1}{N} \sum_i \sigma_i(t)\xi_i^\mu , \tag{10}$$

where $-1 \leq m^\mu(t) \leq +1$. The overlap measures the similarity between the (discretized) experimental and simulated gene expression profiles, and $m^\mu(t) = +1$ means there is perfect agreement between the simulated cell's expression and pattern μ.

A single configuration vector $\sigma_i(t)$ represents the expression profile of a single cell. For many cells κ, let $\sigma_{ik}(t)$ be the expression of gene i in cell k. Define

$$m_k^\mu(t) = \frac{1}{N} \sum_i \sigma_{ik}(t)\xi_i^\mu \tag{11}$$

as the overlap of cell k with attractor μ. Because the microarray and RNA-seq data used here report the gene expression averaged over many cells, it is appropriate to define the population-averaged (i.e. ensemble-averaged) expression,

$$\langle \sigma_i(t) \rangle_{\mathcal{K}} = \frac{1}{\kappa} \sum_{k=0}^{\kappa-1} \sigma_{ik}(t) , \tag{12}$$

which has $-1 \leq \langle \sigma_i(t) \rangle_{\mathcal{K}} \leq +1$. 145

Rather than work with a continuous vector quantity like $m_k^{\mu}(t)$, each cell can simply 146
be identified as being in a discrete phenotypic state at any given time. Define the state 147
of cell k as 148

$$s_k(t) = \arg\max_{\mu} m_k^{\mu}(t) \ , \tag{13}$$

i.e. the index of the attractor with maximum overlap, which may be interpreted as cell 149
k's phenotype. To better understand population-level dynamics, define the discrete 150
probability distribution $P_{\mu}(t)$ as the fraction of κ cells with $s_k(t) = \mu$; that is, $P_{\mu}(t)$ is 151
the probability that a randomly chosen cell is in state μ at time t. Finally, define the 152
time-averaged distribution of states as 153

$$\langle P_{\mu} \rangle_{\mathcal{T}} = \frac{1}{\tau} \sum_{t=0}^{\tau-1} P_{\mu}(t) \tag{14}$$

for a window of time τ. 154

For each data set, T and λ were tuned to the "edge of chaos" [47] such that the 155
cyclic attractor was preserved and the time between transitions was approximately 156
constant, but the system was sensitive enough to perturbations that some targeted 157
inhibitions produced noticeable changes in $\langle P_{\mu} \rangle_{\mathcal{T}}$. See S1 Table for a list of parameters 158
used for each data set. 159

Gene inhibition optimization 160

In this application, the goal is to identify perturbations that halt or retard the encoded 161
cyclic attractor. A standard genetic algorithm (GA; explained in S1 Text) was 162
employed to identify an optimal control field h_i^{opt} that maximized a given objective 163
function $f(h_i^{\text{ext}})$, 164

$$h_i^{\text{opt}} = \arg\max_{h_i^{\text{ext}}} f\left(h_i^{\text{ext}}\right) \ , \tag{15}$$

where h_i^{ext} is the control vector given by 165

$$h_i^{\text{ext}} = \begin{cases} -\infty & \text{if gene } i \text{ is targeted} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

for a fixed number of targets (nonzero elements) n_{targ}. Only negative control fields are 166
used here to simulate the effects of targeted gene inhibition from pharmaceutical 167
compounds. The objective function used here is $\langle P_{\mu} \rangle_{\mathcal{T}}$, meaning that the optimal 168
control field maximizes the time-averaged number of cells occupying a particular 169
attractor state μ. This search was conducted across all attractors μ to determine the 170
controllability of each attractor state. 171

Results and Discussion 172

Note: The supporting information can be downloaded from this temporary Dropbox 173
link: http://bit.ly/2tEekWK. For convenience, the four supplementary videos can 174
also be viewed using the following unlisted YouTube links: 175

1. S1 Video: https://youtu.be/LOUjRftAeYM 176
2. S2 Video: https://youtu.be/pfGbla_LeGI 177
3. S3 Video: https://youtu.be/pfatwL7TusQ 178
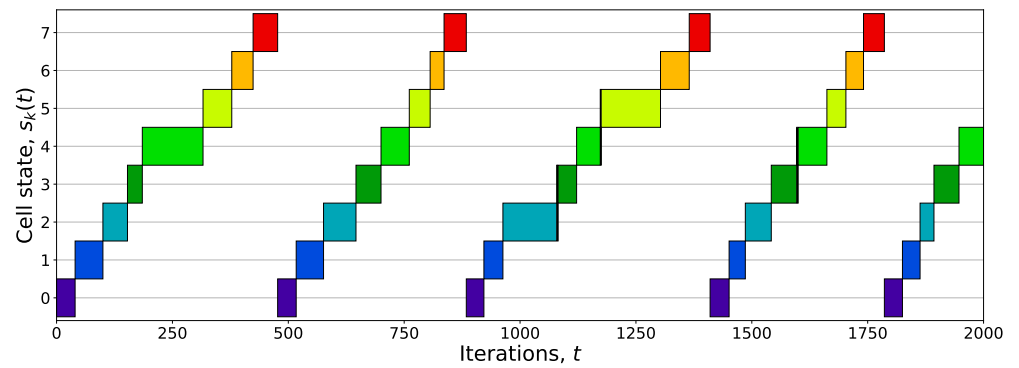4. S4 Video: https://youtu.be/RT7uNAGDcyA 179

Figure 2. Unperturbed cell state versus time. Boxes indicate $s_k(t)$, i.e. the index of the attractor with maximum overlap at time t. The system began with the configuration $\sigma_i(0) = \xi_i^0$ and was allowed to evolve according to the Hopfield signaling rules with zero external field, mapping cyclically through the set of eight attractors. The pattern and cycle durations vary due to the system's stochasticity.

Dynamical behavior

Figure 2 shows the time evolution of $s_k(t)$ for a single simulated cell using the attractors derived from [44]. As expected, the system progresses cyclically through the eight attractor states. The duration of each cycle varies somewhat due to the stochasticity in the update rule from Eq. 2.

Although the gene expression for each simulated cell k has $\sigma_{ik}(t) = \pm 1$, the population-averaged expression has $-1 \leq \langle \sigma_i(t) \rangle_{\mathcal{K}} \leq +1$, and for many cells initially synchronized with $\sigma_{ik}(0) = \xi_i^0$ for all k, $\langle \sigma_i(t) \rangle_{\mathcal{K}}$ successfully recovers the experimentally observed decaying sinusoidal gene expression. Figure 3 shows a comparison between the experimental expression $x_i(t)$ from the Eser data set and the mean simulated expression $\langle \sigma_i(t) \rangle_{\mathcal{K}}$ with $\kappa = 50$ for $i = SLD2$, one of the genes responsible for initiating DNA replication in *S. cerevisiae* [48, 49]. The simulation time t was rescaled by eye to align the simulated and experimental curves.

Trajectories can be visualized by projecting them onto the first two principal components (PCs) of the attractor configurations. Figure 4 shows the eight attractors as stars, and a single cell trajectory (left panel) and 100 cell trajectories (right panel) with random initial states as curves with line segments colored according to $s_k(t)$ (as computed in the full N-dimensional space). Although the cells begin nearly equidistant from all ξ_i^μ, they quickly relax into encoded cycle. S1 Video shows an animation of a system of $\kappa = 50$ cells with random initial conditions projected onto the same PCs, where cells (circles) are colored according to $s_k(t)$. As with the cells shown in Figure 4, all initially random configurations eventually converge to the cycle. S2 Video shows an animation of $\kappa = 50$ cell trajectories with $\sigma_{ik}(0) = \xi_i^0$. As time progresses, the phases of the initially synchronized cells slowly decohere because cells stochastically and independently transition between attractors due to the finite temperature in Eq. 2.

S3 Video demonstrates the effect of temperature on the dynamics. 50 cells were given random initial states, and the temperature was increased and decreased in steps. Cells rarely escape the eight attractor states for $T = 0.045$, and one cell becomes stuck near the center in a spurious attractor (unintentional metastable states that arise from the model's nonlinearity). At $T = 0.06$, fluctuations allow the cells to transition somewhat regularly through the encoded cycle, and the cell trapped in the spurious attractor eventually escapes and joins the cycle. At $T = 0.09$ the cells begin to
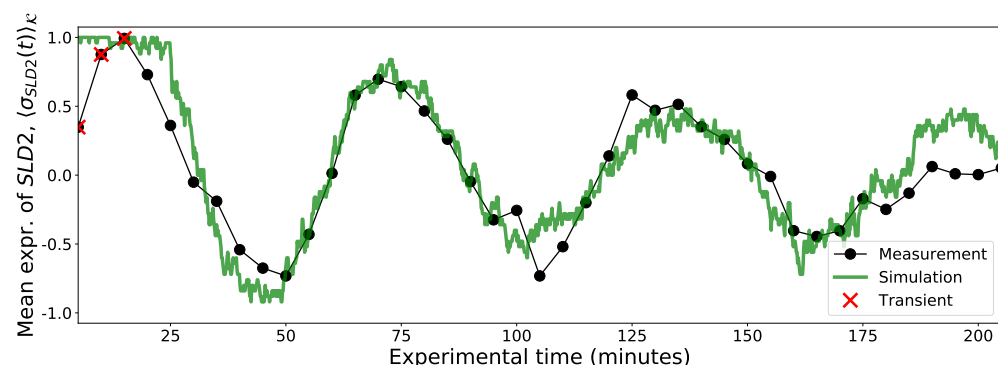
Figure 3. Measured and simulated gene expression. The measured expression of the *S. cerevisiae* gene *SLD2* from [44] was scaled to the range $[-1, +1]$ and is shown as a black beaded curve, and the population-averaged expression of the same gene as defined in Eq. 12 for $\kappa = 50$ cells is shown in green, with the t axis rescaled by eye to match experimental time. Transient points (red X's) were ignored when fitting Eq. 17.

noticeably diverge from the eight attractor states, but still collectively display a net clockwise flow. The noise is too great for the cells to follow the cycle for $T = 0.15$, but lowering the temperature again returns cells to the cycle. This illustrates the fact that the cycle is preserved only for intermediate temperatures: cells become "frozen" in intended or spurious attractor states at low temperatures, but at high temperatures the noise is too great and the couplings between genes become irrelevant to the dynamics.

Optimal control fields

The GA was used to identify some effective combinations of gene targets that slowed progress through the cyclic attractor for varying numbers of targets, n_{targ}. Because each gene has one of eight discrete phases, there can be multiple equivalent optimal control sets. Here we present and discuss only some of the optimal sets. Extensive tables of results can be found in S1 File.

The GA found that inhibiting the set of eight *S. cerevisiae* genes *HEK2*, *PRR1*, *QRI1*, *RFC4*, *STB1*, *TDA7*, *VPS17*, and *ZIM17* was sufficient to trap ∼95% of cells in the $\mu = 7$ state. The effects of this control field on the time evolution of $P_\mu(t)$ for $\kappa = 50$ and $\kappa = 5000$ are shown in Figure 5. Cells were given random, independent initial states at $t = -200$ (not shown), quickly settling into the cyclic attractor with evenly distributed phases so that $P_\mu(0 \leq t < 200) \approx 1/8$. The control field was activated at $t = 200$, causing the cells to accumulate in the $\mu = 7$ state. The field was then disabled at $t = 1000$, allowing the cells to resume cycling with initially synchronized phases, as shown by the sequence of oscillations in $P_\mu(t > 1000)$. The stochastic nature of the transitions causes the cells' phases to slowly spread so that $P_\mu(t \to \infty) \approx 1/8$, eventually returning the system to a desynchronized state.

The effects of this control field can also be visualized using a PC projection as shown in Figure 6 and S4 Video. The same set of $\kappa = 50$ trajectories from Figure 5 was projected onto the attractors' PCs, with cells colored according to $s_k(t)$. The control field manages to fix most cells near the $\mu = 7$ state, but as shown in the $t = 910$ panel in Figure 6, fluctuations occasionally push individual cells out of the $\mu = 7$ basin and back into the cycle.

Four further searches were constrained to inhibiting 2, 4, 6, and 8 out of the 24 periodic kinases [50, 51] in HeLa cells. In all cases, the GA found $\mu = 2$ (M phase) to be
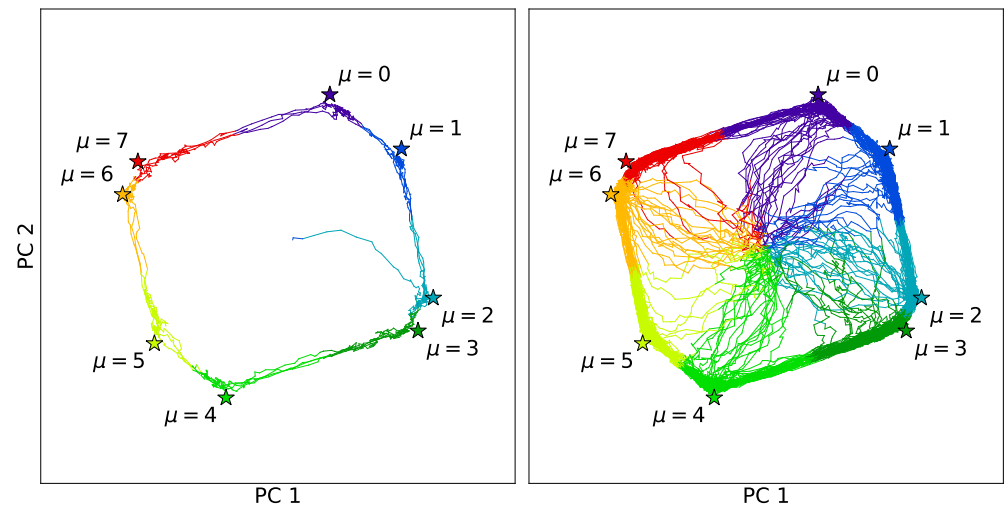
Figure 4. Principal component projection of unperturbed cell trajectories.
The simulated single cell (left panel) and 100 cells (right panel) began with random initial states (projected near the center of the plot), but quickly settled into the encoded cycle. Line segments were colored according to $s_k(t)$, i.e. which of the eight attractors (labeled stars) had maximum overlap at time t.
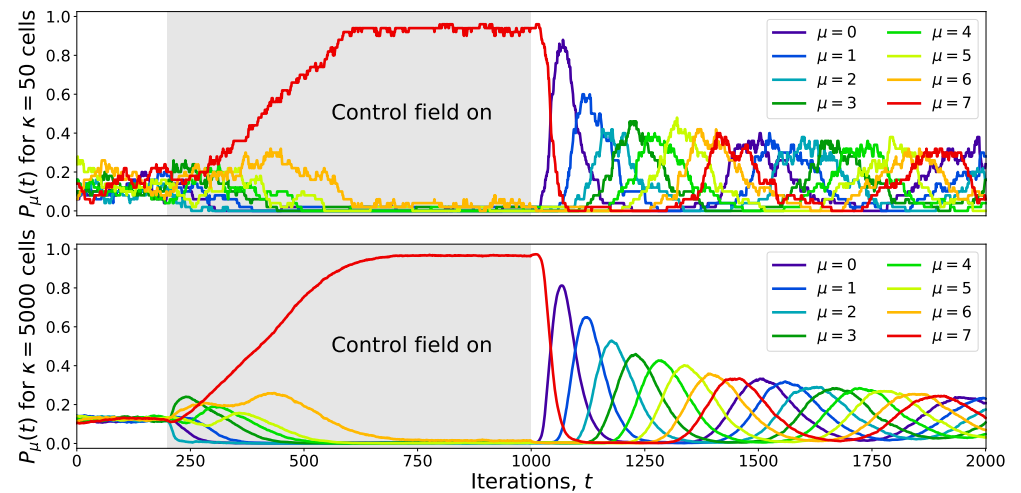

Figure 5. Cell state synchronization by targeted inhibition for 50 and 5000 cells. Cells were initialized with random states at $t = -200$ (not shown) and allowed to relax into the cyclic attractor so that $P_\mu(0 \leq t < 200) \approx 1/8$. A set of eight genes was inhibited with an external control field over the range $200 \leq t < 1000$, fixing most cells near the $\mu = 7$ state. After removing the control field, the cells resumed moving through the cycle with initially synchronized phases that slowly broaden. Eventually the system returns to a desynchronized state, $P_\mu(t \to \infty) \approx 1/8$.
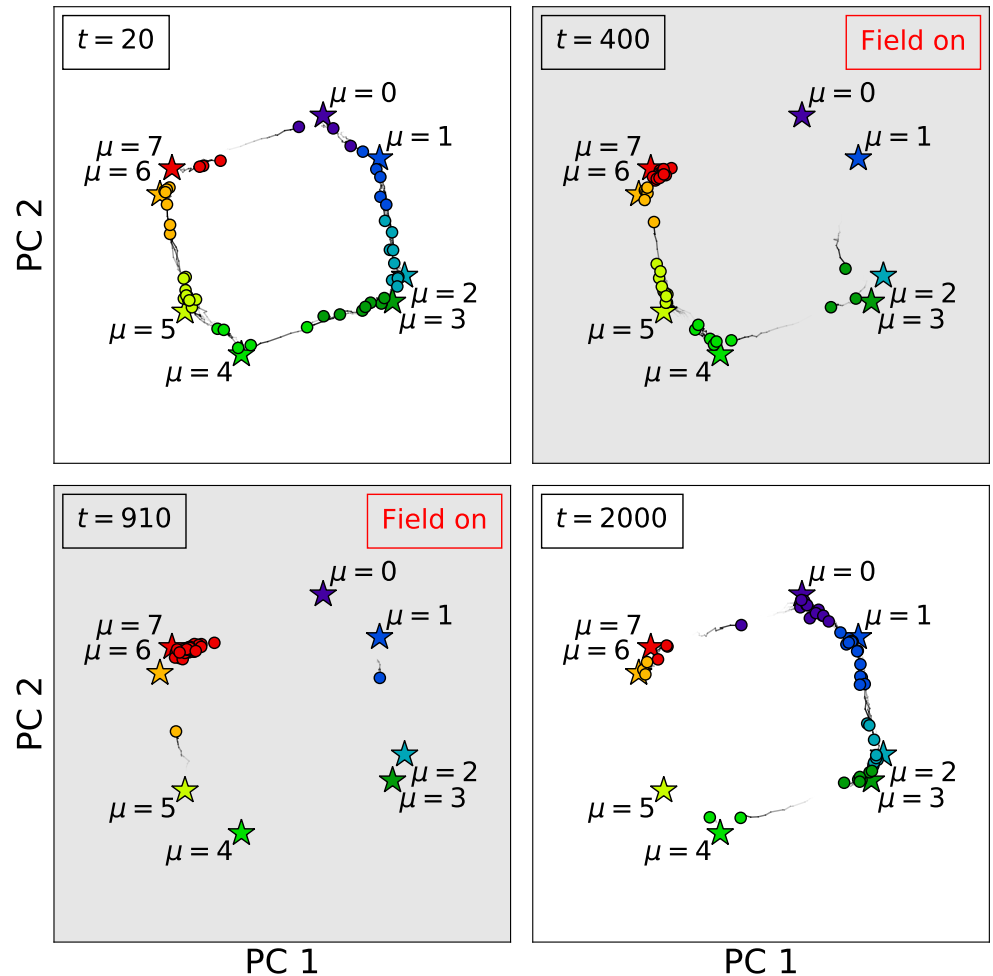
Figure 6. Principal component projection of 50 cell trajectories. The trajectories used to make the $\kappa = 50$ panel of Figure 5 were projected onto the the first two principal components (PCs) of the attractor array ξ_i^μ (labeled stars). Cells (circles) are colored according to the closest attractor as computed by Eq. 13. When the external field is activated, most cells become trapped in the $\mu = 7$ state, although occasionally cells break from the group and complete another circuit before becoming trapped again. After the external field is removed, the cells eventually return to a desynchronized state. See S4 Video for an animation of these trajectories.

Attr. index	Approx. CC phase	Kinase inhibition combination				Score
0	G2/M	MINK1	NRBP2	OBSCN	TAF1	0.192
1	M	HIPK1	NEK7	TAF1	YES1	0.190
2	M	BRD4	MAPK1	NEK7	YES1	0.264
3	M/G1*	CDC42BPB	EPHB2	MAPK1	PTK2	0.201
4	M/G1*	EPHB2	MNAT1	PRKDC	PTK2	0.197
5	G1/S	CSNK1D	HIPK1	PANK2	PRKDC	0.203
6	S	CDK10	CDK9	CSNK1D	MAP4K2	0.178
7	S/G2*	CDK10	CDK5	MAP4K2	MINK1	0.229

Table 1. HeLa kinase inhibition search results. The genetic algorithm was used to identify the best combinations of four kinases to inhibit in order to freeze cells in each attractor, along with the optimal score (the time-averaged fraction of cells occupying that attractor state). Phases marked with an asterisk (*) were not found to be significantly enriched for any CC phase, and so are labeled as being between the previous and next known phases.

the most controllable attractor, with an optimal score of $\langle P_2 \rangle_{\mathscr{T}} \approx 0.264$ for $n_{\text{targ}} = 4$ (0.8% of the 519 periodic genes) when inhibiting the genes *BRD4*, *MAPK1*, *NEK7*, and *YES1*; and an optimal score of $\langle P_2 \rangle_{\mathscr{T}} \approx 0.281$ for $n_{\text{targ}} = 8$ when inhibiting the genes *BRD4*, *MAPK1*, *NEK7*, *YES1*, *CDC42BPB*, *PRKDC*, *PTK2*, and *TRIM28*. Table 1 shows the results for $n_{\text{targ}} = 4$ for all attractors.

The structure of the coupling matrix was probed using centrality measures from complex network theory [52] by taking the absolute value of the coupling matrix's elements as edge weights. Katz centrality and PageRank were found to be poor predictors of optimal target sets, but betweenness centrality proved to be a very effective predictor. *BRD4*, *MAPK1*, *NEK7*, and *YES1* have the four highest betweenness centralities in the network (with a mean betweenness centrality of 2.4×10^{-3} and a p-value of 4.7×10^{-4}, using the inverse of the absolute value of the coupling strengths as weights), indicating that this set of kinases forms a kind of bottleneck in the transmission of signals through the network. Structural network measures, however, do not account for the time-dependent expression of targeted genes, nor how downstream gene expression reacts to upstream perturbations. Controlling nonlinear dynamical systems requires investigating both the structure of the underlying network and the specific form of interactions as defined through the signaling rules.

Conclusions

Above we presented a delayed cyclic Hopfield model designed to store CC time series gene expression data from synchronized *S. cerevisiae* and HeLa cells, and the behaviors of both individual cells and populations of cells were studied. The dynamics of populations of cells successfully recreated the experimental gene expression data, including the slow decoherence of initially synchronized cells due to the stochastic transitions between attractors. Optimal control fields that freeze or stall the cyclic attractor by inhibiting only a small number of genes were identified. These predictions could be experimentally validated or invalidated using kinase inhibitors or knockout studies.

Admittedly, there are several limitations to this model. The specific results reported here depend to some degree on the free parameters T, λ, p, and the node update probability. Tuning the system to match the behavior of the underlying data places some constraints on these parameters, but a more detailed study of the sensitivity of the results to these parameters could prove useful. Additionally, although using the temporal ordering of the time series gene expression samples provides more information

about potentially causal relationships than static samples, the Hopfield model is $\quad$ 277
ultimately an effective model that builds gene-gene couplings from pairwise correlations $\quad$ 278
in gene expression, thereby capturing direct, indirect, and spurious relationships $\quad$ 279
between genes. Independently derived network information with experimentally $\quad$ 280
confirmed molecular regulatory interactions could perhaps be used to refine the $\quad$ 281
construction of the coupling matrix. $\quad$ 282

$\quad$ Our approach can be generalized and improved in many ways. This incarnation of $\quad$ 283
the model causes simulated cells to continuously undergo CC with no G0 (resting) $\quad$ 284
phase. Adding a relatively stable G0 attractor between the M and G1 phases could $\quad$ 285
cause cells to pause between cycles. A GA search could then be conducted to find the $\quad$ 286
best sets of inhibition targets to freeze cells in the G0 state, or to find the best sets of $\quad$ 287
targets to stimulate entry into CC, mimicking the effects of environmental signals such $\quad$ 288
as growth factors. $\quad$ 289

$\quad$ We chose to discretize the continuous gene expression data using a traditional $\quad$ 290
two-state model, which assumes that each gene is either fully activated or fully $\quad$ 291
deactivated. Using a multi-level Hopfield model [53] could better reflect the continuous $\quad$ 292
nature of gene expression data and potentially improve the search results. This model $\quad$ 293
can also incorporate additional omics information, e.g. proteomics and metabolomics, $\quad$ 294
simply by increasing the number of nodes in the system. We plan to explore this option $\quad$ 295
as more multi-omics time series data sets become available. Single-cell experimental $\quad$ 296
techniques and analytical tools are also rapidly improving in quality, decreasing in cost, $\quad$ 297
and gaining in popularity [54–56], and using techniques like pseudo-temporal $\quad$ 298
ordering [57] could allow the Hopfield model to encode single-cell RNA-seq data as well. $\quad$ 299

$\quad$ Although the above simulated populations of cells exhibit intriguing dynamical and $\quad$ 300
statistical properties, they behave as completely homogeneous, non-interacting particles. $\quad$ 301
The importance of cell-cell communication and interactions in populations of cells has $\quad$ 302
been demonstrated in a variety of systems including bacterial quorum sensing [58] and $\quad$ 303
community spatial patterning [59], neuron synchronization in circadian rhythm [60], and $\quad$ 304
various forms of cancer [61–65]. As with many nonlinear systems, even seemingly minor $\quad$ 305
changes can produce dramatically different outcomes. More complex extensions to our $\quad$ 306
model could incorporate cell-cell communication by, for example, adding couplings $\quad$ 307
between known signaling molecules and receptors between different cells, and could even $\quad$ 308
allow for interactions between heterogeneous cell types. This would increase the $\quad$ 309
computational complexity of the model, but could better reflect the underlying biology. $\quad$ 310

Methods $\quad$ 311

Gene expression fitting $\quad$ 312

In order to encode these CC data sets into the Hopfield model, periodic genes needed to $\quad$ 313
be identified, their frequencies and phases computed, and their expression converted $\quad$ 314
from continuous to Boolean form. SciPy's Trust Region Reflective method [66] was used $\quad$ 315
to identify genes i with periodic expression $x_i(t)$ by fitting to the form $\quad$ 316

$$x_i(t) = a_i e^{-b_i t} \cos(\omega_i t - \phi_i) + x_{i0} \tag{17}$$

for amplitude a_i, decay rate b_i, angular frequency ω_i, phase ϕ_i, and asymptotic mean $\quad$ 317
expression x_{i0}. (Because the HeLa data set has fewer time points (14) than the *S.* $\quad$ 318
cerevisiae data set (41), analysis of the HeLa data set was preceded by a smoothing step $\quad$ 319
using a simple box filter to aid in fitting.) The first several time points were ignored to $\quad$ 320
avoid fitting the parameters of Eq. 17 to chemically perturbed (transient) states. A gene $\quad$ 321

was labeled periodic if the maximum relative uncertainty in its parameters from the fit,

$$r_i^{\max} = \max\left\{\frac{\delta x_{i0}}{x_{i0}}, \frac{\delta a_i}{a_i}, \frac{\delta b_i}{b_i}, \frac{\delta \omega_i}{\omega_i}, \frac{\delta \phi_i}{2\pi}\right\}\ , \tag{18}$$

was less than the thresholds defined in S1 Table. Once all frequencies $\{\omega_i\}$ for periodic genes were computed, the frequency was fixed to the mean frequency $\bar{\omega}$ and the fits were recomputed for each periodic gene using the form

$$x_i(t) = a_i e^{-b_i t} \cos\left(\bar{\omega} t - \phi_i\right) + x_{i0}\ , \tag{19}$$

thus producing the final set of continuous phases $\{\phi_i\}$. Figure 1A shows a heat map of the expression of all periodic genes detected in the Eser data set sorted by their fitted phases, and Figure 1B shows the same genes with the fitted expression curves. These fitted curves were converted from continuous values $x_i(t) \geq 0$ to Boolean values $\xi_i(t) = \pm 1$ (over/underexpressed) by assigning

$$\xi_i(t) = \text{sign}(x_i(t) - x_{i0}) \tag{20}$$

as shown in Figure 1C. Finally, one CC period was divided into eight uniformly spaced states $\{\xi_i^\mu\} = \{\xi_i^0, \xi_i^1, \dots, \xi_i^7\}$. These states, shown in Figure 1D, were used as attractors in the Hopfield model.

Determining cell cycle phase

To determine the approximate CC phases for each attractor μ in the HeLa data set, over-representation analysis was conducted using the hypergeometric distribution to calculate p-values with the Benjamini-Hochberg procedure [67] to correct for multiple hypothesis testing with false discovery rate < 0.05, using all genes i with $\xi_i^{\mu-1} = -1$ and $\xi_i^\mu = +1$ as the μth input set and using all detected cyclic genes as the background. In the Dominguez HeLa data set, $\mu = 0$ was enriched for the gene ontology (GO) terms "negative regulation of cell proliferation" ($p = 2.3 \times 10^{-2}$) and "DNA double-strand break repair" ($p = 3.1 \times 10^{-2}$), corresponding to the G2/M checkpoint. $\mu = 2$ was enriched for the GO term "nuclear envelope breakdown" ($p = 4.2 \times 10^{-3}$), corresponding to the preparation of chromosome condensation and cellular mitosis. $\mu = 5$ ("*TP53* regulates transcription of DNA repair genes," $p = 5.0 \times 10^{-3}$) and $\mu = 6$ ("DNA strand elongation," $p = 2.0 \times 10^{-3}$) correspond to the G1/S phase checkpoints and the elongation of DNA in S-phase respectively. The database `yeastgenome.org` [68] was used to determine the CC phases for the Eser *S. cerevisiae* data set. $\mu = 0$ is enriched for the GO term "DNA replication" ($p = 2.02 \times 10^{-12}$), indicating an attractor in the S phase of CC. $\mu = 2$ is enriched for "mitotic spindle organization" ($p = 2.3 \times 10^{-3}$) indicating the beginning of mitosis in *S. cerevisiae*. $\mu = 6$ from Eser is enriched for the GO term "pre-replicative complex assembly involved in nuclear cell cycle DNA replication" ($p = 3.0 \times 10^{-5}$), indicating an attractor at the end of G1 phase as the cells prepare for DNA replication.

Supporting Information

S1 Video

50 cell trajectories with random initial conditions. Data was projected onto the first two principle components of the attractor array ξ_i^μ. Attractors are shown as stars, and cells are shown as circles. Cell colors are assigned using $s_k(t)$ as measured in the full N-dimensional space. All cells k were given random initial conditions $\sigma_{ik} = \pm 1$ with equal probability for all i and k, but eventually converge to the cyclic attractor.

S2 Video

50 cell trajectories with identical initial conditions. See the caption of S1 Video for an explanation of the projection and colors. All cells k were initially synchronized with $\sigma_{ik}(t) = \xi_i^0$, but progress through the cycle stochastically, causing the distribution of $s_k(t)$ to broaden.

S3 Video

Effects of temperature on 50 cell trajectories. See the caption of S1 Video for an explanation of the projection and colors. All cells were given initial random states, and the temperature was increased and decreased in steps as shown in the top panel.

S4 Video

Principal component projection of 50 cells being synchronized. See the caption of S1 Video for an explanation of the projection and colors. This video is an animation of the trajectories used in Figures 5 and 6.

S1 Text

Explanation of genetic algorithm.

S1 Table

List of parameters used for each data set.

S1 File

Zip file containing results of genetic algorithm's gene inhibition searches.

Acknowledgments

References

1. Waddington CH (2014) The Strategy of the Genes, volume 20. Routledge.

2. Kauffman S (1971) Differentiation of malignant to benign cells. Journal of Theoretical Biology 31: 429–451.

3. Huang S, Eichler G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. Physical Review Letters 94: 128701.

4. Alvarez-Buylla ER, Chaos Á, Aldana M, Benítez M, Cortes-Poza Y, et al. (2008) Floral morphogenesis: stochastic explorations of a gene network epigenetic landscape. PLOS ONE 3: e3626.

5. Huang S, Ernberg I, Kauffman S (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. In: Seminars in Cell & Developmental Biology. Elsevier, volume 20, pp. 869–876.

6. Demicheli R, Coradini D (2010) Gene regulatory networks: a new conceptual framework to analyse breast cancer behaviour. Annals of Oncology 22: 1259–1265.

7. Choi M, Shi J, Jung SH, Chen X, Cho KH (2012) Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. Sci Signal 5: ra83–ra83.

8. Yuan R, Zhu X, Radich JP, Ao P (2016) From molecular interaction to acute promyelocytic leukemia: Calculating leukemogenesis and remission from endogenous molecular-cellular network. Scientific Reports 6.

9. Pusuluri ST (2017) Exploring Neural Network Models with Hierarchical Memories and Their Use in Modeling Biological Systems. Ph.D. thesis, Ohio University.

10. Udyavar AR, Wooten DJ, Hoeksema M, Bansal M, Califano A, et al. (2017) Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity. Cancer Research 77: 1063–1074.

11. Wooten DJ, Quaranta V (2017) Mathematical models of cell phenotype regulation and reprogramming: Make cancer cells sensitive again! Biochimica et Biophysica Acta (BBA)-Reviews on Cancer 1867: 167–175.

12. Yuan R, Zhu X, Wang G, Li S, Ao P (2017) Cancer as robust intrinsic state shaped by evolution: a key issues review. Reports on Progress in Physics 80: 042701.

13. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 79: 2554–2558.

14. Peretto P (1988) On learning rules and memory storage abilities of asymmetrical neural networks. Journal de Physique 49: 711–726.

15. Kanter I, Sompolinsky H (1987) Associative recall of memory without errors. Physical Review A 35: 380.

16. Amit DJ, Gutfreund H, Sompolinsky H (1985) Storing infinite numbers of patterns in a spin-glass model of neural networks. Physical Review Letters 55: 1530.

17. Derrida B, Gardner E, Zippelius A (1987) An exactly solvable asymmetric neural network model. Europhysics Letters 4: 167.

18. Nishimori H, Nakamura T, Shiino M (1990) Retrieval of spatio-temporal sequence in asynchronous neural network. Physical Review A 41: 3346.

19. Personnaz L, Guyon I, Dreyfus G (1986) Collective computational properties of neural networks: New learning mechanisms. Physical Review A 34: 4217.

20. Nasrabadi NM, Li W (1991) Object recognition by a Hopfield neural network. IEEE Transactions on Systems, Man, and Cybernetics 21: 1523–1535.

21. Cote S, Tatnall A (1997) The Hopfield neural network as a tool for feature tracking and recognition from satellite sensor images. International Journal of Remote Sensing 18: 871–885.

22. Maetschke SR, Ragan MA (2014) Characterizing cancer subtypes as attractors of Hopfield networks. Bioinformatics 30: 1273–1279.

23. Lang AH, Li H, Collins JJ, Mehta P (2014) Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. PLOS Comput Biol 10: e1003734.

24. Szedlak A, Paternostro G, Piermarocchi C (2014) Control of asymmetric Hopfield networks and application to cancer attractors. PLOS ONE 9: e105842.

25. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57–63.

26. Marguerat S, Bähler J (2010) RNA-seq: from technology to biology. Cellular and Molecular Life Sciences 67: 569–579.

27. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7: 562–578.

28. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7: S7.

29. Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R (2009) Gene regulatory network inference: data integration in dynamic models – a review. Biosystems 96: 86–103.

30. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Research 21: 1109–1121.

31. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, et al. (2012) Wisdom of crowds for robust gene network inference. Nature Methods 9: 796–804.

32. Cahan P, Li H, Morris S, LummertzdaRocha E, Daley G, et al. (2014) CellNet: network biology applied to stem cell engineering. Cell 158: 903–915.

33. Ong E, Szedlak A, Kang Y, Smith P, Smith N, et al. (2015) A scalable method for molecular network reconstruction identifies properties of targets and mutations in acute myeloid leukemia. Journal of Computational Biology 22: 266–288.

34. Szedlak A, Smith N, Liu L, Paternostro G, Piermarocchi C (2016) Evolutionary and topological properties of genes and community structures in human gene regulatory networks. PLOS Computational Biology 12: e1005009.

35. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, et al. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148: 1293–1307.

36. Assaf M, Roberts E, Luthey-Schulten Z, Goldenfeld N (2013) Extrinsic noise driven phenotype switching in a self-regulating gene. Physical Review Letters 111: 058102.

37. Evan GI, Vousden KH (2001) Proliferation, cell cycle and apoptosis in cancer. Nature 411: 342.

38. Malumbres M, Barbacid M (2009) Cell cycle, CDKs and cancer: a changing paradigm. Nature Reviews Cancer 9: 153.

39. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646–674.

40. Diaz-Moralli S, Tarrado-Castellarnau M, Miranda A, Cascante M (2013) Targeting cell cycle regulation in cancer therapy. Pharmacology & Therapeutics 138: 255–271.

41. Feala JD, Cortes J, Duxbury PM, Piermarocchi C, McCulloch AD, et al. (2010) Systems approaches and algorithms for discovery of combinatorial therapies. Wiley Interdisciplinary Reviews: Systems Biology and Medicine 2: 181–193.

42. Csermely P, Agoston V, Pongor S (2005) The efficiency of multi-target drugs: the network approach might help drug design. Trends in Pharmacological Sciences 26: 178–182.

43. Ágoston V, Csermely P, Pongor S (2005) Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example. Physical Review E 71: 051909.

44. Eser P, Demel C, Maier KC, Schwalb B, Pirkl N, et al. (2014) Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. Molecular Systems Biology 10: 717.

45. Dominguez D, Tsai YH, Gomez N, Jha DK, Davis I, et al. (2016) A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. Cell Research .

46. Kohonen T (1974) An adaptive associative memory principle. IEEE Transactions on Computers 100: 444–445.

47. Kauffman SA, Johnsen S (1991) Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches. Journal of Theoretical Biology 149: 467–505.

48. Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H (2010) CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ϵ, and GINS in budding yeast. Genes & Development 24: 602–612.

49. Bruck I, Kaplan DL (2014) The replication initiation protein Sld2 regulates helicase assembly. Journal of Biological Chemistry 289: 1948–1959.

50. Yang CY, Chang CH, Yu YL, Lin TCE, Lee SA, et al. (2008) PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. Bioinformatics 24: i14–i20.

51. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, et al. (2011) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Research 40: D261–D270.

52. Newman M (2010) Networks: An Introduction. Oxford University Press.

53. Zurada JM, Cloete I, Van der Poel E (1996) Generalized Hopfield networks for associative memories with multi-valued stable states. Neurocomputing 13: 135–149.

54. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344: 1396–1401.

55. Scialdone A, Natarajan KN, Saraiva LR, Proserpio V, Teichmann SA, et al. (2015) Computational assignment of cell-cycle stage from single-cell transcriptome data. Methods 85: 54–61.

56. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. (2017) Massively parallel digital transcriptional profiling of single cells. Nature Communications 8: 14049.

57. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, et al. (2014) Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. Nature Biotechnology 32: 381.

58. Waters CM, Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21: 319–346.

59. Momeni B, Brileya KA, Fields MW, Shou W (2013) Strong inter-population cooperation leads to partner intermixing in microbial communities. eLife 2: e00230.

60. Kalsbeek A, Merrow M, Roenneberg T, Foster R (2012) Suprachiasmatic nucleus: cellular clocks and networks. The Neurobiology of Circadian Timing 199: 129.

61. Skog J, Würdinger T, Van Rijn S, Meijer DH, Gainche L, et al. (2008) Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nature Cell Biology 10: 1470–1476.

62. Hong BS, Cho JH, Kim H, Choi EJ, Rho S, et al. (2009) Colorectal cancer cell-derived microvesicles are enriched in cell cycle-related mRNAs that promote proliferation of endothelial cells. BMC Genomics 10: 556.

63. Renzulli JF, Del Tatto M, Dooner G, Aliotta J, Goldstein L, et al. (2010) Microvesicle induction of prostate specific gene expression in normal human bone marrow cells. The Journal of Urology 184: 2165–2171.

64. Del Tatto M, Ng T, Aliotta JM, Colvin GA, Dooner MS, et al. (2011) Marrow cell genetic phenotype change induced by human lung cancer cells. Experimental Hematology 39: 1072–1080.

65. Tetta C, Ghigo E, Silengo L, Deregibus MC, Camussi G (2013) Extracellular vesicles as an emerging mechanism of cell-to-cell communication. Endocrine 44: 11–19.

66. Jones E, Oliphant T, Peterson P, et al. (2001–). SciPy: Open source scientific tools for Python. URL http://www.scipy.org/.

67. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Statistics in Medicine 9: 811–818.

68. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, et al. (2011) Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Research 40: D700–D705.