# A model of phenotypic state dynamics initiates a promising approach to control heterogeneous malignant cell populations

Margaret P. Chapman[1], Tyler T. Risom[2], Anil Aswani[3], Roel Dobbe[1], Rosalie C. Sears[2], and Claire J. Tomlin[1]

***Abstract*— A growing body of experimental evidence indicates a strong link between intratumoral heterogeneity and therapeutic resistance in cancer. In particular, tumor cells may survive therapy by switching their phenotypic identities to more resistant, drug-tolerant states. Computational models of phenotypic plasticity in response to cytotoxic therapy are needed: (1) to strengthen understanding of the interplay between phenotypic heterogeneity and therapeutic resistance, and (2) to identify potential strategies *in silico* that weaken resistance prior to *in vitro* testing. This work presents a linear time-invariant model of phenotypic state dynamics to deduce subpopulation-level behavior likely to affect temporal phenotypic composition and thus drug resistance. The model was identified under different therapeutic conditions with authentic biological data from a breast cancer cell line. Subsequent analysis suggested drug-induced effects on phenotypic state switching that could not be deduced directly from empirical observations. A bootstrap algorithm was implemented to identify statistically significant results: reduction in cell division under each therapeutic condition versus control. Further, Monte Carlo simulation was used to evaluate quality of model fit for two-way switching and net switching on synthetically generated data to determine the limitations of the latter assumption for subsequent modeling. Most importantly, the simple model structure initiated a control-theoretic approach for identifying promising combination treatments *in silico* to guide future laboratory testing.**

## I. Introduction

There is a growing appreciation of the role of intratumoral heterogeneity in therapeutic resistance in cancer. Breast tumors in particular can be heterogeneous in tumor cell genomes and cell phenotypes, and importantly, both of these heterogeneities promote therapeutic resistance [1][2]. Intratumoral genetic heterogeneity populates the tumor with subclones of differing therapeutic sensitivity, which can drive tumor resistance through a Darwinian selection process [3][4][5]. Phenotypic heterogeneity can promote resistance through two mechanisms. First, similar to genetic heterogeneity, phenotypic subpopulations can differ in therapeutic sensitivity and those with innate therapeutic resistance can be selected for throughout treatment [6][7][8]. Secondly, tumor cell phenotypes are inherently plastic, and thus cells can adapt to therapy by switching their phenotype to a more resistant identity [9][10][11][12].

[1]M.C., R.D., and C.T. are with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, 94709, USA. chapmanm@berkeley.edu

[2]T.R. and R.S. are with the Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR, 97239, USA.

[3]A.A. is with the Department of Industrial Engineering and Operations Research at UC Berkeley.

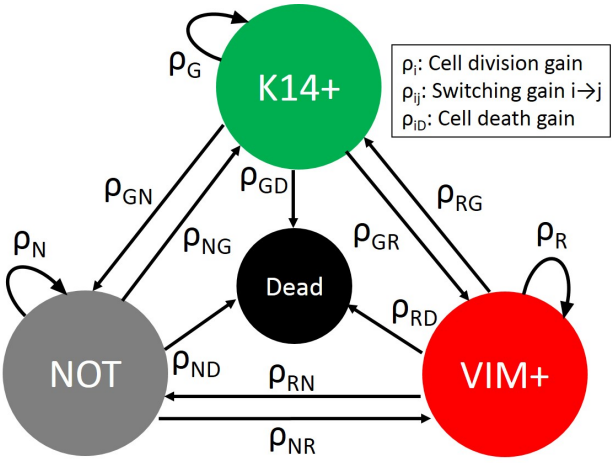


Fig. 1. A schematic of the phenotypic state dynamical model is shown. The cell types of interest are death and three phenotypic states (K14+ *green*, VIM+ *red*, and neither *not*). The model deduces subpopulation-level behavior likely to affect temporal phenotypic composition. Behaviors of interest (cell division, death, and switching) are represented via time-invariant gains identified from empirical observations.

Numerous computational models of fitness competition and Darwinian selection have aided our understanding of how innate therapeutic resistance in tumor subpopulations can drive tumor resistance [3][13][14][15]. Conversely, modeling phenotypic plasticity in response to cancer therapy is a relatively new area. More research is needed to exploit the potential of data-driven systems-based approaches for controlling malignant heterogeneous populations.

*Prior work.* We describe a sample of the existing models of phenotypic plasticity and how our work is distinct. The phenotypic state dynamical model from Goldman et al. is the closest precursor to our work. They use a linear time-invariant deterministic model in continuous time and identify net proliferation and switching rates from experimental data in order to test drug-induced phenotypic plasticity versus clonal selection [9]. However, we use explicit measurements of percent death and S-phase content to identify cell division and death gains. Further, we analyze statistical significance, compare two distinct switching assumptions *in silico*, and explore controller design for steering the cancer system to a more treatable state.

Moreover, a Markov model of phenotypic state switching developed by Gupta and colleagues does not permit change in total cell population size over time and is identified using two time points [12]. In contrast, our model permits temporal change in total cell population size by representing cell

division and death, and is identified using 6 time points over a 60-hour time horizon.

Further, Almendro et al. represents spatial distribution of phenotypes over time in addition to switching. Rates of phenotype switching were chosen to fit the distribution of cell types found in post-treatment tumor samples [16]. Conversely, our model is identified using linear regression on time course data.

*Contributions.* Motivated by prior work, we identify a linear time-invariant model of phenotypic state dynamics (Fig. 1) using biological data (Sec. II). Frequencies of cell division and death in addition to switching between phenotypic state pairs are deduced as described in Sec. III, and therapeutic effects on subpopulation-level behavior relative to normal growth are hypothesized (Sec. IV). Further, we perform statistical analysis on the model parameters (Sec. V), evaluate two variants of switching on synthetic data (Sec. VI), and initiate a data-driven approach rooted in systems theory to control heterogeneous malignant cell populations (Sec. VII). The methods we use generally apply to any cancer system with phenotypic heterogeneity, if sufficient time series data of phenotypic state counts and percent death are available.

## II. Biological background

We discuss the data collection procedure and how we constructed data sets suitable for modeling below.

### A. Biological experiment

Total cell count, dead cell count, phenotypic state composition, and *S-phase* content in the Triple Negative Breast Cancer cell line, HCC1143, were measured from 4 replicate *wells* every 12 hours over a 60-hour time horizon under a specific condition (4 wells $\times$ 6 time points $=$ 24 samples). The phenotypic states were classified as *basal* and *mesenchymal* via detection of biological markers, cytokeratin 14 (K14+) and vimentin (VIM+), respectively. The basal state is called *green*, the mesenchymal state is called *red*, and the state of left-over cells is called *not*. Wells are physically separate "buckets" of cells, assumed to be independent and identically distributed for each condition. The conditions were initial input of a PI3K/mTOR inhibitor BEZ235, MEK inhibitor GSK1120212, PI3K/mTORi-MEKi combination, and control with the drug vehicle DMSO but no therapeutic agent. S-phase content refers to DNA-synthesizing cells, and is a proxy for cell division. Phenotypic identity and vitality of a cell could not be detected simultaneously, so the quantity of dead cells in each phenotypic state was unknown. However, the total quantity of dead cells and the quantity of (dead or alive) cells in each phenotypic state were known.

### B. Construction of data sets

For each condition, the time courses were integrated into a cohesive data set of quantities of 4 cell types: green, red, not, and dead. Equal proportion of death was assumed across phenotypic states in the construction, since the true distribution was unknown *a priori*. For each sample (well and time point),

1) The dead cell fraction, $F_{death} = \frac{\text{Dead cell count}}{\text{Total cell count}}$.
2) Phenotypic state composition was modified due to death: $x_i = (1 - F_{death})\tilde{x}_i$, such that $\tilde{x}_i$ is the measured count and $x_i$ is the modified count in phenotypic state $i$ (green, red, or not).
3) The dead cell count was computed: $F_{death} \sum_{i=1}^{3} \tilde{x}_i$.

Then, dead cell counts were adjusted for each well to ensure increasing death over time: If the dead cell count at time $k+1$ was less than the dead cell count at time $k$, the former value was reset to the latter value to enforce constant death. Green cell counts in the cohesive data set are shown in Fig. 2.

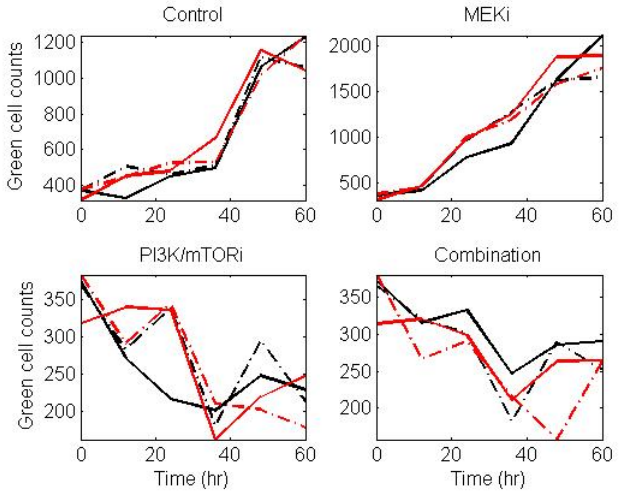


Fig. 2. Green cell counts, realizations of $G_k$ in Eqn. (2), are plotted from four replicate wells over time for each condition; each line style corresponds to a replicate well. The counts are derived from empirical observations, as described in Sec. II-B. These counts and those for red, not, and dead cell types (not shown) were used to identify the dynamics for each condition.

## III. Mathematical model

### A. Structure

We assume a linear time-invariant stochastic model to describe phenotypic state dynamics in discrete time,

$$X_{k+1} = AX_k + \epsilon_k. \tag{1}$$

$X_k \in \mathbb{R}^n$ is a random vector of cell type quantities at time $k$, $A \in \mathbb{R}^{n \times n}$ is the dynamics matrix, $n$ is the number of cell types, and $\epsilon_k \sim \mathcal{N}(0_n, \sigma^2 I_n)$ i.i.d. with unknown $\sigma^2 \in \mathbb{R}$. A stochastic process generates the cell type vectors, and realizations of this process are observed: $x_k^i$ is a realization of $X_k$, computed from measurements in well $i$ (Sec. II-B). $A$ is unknown, to be identified using the realizations.

### B. Rationale

Our choice of model is justified as follows.

(1) *Linear time-invariance captures what is essential for analysis and control.* Modeling constant first-order dynamics is adequate for general analysis of drug-induced phenotypic state division, death, and switching; this level of abstraction sufficiently captures basic trends in system response. Further, controlling phenotypic state dynamics using systems theory

is an open research problem. Our model is appropriate to initiate a broad control framework that can be developed when more data is available.

(2) *A simple model is appropriate for a small data set.* 24 samples were available to identify 12 parameters for each condition (Sec. II-A, Eqn. (3)). Cell populations (wells of cells) were fixed manually at each time point, so observing many wells several times daily was impractical. Fitting a more complex model to the data (e.g., higher-order autoregressive model) would represent abnormalities more strongly than the dynamics of interest.

(3) *Limitations of our model guide future progress*, particularly in the representation of stochasticity (Sec. VIII).

### C. Detailed form

The state vector is:

$$X_k = (G_k, R_k, N_k, D_k)^T \in \mathbb{R}^4, \tag{2}$$

where $G_k$, $R_k$, $N_k$, and $D_k$ denote random quantities of green, red, not, and dead cells at time $k$, respectively. The dynamics matrix is:

$$A = \begin{bmatrix} \alpha_G & \rho_{RG} & \rho_{NG} & 0 \\ \rho_{GR} & \alpha_R & \rho_{NR} & 0 \\ \rho_{GN} & \rho_{RN} & \alpha_N & 0 \\ \rho_{GD} & \rho_{RD} & \rho_{ND} & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4},$$

such that

$$\begin{array}{rcl} \alpha_G &=& \rho_G - \rho_{GR} - \rho_{GN} - \rho_{GD} \\ \alpha_R &=& \rho_R - \rho_{RG} - \rho_{RN} - \rho_{RD} \\ \alpha_N &=& \rho_N - \rho_{NG} - \rho_{NR} - \rho_{ND} \end{array} . \tag{3}$$

Terms of the form $\rho_{ij}$ such that $i \neq j$, $\rho_i$, and $\rho_{iD}$ are phenotypic state switching, division, and death gains, respectively. A *switching gain* is the proportion of cells in a phenotypic state that switch to another phenotypic state during a time interval (12 hours), averaged over the time horizon (60 hours). A *death gain* is the proportion of cells in a phenotypic state that die during a time interval, averaged over the time horizon. A *division gain* is the ratio of cells in a phenotypic state at the end of an interval, to those at the start of that interval, assuming only replication may occur, averaged over the time horizon.

Here we justify the phenotypic state dynamics equations by example. Consider the green phenotypic state. Cell division within the green state and transition of cells from other phenotypes to the green state increases subpopulation size, while cell death within the green state and transition from the green state to other phenotypes decreases subpopulation size. The dynamics equation for green cells is, quite naturally,

$$G_{k+1} = \alpha_G G_k + \rho_{RG} R_k + \rho_{NG} N_k + 0 D_k.$$

The final term, $0D[k]$, indicates that dead cells stay dead and is shown for completeness. Logic for the remaining phenotypic state dynamics equations is equivalent.

Conversely, the dead cell dynamics equation is,

$$D_{k+1} = \rho_{GD} G_k + \rho_{RD} R_k + \rho_{ND} N_k + D_k.$$

TABLE I

CONSTANTS FOR AUTHENTIC AND SYNTHETIC DATA SETS

| Symbol | Description | Authentic value | Synthetic value |
|---|---|---|---|
| $n$ | Number of cell types | 4 | 4 |
| $T$ | Length of time horizon | 6 | 6 |
| $W$ | Number of wells | 4 | 100 |

TABLE II

PROPERTIES OF $A \in \mathcal{A}$

Note: Each subscript $i$, $j \in \{R, G, N\}$ corresponds to a red, green, or not phenotypic state, respectively. Refer to (3).

| Label | Constraint | Rationale |
|---|---|---|
| 1 | $\rho_i \geq 1$ | Cell division can only increase subpopulation size of phenotypic state $i$. |
| 2 | $\rho_i = \rho_j$, $i \neq j$ | Biological evidence (S-phase content, not shown) suggests that division may be equivalent across phenotypic states. |
| 3 | $\rho_{ij} \in [0, 1]$, $i \neq j$ | $\rho_{ij}$ is the average proportion of cells in phenotypic state $i$ that switch to phenotypic state $j$ during a time interval. |
| 4 | $\rho_{iD} \in [0, 1]$ | $\rho_{iD}$ is the average proportion of cells in phenotypic state $i$ that die during a time interval. |
| 5 | $\alpha_i \geq 0$ | Cells that exit phenotypic state $i$ during a time interval must be there initially. |
| 6 | Last column of $A$ is fixed | Dead cells accumulate over time, and death is irreversible. |

The intuition is that dead cells accumulate over time, as cells in any phenotypic state die during each time interval. Stacking the element-wise dynamics equations and applying standard rules of matrix multiplication yields (3).

### D. Identification

Linear least-squares was used to identify the dynamics matrix of (1), since this method produces the maximum likelihood estimate under the assumed error distribution (Example 7.1 [17]),

$$\hat{A} = \underset{A\in\mathcal{A}}{\operatorname{argmin}} \; ||\mathbb{X}^+ - A\mathbb{X}||_F. \tag{4}$$

$F$ denotes the Frobenius norm, $\mathbb{X}^+$ and $\mathbb{X}$ are known matrices of the same size computed from data, and $\mathcal{A} \subset \mathbb{R}^{n\times n}$ is a convex set of biologically plausible dynamics matrices. Realizations $x^i_{k+1}$ and $x^i_k$ form the $j^{th}$ columns of matrices $\mathbb{X}^+$ and $\mathbb{X}$, respectively, where index $j$ corresponds to a unique time point $k$ and well $i$ pair. $\mathbb{X}^+$ and $\mathbb{X}$ belong to $\mathbb{R}^{n\times(T-1)W}$, where $n$ is the number of cell types, $T$ is the length of the time horizon, and $W$ is the number of wells (Table I). Justification for the properties of matrices in $\mathcal{A}$ is presented in Table II. CVX software in MATLAB was employed to solve this optimization program [18], and MATLAB was used for all computations in this paper.

## IV. BIOLOGICAL INSIGHT

A schematic of gain values identified from data for each condition is provided in Fig. 3, and numerical values are listed in the Appendix. Biological insights suggested by our model are described below.

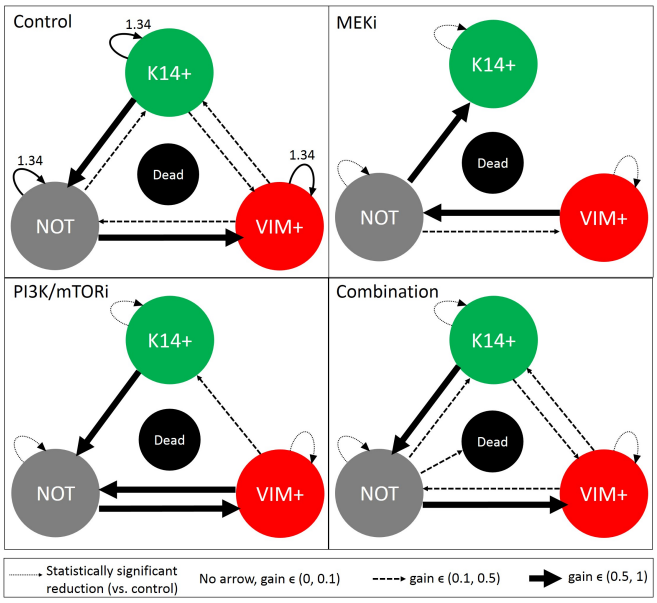


Fig. 3. A schematic of gain values identified from data. Arrow position indicates gain type (see Fig. 1), and arrow style depicts gain magnitude, rounded to two decimal places. Study of control-vs.-therapy subfigures illuminates cytotoxic effect on phenotypic state division, death, and switching.

The green subpopulation increases in size under the MEK inhibitor over time and surpasses the green cell count of the control condition by time 60 hours (Fig. 2). The identified gains suggest that more switching from not to green subpopulations under MEKi (vs. control) may explain this effect. Observe Fig. 3: Under MEKi, the arrow from not to green phenotypic states is thick and no reverse arrow is present; but, in the control, the arrow from not to green phenotypic states is thin and a thick arrow is present in reverse.

Further, the green subpopulation is observed to decrease in size and remains below the green cell count of the control condition over time under PI3K/mTORi (Fig. 2). Less switching from not to green subpopulations and less cell division under PI3K/mTORi (vs. control) may drive this effect, according to the identified gains. Observe Fig. 3: No arrow from not to green phenotypic states is present under PI3K/mTORi; but, an arrow is present from not to green phenotypic states under the control.

Brief comparison of the control condition and MEKi (or PI3K/mTORi) in Fig. 3 indicates that the single-agent therapy exhibits little influence on death. The values of any death gain under the control and MEKi (or PI3K/mTORi) differ by no more than approximately $0.02$ (Table IV). Conversely, the values of some switching gains exhibit relatively large differences between the control and the single-agent therapies; for example, see $\rho_{GN}$ under MEKi vs. control and $\rho_{GR}$ under PI3K/mTORi vs. control (Table V). Also, substantially less division occurs under each therapy compared to control (a statistically significant reduction, to be shown in Sec. V) (Table IV). Together, these model-based findings suggest that MEKi and PI3K/mTORi exhibit greater effects on switching and division relative to death. However, the combination therapy appears to induce some death (see $\rho_{ND}$, combination vs. control, Table IV). A concept for empirical validation of our hypotheses is discussed in future work (Sec. VIII).

## V. Statistical Analysis

The analysis in this section provides an indication of parameter variation due to process noise. *Wild bootstrap* [19][20] was implemented to identify statistically significant differences between therapeutic and control conditions. This flavor of bootstrap was chosen because it has been designed to provide improved results with smaller sample sizes when the data displays heteroskedasticity [19][20], which is the case for this data because variance of each cell type vector depends on its magnitude. Our implementation assumes independent error among cell types for a fixed time point and a fixed well. These steps were executed for each condition:

1) Identify the dynamics matrix, $\hat{A}$, via (4) using data from all observations (4 wells), called *authentic data*.
2) Fit authentic data to $\hat{A}$. Set $\hat{x}^i_{k+1} = \hat{A}x^i_k$, such that $x^i_k$ and $\hat{x}^i_k$ are authentic data and fitted data at time $k$ in well $i$, respectively, and $\hat{x}^i_0 = x^i_0$ for each well $i$.
3) Compute a set of bootstrap samples. Calculate absolute values of the residuals between authentic data and fitted data, $|r^i_{t,k}| = |x^i_{t,k} - \hat{x}^i_{t,k}|$ for all times $k$, wells $i$, and cell types $t$. Initialize the bootstrap set to the fitted set, $z^i_k = \hat{x}^i_k \in \mathbb{R}^n$ for all times $k$ and wells $i$. For each well $i$ such that $k > 0$,
   a) Fix a cell type $t$.
   b) Select random sign $S_t = s_t$, such that $S_t = -1$ or $S_t = 1$ with equal probability.
   c) Set $z^i_{t,k} = z^i_{t,k} + s_t|r^i_{t,k}|$.
   d) Repeat for all cell types, $t = 1, \ldots, n$.
4) Identify the dynamics of the bootstrap set via (4).
5) Repeat steps 3-4 several times (we chose 200). Compute a $95\%$ confidence interval for each gain. Non-overlapping intervals of two distinct conditions for the same gain indicate a statistically significant difference.

Less frequent cell division under each therapeutic condition (vs. control) was found to be statistically significant (Fig. 4). This is an intuitive finding since cancer drugs are expected to suppress proliferation.

## VI. Monte Carlo Evaluation of Switching Type

Here we study two distinct model assumptions: *two-way* switching and *net* switching. In a two-way switching model, there exists two bidirectional switching gains between each phenotypic state pair. Some cells may switch from phenotypic state $i$ to $j$, while other cells may switch from $j$ to $i$, during the same time interval. Two-way switching, an abstraction that reflects plasticity of cancer systems [21][22], is assumed in the proposed model (1), and is defined formally in constraint 3 of Table II. Conversely, in a net switching model, there exists one switching gain that represents the primary switching direction for each phenotypic state pair. For example, if switching occurs from green to red phenotypic states, then it cannot occur from red to green states. Net

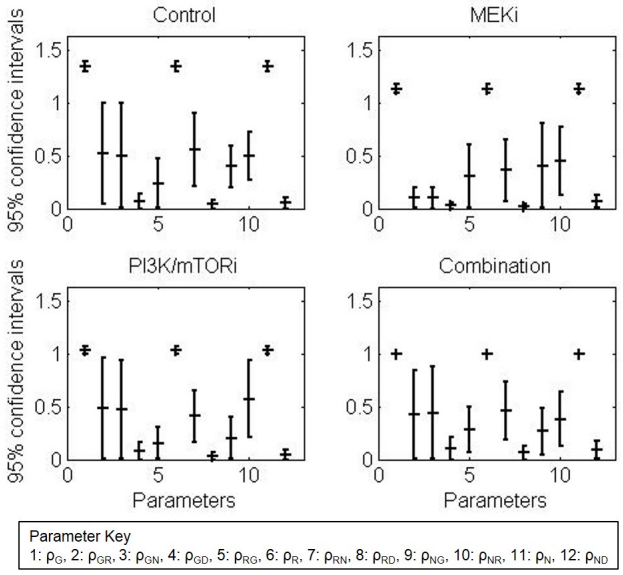


Fig. 4. A 95% confidence interval for each gain, computed via wild bootstrap [19][20], is labeled by x-axis index for each condition. See Tables IV and V for numerical intervals. Recall that division gains for a fixed condition were forced to be equal (Table II). Confidence intervals of division gains under the control condition do not overlap those of any therapeutic condition, indicating a statistically significant effect.

switching imposes a non-convex constraint on non-diagonal terms of the dynamics matrix, $A \in \mathbb{R}^{n \times n}$,

$$(A_{i,j} \in [0,1] \wedge A_{j,i} = 0) \vee (A_{j,i} \in [0,1] \wedge A_{i,j} = 0) \quad (5)$$

for all $i \neq j$, $i \neq n$, $j \neq n$. We denote the non-convex set that satisfies (5) and constraints 1-2, 4-6 of Table II, $\mathcal{A}_{\text{Net}}$.

While two-way switching is more natural, the parameter analysis can be quite complex, especially if many phenotypic states are modeled. Thus, net switching may be desirable to assume in future work. To understand its limitations, we evaluated the quality of model fit for net switching versus two-way switching on synthetically generated data (Table I). A Monte Carlo method was used to assess how the "best" net switching model represented "biologically plausible" data (see Steps 2 and 3 below). The focus was on the best-case scenario to reveal limitations that arise under perfect information; these limitations would only become more severe in practice under imperfect information. The subsequent steps were implemented for each condition:

1) Let the dynamics identified from data via (1) be the synthetic truth for two-way switching, $A^*_{\text{Two-way}} \in \mathcal{A}$.
2) Identify the "best" net switching dynamics matrix using data and a brute force search through the net switching options, $A^*_{\text{Net}} \in \mathcal{A}_{\text{Net}}$.
3) Generate "biologically plausible" synthetic data assuming two-way switching.
   a) Compute the sample mean, $\bar{x}_0$, and the sample covariance, $\Sigma_0$, of authentic data at time 0.
   b) Let $\bar{x}_0$ be the synthetic initial condition.
   c) Set $\sigma^2 = \frac{1}{4r}\sum_{i=1}^{4}\Sigma_{0,ii}$, average of the initial variances for each cell type, scaled by $\frac{1}{r}$. (We chose $r = 50$ by trial and error, so simulated trajectories displayed clear trends with some noise.)
   d) Let $X_0 = \bar{x}_0$, $X_{k+1} = A^*_{\text{Two-way}}X_k + \epsilon_k$, and $\epsilon_k = \sigma\nu_k$, such that $\nu_k \sim \mathcal{N}(0_n, I_n)$ i.i.d. Generate a realized time-trajectory for each synthetic well $i$, $\{x^i_k\}_{k=0}^{T-1}$, using MATLAB function **randn**$(n, 1)$.
   e) Adjust dead cell counts to enforce increasing death over time.
4) Separate synthetic data into train and test sets with 67 and 33 wells, respectively.
5) Estimate the two-way switching dynamics via (4) using synthetic train data, $\hat{A}_{\text{Two-way}}$.
6) Fit synthetic test data to $\hat{A}_{\text{Two-way}}$. $\hat{x}^i_{k+1,\text{Two-way}} = \hat{A}_{\text{Two-way}}x^i_k$, where $\hat{x}^i_{k,\text{Two-way}}$ and $x^i_k$ are the fitted and synthetic test realizations, respectively, at time $k$ in well $i$. $\hat{x}^i_{0,\text{Two-way}} = x^i_0$ for all $i$.
7) Compute the generalization error for two-way switching, $E_{\text{Two-way}}$. Let $\mathbb{X}_{\text{Two-way}}$ and $\hat{\mathbb{X}}_{\text{Two-way}}$ be matrices of synthetic test data, $\{x^i_k\}$, and fitted data, $\{\hat{x}^i_{k,\text{Two-way}}\}$, respectively. $E_{\text{Two-way}} = ||\hat{\mathbb{X}}_{\text{Two-way}} - \mathbb{X}_{\text{Two-way}}||_F$, where $F$ denotes the Frobenius norm.
8) Fit synthetic test data to $A^*_{\text{Net}}$. $\hat{x}^i_{k+1,\text{Net}} = A^*_{\text{Net}}x^i_k$, where $\hat{x}^i_{k,\text{Net}}$ and $x^i_k$ are the fitted and synthetic test realizations, respectively, at time $k$ in well $i$. $\hat{x}^i_{0,\text{Net}} = x^i_0$ for all $i$.
9) Compute the generalization error for net switching, $E_{\text{Net}} = ||\hat{\mathbb{X}}_{\text{Net}} - \mathbb{X}_{\text{Two-way}}||_F$. $\hat{\mathbb{X}}_{\text{Net}}$ is a matrix of fitted data, $\{\hat{x}^i_{k,\text{Net}}\}$, and $\mathbb{X}_{\text{Two-way}}$ was defined earlier.
10) Repeat steps 3-9 for several trials (we chose 10).

Synthetic test data and fitted data for two-way and net switching are shown in Fig. 5 from a trial in a sample execution of the code. The maximum generalization error for two-way switching (Step 7) and for net switching (Step 9) over all trials in a sample execution are provided in Table III. As expected, two-way switching achieves a closer fit to the test data compared to net switching for every condition, evident by better qualitative matching (Fig. 5) and lower error (Table III). Interestingly, fitted data of net switching and two-way switching may differ in derivative magnitude (estimated visually), but generally not in derivative sign, for a given condition, state, and interval (Fig. 5). That is, these fitted data generally increase (or decrease) over the same intervals, but possibly at different time-varying rates, for a given condition and state (Fig. 5). Consequently, the net switching model may be used, if accurate rates of change in subpopulation size are not imperative. However, net switching is generally more expensive than two-way switching due to the non-convex constraints (5) that characterize the former model. A less costly alternative to the bootstrap algorithm of Sec. V may be required to assess parameter variation due to process noise, if net switching is adopted in future work.

## VII. Initial Steps for Controller Design

The long-term goal of this work is to improve the trial-and-error paradigm used in the laboratory to test anti-cancer drugs. Fortunately, the simple structure of our model

TABLE III

MAXIMUM GENERALIZATION ERROR (SAMPLE EXECUTION)

| Condition | Two-way switching | Net switching |
|---|---|---|
| Control | 154 | 673 |
| MEKi | 153 | 371 |
| PI3K/mTORi | 157 | 1180 |
| Combination | 157 | 467 |

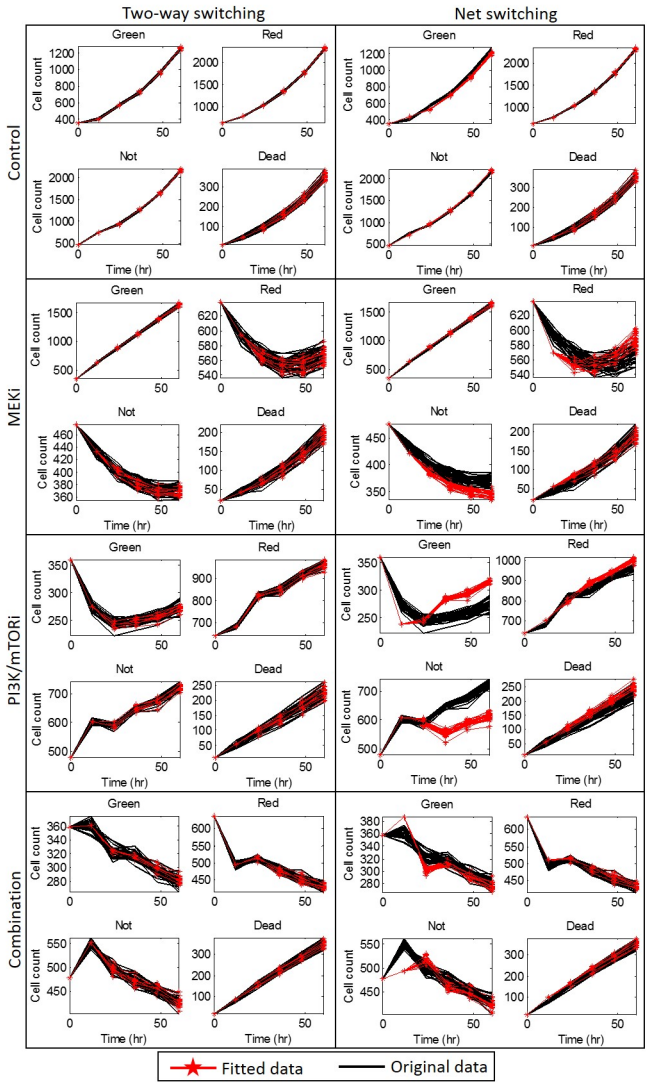


Fig. 5. Synthetic test data and fitted data for two-way switching (left) and net switching (right) are shown for each condition from a sample trial. Trajectories from all 33 synthetic wells are plotted to illuminate the assumed error distribution. Black lines are synthetic test data, called "original" in the legend, and red starred lines are fitted data. The net and two-way switching fitted data generally increase (or decrease) over the same intervals, but possibly at different time-varying rates, for a given condition and state.

naturally motivates a control-theoretic approach to identify potential combination treatments *in silico* to guide future experiments. Specifically, the control input is the drug perturbation, and the quantities of cells in each type (green, red, not, and dead) need to be controlled. Matrix $A$ of (1) represents the phenotypic state dynamics of the cancer system subject to the control input, as the drug-induced system response was used to identify these dynamics. Further, a finite composition of dynamics matrices, each associated with a drug perturbation, represents the phenotypic state dynamics subject to a sequence of control inputs over time. This motivates the goal to identify a therapeutic schedule that is projected to weaken resistance of malignant populations *in silico*. One way to weaken resistance may be to drive the cancer system to a more homogeneous composition, in which one phenotypic state with special properties (e.g., particular drug sensitivity) is dominant. The questions of interest are:

- Using available data, how can we identify a finite sequence of drugs that drives subpopulation sizes of all but one phenotypic state sufficiently close to zero?
- More generally, what mathematical conditions guarantee existence of such a sequence?

An initial step for controller design is to examine the eigenvectors and eigenvalues of the dynamics matrices identified for each therapeutic condition via (4). For example, consider the dynamics matrix identified for MEKi and the eigenvector-eigenvalue pairs of the phenotypic states. Only one phenotypic state eigenvalue is unstable (outside the unit circle), and its corresponding eigenvector is $(0.94, 0.21, 0.12, 0.25)^T$, rounded to two decimal places. Notice that the first element of the eigenvector is much larger than the second and third elements. Hence, the phenotypic state associated with the first element (green) will tend to dominate the remaining states. This motivates a potential therapeutic schedule: first, apply MEKi daily for many days to expand the green subpopulation; then, apply a drug that selectively kills green cells. For example, applying MEKi once per day for ten days from a plausible initial condition (the median of all available realizations at time 0) produces the composition $(3300, 800, 480, 550)^T$, in which roughly 64% of the cell population belongs to the green state. Among other challenges, sufficient time horizon length and existence of a drug that selects green cells are unknown. Yet, we show that our data-driven systems-based framework has the potential to identify new temporal treatments computationally in order to direct future *in vitro* cancer studies.

## VIII. FUTURE WORK

*Model extension.* The current model predicts cytotoxic effect on subpopulation-level behavior and facilitates a natural framework to control populations of *any* cancer with high phenotypic state heterogeneity. However, the representation of stochasticity could be improved by introducing parameters that signify cell-type-specific variance of process noise. Strong evidence for the prominence of noise in biological systems [23][24][25][26] motivates this extension.

TABLE IV

CELL DIVISION AND DEATH GAINS: IDENTIFIED VALUES AND CIS

| | Division | $\rho_{GD}$ | $\rho_{RD}$ | $\rho_{ND}$ |
|---|---|---|---|---|
| Control | 1.34; [1.30, 1.39] | 0.05; [0, 0.14] | 0; [0, 0.08] | 0.04; [0, 0.10] |
| MEKi | 1.14; [1.09, 1.18] | 0.03; [0, 0.06] | 0; [0, 0.04] | 0.02; [0, 0.12] |
| PI3K/mTORi | 1.08; [1.00, 1.07] | 0.07; [0, 0.16] | 0.02; [0, 0.06] | 0.02; [0, 0.09] |
| Combination | 1.00; [1.00, 1.00] | 0; [0, 0.21] | 0.01; [0, 0.13] | 0.13; [0, 0.18] |

TABLE V

SWITCHING GAINS: IDENTIFED VALUES AND CIS

| | $\rho_{GR}$ | $\rho_{GN}$ | $\rho_{RG}$ | $\rho_{RN}$ | $\rho_{NG}$ | $\rho_{NR}$ |
|---|---|---|---|---|---|---|
| Ctrl. | 0.29; [.04, 1] | 0.99; [0, 1] | 0.28; [0, .48] | 0.39; [.21; .91] | 0.46; [.20, .59] | 0.52; [.27, .72] |
| MEKi | 0.06; [0, .20] | 0; [0, .20] | 0; [0, .61] | 0.56; [.06, .66] | 0.53; [0, .81] | 0.42; [.13, .77] |
| PI3K. | 0; [0, .97] | 0.55; [.01, .94] | 0.16; [0, .30] | 0.57; [.17, .66] | 0; [0, .40] | 0.98; [.20, .94] |
| Comb. | 0.10; [0, .84] | 0.85; [0, .88] | 0.38; [.07, .50] | 0.39; [.18, .74] | 0.22; [.05, .49] | 0.65; [.12, .64] |

*Statistical analysis.* Values of the identified gains can fall outside the bootstrap intervals (e.g., see Table V-PI3K/mTORi-$\rho_{NR}$). To mitigate this, we may implement bias-corrected bootstrap in subsequent work.

*Data quality and quantity.* Since phenotypic state dynamics are hard to quantify using modern empirical tools, our data set has reduced quality and size compared to those typically used in systems research, and our model was not validated on real data. Fortunately, time series data from more replicate wells will be available in the future.

*More phenotypic states.* In this paper, cells expressing both basal and mesenchymal markers were improperly classified in both green and red states. A larger data set may enable including more states in the model to correct this issue.

*Experimental validation.* Potentially, a fluorescent reporter system, which would enable cells to fluoresce different colors based on phenotypic expression, could facilitate real-time tracking of single-cell state-specific switching and death.

## ACKNOWLEDGMENT

The authors thank Walid Krichene and Max Balandat for their assistance in probability theory and statistics, respectively. This work is supported in part by NIH under the Stanford Center for Systems Biology, and by NIH under grant 1P01GM09955-01. M.C. is supported by the NSF Graduate Research Fellowship Program.

## APPENDIX

Numerical values of identified gains (Fig. 3) and confidence intervals (CIs) (Fig. 4) are provided in Tables IV and V, rounded to two decimal places.

## REFERENCES

[1] M. D. Brooks *et al.*, "Therapeutic implications of cellular heterogeneity and plasticity in breast cancer," *Cell Stem Cell*, vol. 17, no. 3, pp. 260–271, 2015.

[2] L. Melchor *et al.*, "Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models," *The Journal of Pathology*, vol. 233, no. 2, pp. 124–137, 2014.

[3] H. C. Bhang *et al.*, "Studying clonal dynamics in response to cancer therapy using high-complexity barcoding," *Nature Medicine*, vol. 21, no. 5, pp. 440–448, 2015.

[4] S. Nik-Zainal *et al.*, "The life history of 21 breast cancers," *Cell*, vol. 162, no. 4, p. 924, 2015.

[5] S. Misale *et al.*, "Emergence of kras mutations and acquired resistance to anti-egfr therapy in colorectal cancer," *Nature*, vol. 486, no. 7404, pp. 532–536, 2012.

[6] F. Reim *et al.*, "Immunoselection of breast and ovarian cancer cells with trastuzumab and natural killer cells: selective escape of cd44high/cd24low/her2low breast cancer stem cells," *Cancer Research*, vol. 69, no. 20, pp. 8058–8066, 2009.

[7] C. J. Creighton *et al.*, "Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features," *Proceedings of the National Academy of Sciences*, vol. 106, no. 33, pp. 13 820–13 825, 2009.

[8] H. Easwaran *et al.*, "Cancer epigenetics: Tumor heterogeneity, plasticity of stem-like states, and drug resistance," *Molecular Cell*, vol. 54, no. 5, pp. 716–727, 2014.

[9] A. Goldman *et al.*, "Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition," *Nature Communications*, vol. 6, no. 6139, 2015.

[10] A. O. Pisco *et al.*, "Non-darwinian dynamics in therapy-induced cancer drug resistance," *Nature Communications*, vol. 4, no. 2467, 2013.

[11] Q. Q. Li *et al.*, "Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells," *Clinical Cancer Research*, vol. 15, no. 8, pp. 2657–2665, 2009.

[12] P. B. Gupta *et al.*, "Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells," *Cell*, vol. 146, no. 4, pp. 633–644, 2011.

[13] D. Cacchiarelli *et al.*, "Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency," *Cell*, vol. 162, no. 2, pp. 412–424, 2015.

[14] A. Fischer *et al.*, "High-definition reconstruction of clonal composition in cancer," *Cell Reports*, vol. 7, no. 5, pp. 1740–1752, 2014.

[15] I. Bozic and M. A. Nowak, "Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers," *Proceedings of the National Academy of Sciences*, vol. 111, no. 45, pp. 15 964–15 968, 2014.

[16] V. Almendro *et al.*, "Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity," *Cell Reports*, vol. 6, no. 3, pp. 514–527, 2014.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.

[19] C.-F. J. Wu, "Jackknife, bootstrap and other resampling methods in regression analysis," *The Annals of Statistics*, vol. 14, no. 4, pp. 1261–1295, 1986.

[20] W. Hardle and E. Mammen, "Comparing nonparametric versus parametric regression fits," *The Annals of Statistics*, vol. 21, no. 4, pp. 1926–1947, 1993.

[21] D. Klevebring *et al.*, "Sequencing of breast cancer stem cell populations indicates a dynamic conversion between differentiation states in vivo," *Breast Cancer Research*, vol. 16, no. 4, p. R72, 2014.

[22] M. R. Doherty *et al.*, "Cancer stem cell plasticity drives therapeutic resistance," *Cancers*, vol. 8, no. 1, p. 8, 2016.

[23] H. H. McAdams and A. Arkin, "It's a noisy business! genetic regulation at the nanomolar scale," *Trends in Genetics*, vol. 15, no. 2, pp. 65–69, 1999.

[24] M. Hoffmann *et al.*, "Noise-driven stem cell and progenitor population dynamics," *PLoS One*, vol. 3, no. 8, p. e2922, 2008.

[25] C. V. Rao *et al.*, "Control, exploitation and tolerance of intracellular noise," *Nature*, vol. 420, no. 6912, pp. 231–237, 2002.

[26] M. S. Samoilov *et al.*, "From fluctuations to phenotypes: The physiology of noise," *Science Signaling*, vol. 2006, no. 366, p. re17, 2006.