Moment-Based Estimation of State-Switching Rates in Cell Populations

Michael Saint-Antoine[1] and Abhyudai Singh[2]

Abstract— In isogenic cell populations, cells can switch back and forth between different gene expression states. These expression states can be biologically relevant. For example, a certain expression state may cause a tumor cell to be resistant to treatment, while another state may leave it vulnerable to treatment. However, estimating the rates of state-switching can be difficult, because experimentally measuring a cell's transcriptome often involves destroying the cell, so it can only be measured once. In this paper, we propose a computational method to estimate the rate of switching between expression states, given data from a Luria-Delbrück style fluctuation test that is experimentally simple and feasible. We then benchmark this method using simulated data to test its efficacy, with varying assumptions made about cell cycle timing distribution in the simulations.

I. INTRODUCTION

Even in clonal groups of cells that share the same genetic code, cells sometimes exhibit different phenotypic characteristics, caused by different patterns of gene expression [1]. This phenomenon is referred to as "cellular heterogeneity", and can arise due to non-genetic factors affecting gene expression, such as epigenetic regulation and stochastic variability in the processes of transcription and translation [2]–[14]. Recent developments in experimental technology, such as single-cell RNA-seq (scRNA-seq) have made it feasible to study the transcriptomic profiles of individual cells, yielding new insights about cellular heterogeneity at the single-cell level [15]. Heterogeneity in cell populations is known to play a role in many biological contexts, including drug response in cancer [16]–[18], latency in HIV cells [19], [20], immune response in epithelial tissue [21], determination of cell fate in genetically identical populations [22]–[31], and "bet-hedging" responses to environmental stresses in bacteria [32], [33]. For more information, an excellent review of the study of cellular heterogeneity can be found in the Altschuler and Wu (2010) paper "Cellular Heterogeneity: When Do Differences Make a Difference?" [34].

We are especially interested in the phenotypic state-switching dynamics associated with transient cellular heterogeneity – that is, when cells switch back and forth between different phenotypic states. A fascinating example of this is related to drug resistance in cancer. Recent research has uncovered a phenomenon in which melanoma cells switch back and forth between a common drug-sensitive state and a rare pre-resistant state [35], [36]. When the drug is administered, cells in the sensitive state die off, while cells in the pre-resistant state undergo a process in which they become "locked in" to a state of permanent drug resistance, and no longer switch back to the sensitive state [35], [36]. Similar phenomena have been observed in many different types of cancer [37]–[40] and their significance for treatment scheduling has been explored in theoretical models [41], [42].

In cases like this, where cells switch back and forth between different phenotypic states, it can be quite challenging to estimate the rates of switching between the states. This is because experimentally measuring a cell's transcriptome often involves destroying the cell, so it can only be measured once for each cell. In this paper, we will propose a computational method to estimate the rate of switching between expression states that is experimentally simple and feasible.

II. EXPERIMENTAL SETUP

In 1943, researchers Salvador Luria and Max Delbrück used a fluctuation test experiment to investigate whether bacterial resistance to T1 phage occurred spontaneously or was induced by the virus [43]. The fluctuation test experiment involved growing out many colonies of bacteria, and then exposing each colony to the phage. Luria and Delbrück counted the number of surviving bacteria in each colony, and studied their distribution. They realized that if resistance were being induced by the virus, the numbers of surviving bacteria would follow a Poisson distribution. However, they found instead that the distribution of surviving bacteria showed much greater variance, suggesting that resistance to the virus was caused by a mutation that occurred prior to exposure to the virus. The results of this experiment were evidence of Darwinian selection acting on bacteria by way of random mutations that confer a fitness advantage. Since then, the experiment has been analyzed theoretically, with researchers deriving probability distributions for the number of resistant cells based on different biological assumptions [44]–[46]. An excellent review of some of the theoretical developments is available in the Zheng 1999 paper "Progress of a half century in the study of the Luria-Delbrück distribution" [47].

We can use a modified fluctuation test to study the rate at which cells switch between transient phenotypic states. Consider a population of cells that switch back and forth between two expression states, an ON state and an OFF state. This modified fluctuation test will involve a workflow similar to the original Luria-Delbrück experiment. First, several individual cells will be isolated from the population. Then, each of these isolated cells will be grown out into its

[1]Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE USA 19716. `mikest@udel.edu`

[2]Department of Electrical and Computer Engineering, Biomedical Engineering, Mathematical Sciences, Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE USA 19716. `absingh@udel.edu`
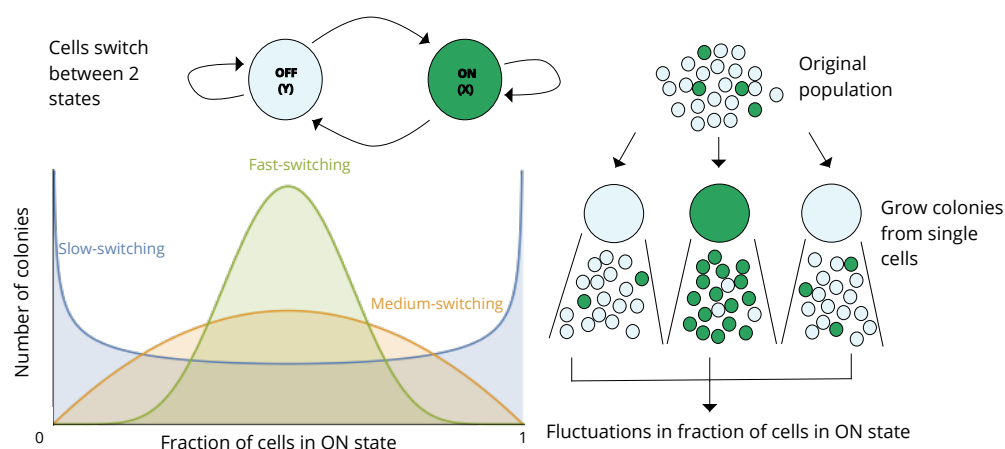
Fig. 1. Workflow of the modified fluctuation test experiment. We begin with an original population of cells that switch back and forth between an OFF (Y) state and an ON (X) state. We isolate many individual cells, and grow each one out into a colony. After some time, we measure the fraction of cells in the ON state in each colony, and calculate the CV^2 in these fractions across colonies. All else equal, populations of cells with faster state-switching rates will tend to have lower variation in these fractions than populations of cells with slower state-switching rates, since the faster-switching populations will tend to converge back to the steady state fraction of ON cells in the original population more quickly.

own colony. After a set amount of time, the fraction of ON cells in each colony will be recorded. This may be done by transcriptome sequencing, using methods such as RNA-Seq. Or, in cases where the ON state corresponds to resistance to a drug or virus, the colonies could simply be exposed to the drug or virus and the survivors counted. Please see Figure 1 for a graphical representation of this experiment.

Once we have recorded the fraction of ON cells in each colony, we can then look at the coefficient of variation squared (CV^2) of these fractions in order to infer the rates of switching between the ON and OFF states. Slow rates of switching between ON and OFF states will tend to yield higher CV^2, while faster rates of switching will tend to yield lower CV^2. The intuition here is that each of the colonies will eventually converge back to the original population average of cells in the ON state. However, faster-switching populations will converge back more quickly, leading to lower CV^2 in fractions across colonies, and the reverse is true for slower-switching populations.

Variations of this modified fluctuation test have previously been used to study HIV [19] and cancer [41], [48], [49]. The test has also been studied theoretically through mathematical analysis [50]. However, to our knowledge, there has not yet been an attempt to benchmark this method on simulated "gold standard" data, for which the true switching rates are already known. As we will see in the next section, the mathematical formula used in this technique is an inexact approximation. So, it is a pressing issue to validate this method through a benchmarking process so that we can be confident in its accuracy as it continues to be used on real-world experimental datasets.

III. FORMULA DERIVATION

In this section, we will derive a formula for the CV^2 of the fractions of ON cells across colonies in the modified

fluctuation test experiment. Let us again consider a population in which cells switch back and forth between an OFF state, which we will call Y, and an ON state, which we will call X. Please see Figure 2 for a graphical representation of these dynamics.
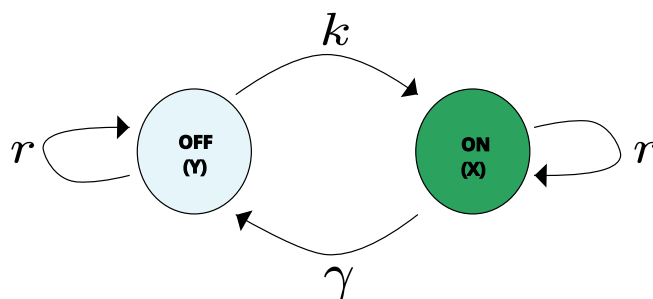

Fig. 2. Cells switch back and forth between states OFF (Y) and ON (X). They switch from OFF to ON with rate k, and from ON to OFF with rate γ. Cells in both states proliferate with rate r.

We assume that the cell cycle time is an exponentially-distributed random variable with a mean generation time of $1/r$. Hence, starting from a single cell the population grows exponentially with rate r. We make two simplifying assumptions: i) The proliferation rate of a cell is the same irrespective of the cellular state; ii) The population remains in the exponential growth phase during the timespan of the experiment. Cells in the OFF state transition to the ON state with rate k, and switch back to the OFF state with rate γ. So, the expected amount of time a cell will spend in the OFF state is $1/k$, and the expected amount of time a cell will spend in the ON state is $1/\gamma$. Since we have defined the cell proliferation rate as r, the expected number of generations a cell will spend in the OFF state is r/k, and the expected number of generations a cell will spend in the ON state is r/γ.

When discussing the fraction of cells in the ON state, we will use f to refer to the average fraction of ON cells in the original population, and $F(t)$ to refer to the fraction of cells in the ON state in cell colonies over time.

f can be written in terms of the state switching rates k and γ as

$$f = \frac{k}{k + \gamma} \quad (1)$$

Individual cells are randomly chosen from the population, and they will be in the ON state with probability f and in the OFF state with probability $1 - f$. Let $X(t)$ and $Y(t)$ denote the number of cells within the colony in the ON state and OFF state, respectively, at time t. The time evolution of integer-valued random processes $X(t)$ and $Y(t)$ is governed by the model illustrated in Table I.

TABLE I

STOCHASTIC MODEL OF CELL PROLIFERATION AND SWITCHING

Stochastic event description	Reset in population counts	Probability event occurs in small time interval $(\mathbf{t}, \mathbf{t} + \mathbf{dt})$
ON cell proliferation	$X \to X + 1$	$rXdt$
OFF cell proliferation	$Y \to Y + 1$	$rYdt$
Cell switching OFF to ON	$Y \to Y - 1$ $X \to X + 1$	$kYdt$
Cell switching ON to OFF	$X \to X - 1$ $Y \to Y + 1$	γXdt

The model consists of four events that occur probabilistically with rates given in the third column, and whenever the event occurs, the cell numbers increase/decrease by one as per the reset map in the second column. Based on this Markov process, we can derive the time evolution of the joint probability density function of $X(t)$ and $Y(t)$ that evolves according to the following Chemical Master Equation (CME)

$$\frac{dP_{i,j}(t)}{dt} = r(i - 1)P_{i-1,j}(t) + r(j - 1)P_{i,j-1}(t)$$
$$+ k(i + 1)P_{i+1,j-1}(t) + \gamma(j + 1)P_{i-1,j+1}(t)$$
$$- P_{i,j}(t)(ri + rj + ki + \gamma j) \quad (2)$$

where $P_{i,j}(t)$ denotes the probability of observing i cells in the OFF state, and j cells in the ON state at time t [51], [52]. Previous research has shown how the time evolution of moments can be derived from the CME [53]–[56].

We are interested in the ratio

$$F(t) = \frac{X(t)}{X(t) + Y(t)} \quad (3)$$

that represents the fraction of cells in the ON state at time t, and our goal is to quantify statistical fluctuations in this ratio across colonies. To do this, we derive the time evolution of the first two statistical moments of $X(t)$ as well as those of the total cell count, which we will refer to as $T(t) = X(t) + Y(t)$. For the sake of simplicity, we will refer to the

random variables $X(t)$, $Y(t)$, $T(t)$, $F(t)$, as simply X, Y, T, and F from now on.

We will use angle brackets to denote the expected value of a random variable. For example, $\langle X \rangle$ will denote the expected value of X, and $\langle X^m \rangle$ will denote the expected value of X^m.

The following equation is a general formula for the moment dynamics ODE of this system [53].

$$\frac{d\langle X^m Y^n \rangle}{dt} = \langle (rX)((X + 1)^m Y^n - X^m Y^n)$$
$$+ \gamma X((X - 1)^m (Y + 1)^n - X^m Y^n)$$
$$+ (rY)(X^m (Y + 1)^n - X^m Y^n)$$
$$+ kY((X + 1)^m (Y - 1)^n - X^m Y^n) \rangle \quad (4)$$

We can use this general formula to find differential equations for the first moments of X and Y, which are shown below.

$$\frac{d\langle X \rangle}{dt} = r\langle X \rangle + k\langle Y \rangle - \gamma\langle X \rangle \quad (5)$$

$$\frac{d\langle Y \rangle}{dt} = r\langle Y \rangle + \gamma\langle X \rangle - k\langle Y \rangle \quad (6)$$

To define the initial conditions for this set of ODEs, we must consider the state of our experiment at time $t = 0$. At time $t = 0$, each colony is still only a single cell that has been sampled from the original population. If this individual cell is ON, then the fraction of cells in the ON state will be 1. If this cell is OFF, then the fraction of cells in the ON state will be 0. In other words, the fraction of cells in the ON state at time $t = 0$ follows a Bernoulli distribution, taking the value 1 with probability $p(1) = f$, and taking the value 0 with probability $p(0) = 1 - f$. So, at time $t = 0$, $\langle X \rangle = f$ and $\langle Y \rangle = 1 - f$. With this information, we can solve this set of ODEs and find equations for $\langle X \rangle$ and $\langle Y \rangle$ as functions of time. These are shown below.

$$\langle X \rangle = \frac{e^{rt} k}{(k + \gamma)} \quad (7)$$

$$\langle Y \rangle = \frac{e^{rt} \gamma}{(k + \gamma)} \quad (8)$$

Adding equations (7) and (8) together yields an expression for $\langle T \rangle$, the total number of cells in both states.

$$\langle T \rangle = e^{rt} \quad (9)$$

To compute the second moments of X and Y, we again use the differential equation technique described in equation (4). After simplifying, this yields:

$$\frac{d\langle X^2 \rangle}{dt} = 2r\langle X^2 \rangle + 2k\langle XY \rangle + r\langle X \rangle + k\langle Y \rangle - 2\gamma\langle X^2 \rangle + \gamma\langle X \rangle \quad (10)$$

$$\frac{d\langle Y^2 \rangle}{dt} = 2r\langle Y^2 \rangle + 2\gamma\langle XY \rangle + r\langle Y \rangle + \gamma\langle X \rangle - 2k\langle Y^2 \rangle + k\langle Y \rangle \tag{11}$$

$$\frac{d\langle XY \rangle}{dt} = 2r\langle XY \rangle - \gamma\langle XY \rangle - k\langle XY \rangle$$
$$+ \gamma\langle X^2 \rangle + k\langle Y^2 \rangle - \gamma\langle X \rangle - k\langle Y \rangle \tag{12}$$

Again, we know the initial conditions for this set of ODEs: at time $t = 0$, $\langle X^2 \rangle = f$, $\langle Y^2 \rangle = 1 - f$, and $\langle XY \rangle = 0$. Using this information, we can again solve these ODEs and find the second order moments as functions of time.

Equation (13), shown at the top of the next page because of its length, is the equation for $\langle X^2 \rangle$. We have omitted $\langle Y^2 \rangle$ and $\langle XY \rangle$ for now, as they are not needed for our fluctuation test analysis. To find the second moment of T, we will again use the differential equation method described in equation (4). However, this time, we will ignore state-switching dynamics and consider only cell population growth, since T is the total number of all cells in the population.

$$\frac{d\langle T^2 \rangle}{dt} = 2r\langle T^2 \rangle + r\langle T \rangle \tag{14}$$

We know the initial condition at time $t = 0$ is $\langle T^2 \rangle = 1$. After solving the ODE with this initial condition, the resulting equation for $\langle T^2 \rangle$ is shown below.

$$\langle T^2 \rangle = 2e^{2rt} - e^{rt} \tag{15}$$

Based on equations (9) and (15), we note that the total number of cells T at a given timepoint follows a geometric distribution.

We have now derived equations for the first and second order moments of X and T. However, in order for this method to be experimentally and computationally feasible, what we are actually interested in is the first and second order moments of the fraction $F = \frac{X}{T}$, which is the fraction of cells in the ON state over time. If we can find the first and second order moments of this fraction, then we can use these moments to compute the CV^2 of the fraction of ON cells across colonies as a function of time.

How can we compute the moments of F? Please consider the following:

$$\langle X \rangle = \langle FT \rangle \tag{16}$$

$$\langle X^2 \rangle = \langle F^2 T^2 \rangle \tag{17}$$

At this point, we will introduce a simplifying assumption that will allow us to derive an approximation of the moments $\langle F \rangle$ and $\langle F^2 \rangle$. We will assume that F and T are independent variables. In other words, the fraction of ON cells is independent of the total number of cells in the colony. In the next section of the paper, we will discuss how accurate

this assumption is. Making this assumption, we can rewrite equations (16) and (17) as, respectively:

$$\langle X \rangle = \langle F \rangle\langle T \rangle \tag{18}$$

$$\langle X^2 \rangle = \langle F^2 \rangle\langle T^2 \rangle \tag{19}$$

Then, simple algebra yields the first and second order moments of F:

$$\langle F \rangle = \frac{\langle X \rangle}{\langle T \rangle} \tag{20}$$

$$\langle F^2 \rangle = \frac{\langle X^2 \rangle}{\langle T^2 \rangle} \tag{21}$$

Please note that we have already derived $\langle X \rangle$, $\langle T \rangle$, $\langle X^2 \rangle$, and $\langle T^2 \rangle$ in equations (7), (9), (13), and (15) respectively. So, we can simply plug in those formulas to get the first and second order moments of F in terms of k, γ, and t. Equation (22) is the first moment of F, which is equal to f, the fraction of ON cells in the original population. Equation (23), shown on the following page because of its size, is the second moment of F.

$$\langle F \rangle = \frac{k}{k + \gamma} \tag{22}$$

Finally, now that we have calculated the first and second order moments of the fraction of ON cells, we can calculate the coefficient of variation squared as follows:

$$CV^2 = \frac{\langle F^2 \rangle - \langle F \rangle^2}{\langle F \rangle^2} \tag{24}$$

The full formula for CV^2 is a bit long and unwieldy, so we will define two new parameters to be able to write it more simply. Let $Z = \frac{1-f}{\gamma}$, and let $\tau = tr$, so that τ is the duration of the experiment normalized to generation time. Then we can write the formula for CV^2 as follows.

$$CV^2 = \left(\frac{2e^{\tau - \frac{2\tau}{Z}}Z - 2 - Z}{(2e^\tau - 1)(Z - 2)} \right) \frac{1 - f}{f} \tag{25}$$

Equation 26 shows the limit of our CV^2 formula as t approaches 0. To understand this, consider that at time $t = 0$, each colony is still only a single cell that has been sampled from the original population. So, the fraction of cells in the ON state will take the value 1 with probability $p(1) = f$, and will take the value 0 with probability $p(0) = 1 - f$. In other words, at time $t = 0$, the fraction of cells in the ON state follows a Bernoulli distribution, and the CV^2 of this distribution is $\frac{1-f}{f}$.

$$\lim_{t \to 0} CV^2 = \frac{1 - f}{f} \tag{26}$$

Conversely, as t goes to infinity, the fraction of ON cells in each of the colonies tends back towards the steady state value of f in the original population, and the CV^2 of the fraction of ON cells in the colonies approaches 0.

$$\langle X^2 \rangle = \frac{ke^{-2t(\gamma+k-r)}\left(2k(2(\gamma+k)-r)e^{2t(\gamma+k)} + (\gamma+k)(2\gamma-2k+r)e^{t(2(\gamma+k)-r)} - 2\gamma r\right)}{(\gamma+k)^2(2(\gamma+k)-r)} \tag{13}$$

$$\langle F^2 \rangle = \frac{k\left(2\gamma^2 + \gamma\left(-2re^{t(r-2(\gamma+k))} + 4ke^{rt} + r\right) + k(2k-r)\left(2e^{rt}-1\right)\right)}{(\gamma+k)^2\left(2e^{rt}-1\right)(2(\gamma+k)-r)} \tag{23}$$
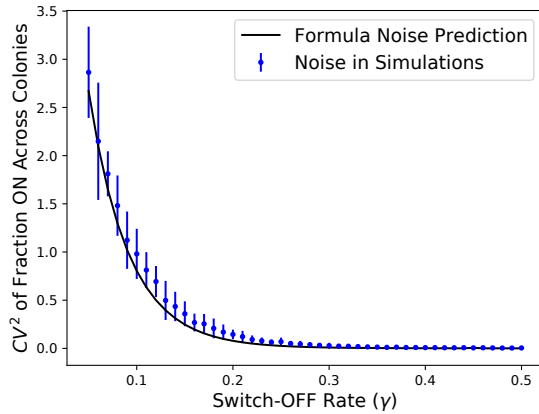

Fig. 3. Comparison between formula noise prediction and noise in simulations. Simulated data was generated with the Gillespie algorithm method, further described in Section V. The fraction of ON cells in the original population (f) was assumed to be 0.10. We set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. The simulations were run for 12 generations ($t = 12$), and CV^2 was calculated across 40 colonies. The simulations were run 50 times per parameter set to generate error bars, which show one standard deviation. Although our formula is not a perfect prediction of the true underlying CV^2, it is a reasonably good approximation.

$$\lim_{t\to\infty} CV^2 = 0 \tag{27}$$

Importantly, at a fixed experimental endpoint t, cell populations with slower switching rates will tend to have a higher CV^2 in the fraction of ON cells across colonies compared to cell populations with faster switching rates. The intuition behind this is that colonies from populations with faster switching rates will tend to converge back to the steady state fraction of ON cells more quickly, leading to less variation across colonies, compared those of a slower-switching population.

Figure 3 shows a comparison of our formula's CV^2 predictions and CV^2 of stochastic simulations of the experiment. Now that we have a formula for the CV^2 of the fraction of ON cells, we can measure the CV^2 of the fraction of ON cells in real cell populations, and compare the formula to the real measurement to fit the state-switching parameters k and γ. Please note that, based on equation (1), if we experimentally measure the fraction of cells in the ON state in the original population, we can write k in terms of γ and this measured fraction.

$$k = \frac{\gamma}{\frac{1}{f}-1} \tag{28}$$

So, if we know the experimentally-measured CV^2, the fraction of ON cells in the original population f, the duration of the experiment t, and have written k in terms of γ, then we are left with only one unknown variable, γ, that we can solve for with a simple error-minimization between the derived CV^2 formula and the experimentally-measured CV^2.

$$\min_{\gamma} \quad |CV^2(\gamma) - CV^2_{data}| \tag{29}$$

In practice, we use the Python SciPy library's **optimize.fmin()** function to do this error minimization and solve for γ.

IV. INDEPENDENT VARIABLE ASSUMPTION

As noted previously, in order to derive the formulas for the first and second order moments of F, we made a simplifying assumption: that F and T are independent variables (which would mean that F^2 and T^2 are also independent variables). To be precise, writing equation (16) as equation (18) is based on the assumption that F and T are independent, and writing equation (17) as equation (19) is based on the assumption that F^2 and T^2 are independent.

However, we tested this assumption by collecting data from stochastic simulations (discussed further in the next section), and found that it is not completely correct. For these simulations, we set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. The fraction of ON cells in the original population (f) was assumed to be 0.10. The switch-off rate (γ) was set at $1/8$, so that the expected number of generations spent in the ON state is 8. The switch-on rate (k) was set according to equation (28), using these values of f and γ. The simulations were run for 7 generations ($t = 7$). We ran $10,000$ simulations, and at the end of each one we recorded the total number of cells (T) and the fraction of cells in the ON state (F), as well as the squares of both (F^2 and T^2).

We observed a very slight Pearson correlation of -0.016 between F and T, which was not statistically significant ($p = 0.1087$). However, we observed a stronger Pearson correlation of -0.0439 between F^2 and T^2, which was statistically significant ($p = 1.1078 \times 10^{-5}$). This negative correlation is shown in Figure 4, a plot of the final fraction of cells in the ON state squared (F^2) against the total number of
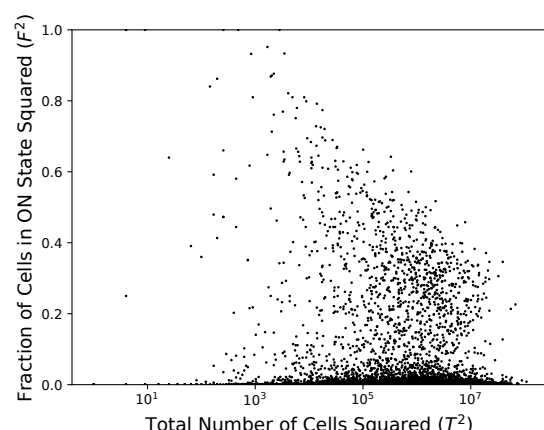
compared to the deterministic steady state fraction ON.

Fig. 4. Fraction of cells in the ON state squared F^2 plotted against the total number of cells squared T^2 for 10,000 simulated cell colonies. For these simulations, we set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. The fraction of ON cells in the original population (f) was assumed to be 0.10. The switch-off rate (γ) was set at $1/8$, so that the expected number of generations spent in the ON state is 8. The switch-on rate (k) was set according to equation (28), using these values of f and γ. The simulations were run for 7 generations ($t = 7$). We ran $10,000$ simulations, and at the end of each one we recorded the total number of cells squared (T^2) and the fraction of cells in the ON state squared (F^2). We observed a Pearson correlation of -0.0439 between F^2 and T^2, which was statistically significant ($p = 1.1078 \times 10^{-5}$).

cells squared (T^2) for $10,000$ simulated cell colonies. This observed statistical dependence between F^2 and T^2 means that there must also be some level of dependence between F and T, even if it was not detected to a statistically significant extent in our simulation analysis.

So, although the derivation of our CV^2 formula was based on a simplifying assumption of independence between F and T, it appears that there is in fact a very slight level of statistical dependence between them. Therefore, our derived formula for the CV^2 should be thought of as an approximation, not as an exact analytical formula for the experimental CV^2. Can this approximation of the CV^2 still be useful in a biological context, even if it is not exact? In the following sections, we will benchmark our method on simulated data to test its efficacy.

V. SIMULATING CELL COLONIES

A. Simulations with Gillespie Algorithm

In order to test if our fluctuation test method, based on the approximation of the CV^2, is useful, we first generated simulated data by using the Gillespie algorithm [57] to simulate the stochastic model described previously in Table I.

Please note that for all of our simulations, we set the proliferation rate $r = 1$, so the average cell cycle time $1/r$ will also be 1 and time is normalized to the generation time of the cells. This will allow us to discuss results for $1/\gamma$ both as predicted time in the ON state and predicted number of generations in the ON state.

Figure 5 shows the evolution of the fraction ON, $F = \frac{X}{T}$, over time in the stochastic simulations (blue dotted curves),
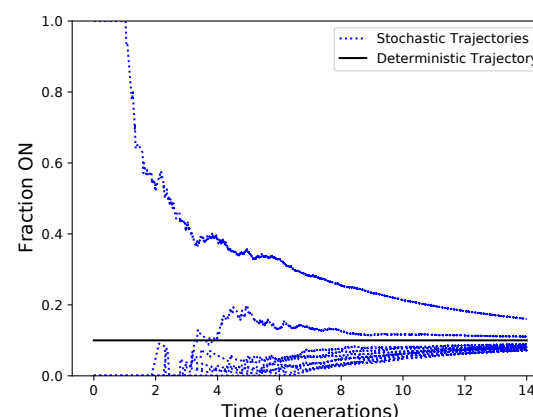

Fig. 5. This plot shows the evolution of the fraction ON, $F = \frac{X}{T}$, over time in the stochastic simulations (blue dotted curves), compared to the deterministic steady state fraction ON. The fraction of ON cells in the original population (f) was assumed to be 0.10. We set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. Most colonies begin with a single OFF cell, and the fraction ON increases up to 0.10 over time as the colony grows. A small number of colonies begin with a single ON cells, and the fraction ON decreases down to 0.10 over time as the colony grows.

To simulate a colony of cells, we will run this simulation starting with one cell that will be randomly chosen to either be in the X (ON) state with probability f, or the Y (OFF) state with probability $1 - f$. Then we will run the simulation to some specified end point to simulate the colony of cells proliferating and switching between the ON and OFF states. To simulate the modified fluctuation test experiment, we will simulate many of these colonies so that we can measure the CV^2 of the final fraction of ON cells in each colony.

B. Non-Exponential Cell Cycle Time

In our simulations with the Gillespie algorithm, it is assumed that the timing between events follows an exponential distribution. However, in the biological context, cell cycle timing does not typically follow an exponential distribution. More realistic distributions for modeling cell cycle timing are the log-normal distribution and gamma distribution. So, in addition to using the Gillespie algorithm, we also developed an algorithm for simulating cell colonies that can account for other distributions of cell cycle timing besides the exponential distribution. More information about these simulations is available in the Appendix section of this paper, and the code for these simulations is available upon request from the authors.

VI. RESULTS

As discussed previously, the formula we derived for the CV^2 of the fraction of ON cells across colonies over time is an inexact approximation. Is this approximation accurate enough to be useful to biologists in an experimental setting? To investigate this, we decided to benchmark our formula by generating simulated data and seeing how effective our

method was at estimating the true, underlying rates of state-switching.

We considered two hypothetical cell populations: a fast-switching population and a slow-switching population. Both populations have an average of 10% ON cells in the original populations, so $f = 0.10$. However, cells in the fast-switching population spend an expected 4 generations in the ON state, while cells in the slow-switching population spend an expected 10 generations in the ON state, so $\gamma = 1/4$ for the fast-switching population and $\gamma = 1/10$ for the slow-switching population. We assumed that our simulated experiment would run for 12 generations ($t = 12$), and CV^2 would be calculated across 40 colonies. We set the proliferation rate $r = 1$, so the average cell cycle time $1/r$ will also be 1 and time is normalized to the generation time of the cells, allowing us to discuss results for $1/\gamma$ both as predicted time in the ON state and predicted number of generations in the ON state.

We used the stochastic simulation methods described in the previous section to simulate the experimental workflow of the fluctuation test, generating many simulated datasets for both the fast-switching and slow-switching scenarios. We then applied our switch-rate estimation method to each of these "experimental" datasets, to see how close the predicted switching rates came to the actual switching rates. To make our results more intuitive, we have chosen to report $1/\gamma$, which is the expected number of generations in the ON state, rather than γ itself.

The histograms in the following subsections show the predictions of $1/\gamma$ (expected number of generations in the ON state) for the fast-switching datasets in blue, and for the slow-switching datasets in red. There are also vertical lines at 4 and 10 on the horizontal axis, to mark the true, underlying values of $1/\gamma$ in the fast-switching and slow-switching scenarios, respectively.

A. Exponential Distribution Cell Cycle Time

Figure 6 shows the results of the benchmarking analysis, for cell colonies simulated with the Gillespie algorithm, assuming exponentially-distributed cell cycle time. As you can see, our method's predictions for the fast-switching scenario cluster around the true $1/\gamma$ value of 4 (meaning that cells spend an average of 4 generations in the ON state), while the predictions for the slow-switching scenario cluster around the true $1/\gamma$ value of 10 (meaning that cells spend an average of 10 generations in the ON state). Table II shows relevant statistics for the predictions. The mean predicted $1/\gamma$ value for the fast-switching scenario was 4.67, while the mean predicted $1/\gamma$ value for the slow-switching scenario was 11.01.

For both the fast-switching and slow-switching scenarios, our formula had a tendency to slightly overestimate the number of generations in the ON state. So, its performance was not perfect. However, we believe that it was still close enough to potentially be useful to biologists an approximation.
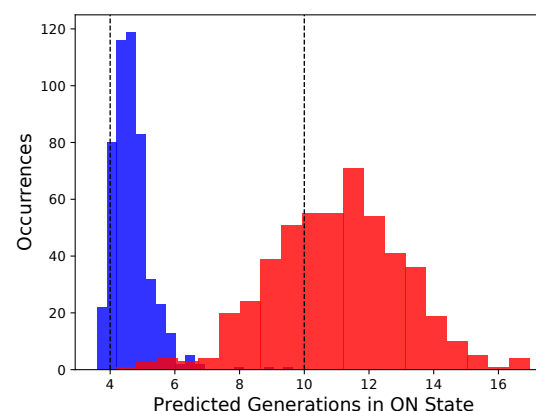

Fig. 6. Benchmarking with exponentially-distributed cell cycle time, using data generated with the Gillespie algorithm. The fraction of ON cells in the original population (f) was assumed to be 0.10. We consider a fast-switching scenario, in which cells spend an average of 4 generations in the ON state ($1/\gamma = 4$), and a slow-switching scenario, in which cells spend an average of 10 generations in the ON state ($1/\gamma = 10$). We set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. The simulations were run for 12 generations ($t = 12$), and CV^2 was calculated across 40 colonies. Our method's predictions for the fast-switching scenario (shown in blue) cluster around the true $1/\gamma$ value of 4 (shown with a dotted line), while the predictions for the slow-switching scenario (shown in red) cluster around the true $1/\gamma$ value of 10 (shown with a dotted line).

TABLE II

BENCHMARKING RESULTS – EXPONENTIAL DISTRIBUTION OF CELL CYCLE TIME

Scenario	True Generations ON	Predicted Generations ON	Prediction Std. Dev.
Fast-Switching	4	4.67	0.63
Slow-Switching	10	11.01	2.03

B. Log-Normal Distribution Cell Cycle Time

Figure 7 shows the results of the benchmarking analysis, assuming a log-normal distribution of cell cycle time. Again, we observe that our method's predictions for the fast-switching scenario cluster around the true $1/\gamma$ value of 4, while the predictions for the slow-switching scenario cluster around the true $1/\gamma$ value of 10. The mean predicted $1/\gamma$ value for the fast-switching scenario was 4.77, and the mean predicted $1/\gamma$ value for the slow-switching scenario was 10.97 (listed in Table III). Just as with the results in the previous sub-section, our formula again had a tendency to slightly overestimate the number of generations in the ON state in both the fast-switching and slow-switching scenarios. Again, we note that although our formula is not perfect, it may still be accurate enough to be useful to a biologist as an approximation.

C. Gamma Distribution Cell Cycle Time

Figure 8 shows the results of the benchmarking analysis, assuming a gamma distribution of cell cycle time. The mean predicted $1/\gamma$ value for the fast-switching scenario
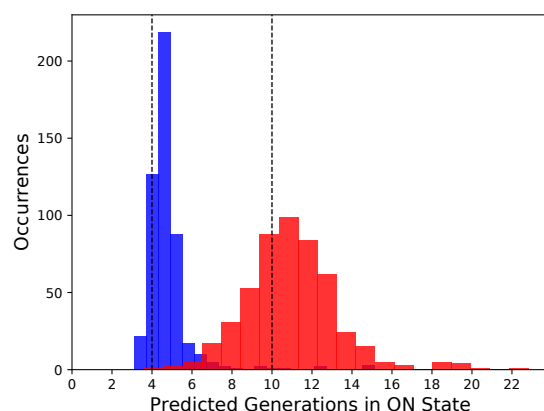
Fig. 7. Benchmarking with log-normal-distributed cell cycle time. The fraction of ON cells in the original population (f) was assumed to be 0.10. We consider a fast-switching scenario, in which cells spend an average of 4 generations in the ON state ($1/\gamma = 4$), and a slow-switching scenario, in which cells spend an average of 10 generations in the ON state ($1/\gamma = 10$). We set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. The simulations were run for 12 generations ($t = 12$), and CV^2 was calculated across 40 colonies. Our method's predictions for the fast-switching scenario (shown in blue) cluster around the true $1/\gamma$ value of 4 (shown with a dotted line), while the predictions for the slow-switching scenario (shown in red) cluster around the true $1/\gamma$ value of 10 (shown with a dotted line).
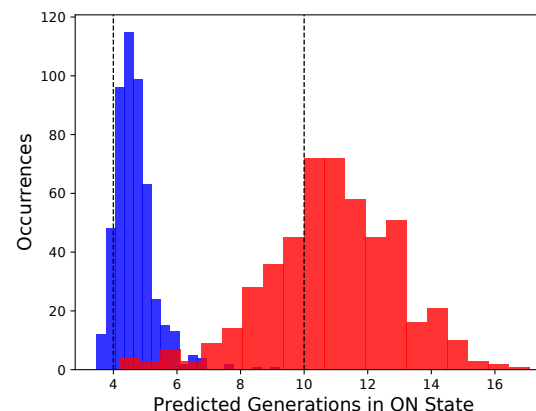

Fig. 8. Benchmarking with gamma-distributed cell cycle time. The fraction of ON cells in the original population (f) was assumed to be 0.10. We consider a fast-switching scenario, in which cells spend an average of 4 generations in the ON state ($1/\gamma = 4$), and a slow-switching scenario, in which cells spend an average of 10 generations in the ON state ($1/\gamma = 10$). We set the proliferation rate (r) to be 1, so that time is normalized to the generation time of the cells. The simulations were run for 12 generations ($t = 12$), and CV^2 was calculated across 40 colonies. Our method's predictions for the fast-switching scenario (shown in blue) cluster around the true $1/\gamma$ value of 4 (shown with a dotted line), while the predictions for the slow-switching scenario (shown in red) cluster around the true $1/\gamma$ value of 10 (shown with a dotted line).

TABLE III

BENCHMARKING RESULTS – LOG-NORMAL DISTRIBUTION OF CELL CYCLE TIME

Scenario	True Generations ON	Predicted Generations ON	Prediction Std. Dev.
Fast-Switching	4	4.77	1.23
Slow-Switching	10	10.97	2.33

was 4.67, and the mean predicted $1/\gamma$ value for the slow-switching scenario was 10.87 (listed in Table IV). Just as with the exponential distribution and log-normal distribution simulations, we again observe a tendency of our formula to slightly overestimate the true number of generations in the ON state. However, yet again the formula estimates of the number of generations ON are close enough to the true underlying values to potentially be useful in a biological context, even though they are not exact.

TABLE IV

BENCHMARKING RESULTS – GAMMA DISTRIBUTION OF CELL CYCLE TIME

Scenario	True Generations ON	Predicted Generations ON	Prediction Std. Dev.
Fast-Switching	4	4.67	0.65
Slow-Switching	10	10.87	2.06

D. Conclusion

Overall, we conclude that this method of switch-rate estimation is efficacious enough to be useful in certain

contexts. Our formula is imperfect and has a tendency to slightly overestimate the number of generations spent in the ON state. However, it is close enough to potentially be useful to biologists as an approximation. Also, our formula has the advantage of being robust to different distributions of cell cycle time. This is a non-trivial result, since the moment analysis techniques used to derive our formula typically assume an exponential distribution of time between events in the simulation, and is an important advantage in real biological contexts, in which the underlying distribution of cell cycle timing may not be known.

VII. DISCUSSION

In this paper, we described a modified fluctuation test that can be done to estimate the rates of transient state-switching in cell populations. We analyzed a stochastic model of the experiment to derive a moment-based formula for the CV^2 of the fraction of cells in the ON state over time. One step in this derivation involved making a simplifying assumption, that the fraction of cells in the ON state F and the total number of cells T are independent, uncorrelated variables. This assumption is not completely correct. In fact, there is a very slight negative correlation between these variables. So, our formula is not exact, but rather an approximation. Despite this imperfection, we wondered if this approximation could still be useful. We benchmarked it on simulated cell populations, using varied assumptions for the distribution of cell cycle time. We found that our method was able to differentiate between fast-switching and slow-switching simulated cell populations to a reasonable degree, suggesting that it may be a useful tool for biologists despite its imperfection.

That the method works well for simulations with gamma-distributed and log-normal-distributed cell cycle times is a non-trivial result, since the moment analysis techniques used to derive the formula assumed exponentially-distributed time between events. Being robust to many potential distributions of cell cycle time is an important advantage, since in a real biological setting the exact underlying distribution of the cell cycle time may not be known.

Future work on this topic will include applying our method to real biological datasets in collaboration with experimental biologists working on topics such as drug resistance in cancer. Also, we will further explore the theory of this method by relaxing some of the assumptions in our problem formulation. For example, in this paper we assumed that cells in both states grow at the same rate, so it will be interesting to explore the theoretical implications of differences in growth rates between the states. Furthermore, we hope to expand this model beyond two states to study systems with three or more states.

APPENDIX

For readers who are not interested in computer programming or familiar the Python language, this section can be skipped. Just note that we used a simulation of cell colonies in which cells switch back and forth between two states, and divide with cell cycle times that can follow log-normal or gamma distributions. However, for readers interested in computer programming and familiar with Python, we will now describe our simulation code. Please note that this code is available upon request from the authors. The Python code is listed below:

```python
import numpy as np

tend = 12
end_states = []

def cell(latest_t, tend, state):
    tau = np.random.gamma(1)

    next_t = latest_t + tau

    if next_t <= tend:

        dt = tau

        if state == 0:
            switch_on_prob = -((-1 + math.e**((-off_rate-on_rate)*dt))*
                on_rate)/(off_rate+on_rate)
            if random.random() <= switch_on_prob:
                state = 1

        elif state == 1:
            switch_off_prob = -((-1 + math.e**((-off_rate-on_rate)*dt))*
                off_rate)/(off_rate+on_rate)
            if random.random() <= switch_off_prob:
                state = 0

    else:
        dt = tend-latest_t

        if state == 0:
            switch_on_prob = -((-1 + math.e**((-off_rate-on_rate)*dt))*
                on_rate)/(off_rate+on_rate)
            if random.random() <= switch_on_prob:
                state = 1

        elif state == 1:
            switch_off_prob = -((-1 + math.e**((-off_rate-on_rate)*dt))*
                off_rate)/(off_rate+on_rate)
```

```python
            if random.random() <= switch_off_prob:
                state = 0


    if next_t < tend:
        recursion_attempt(next_t, tend, state)
        recursion_attempt(next_t, tend, state)
    else:
        end_states.append(state)
```

This simulation algorithm is based on a recursive function **cell()** which represents a single cell. The function takes three arguments: **latest_t** is the timepoint at which the cell is instantiated, **tend** is the global variable that defines the end time of the simulated experiment, and **state** is the state of the cell, where 0 represents the OFF state and 1 represents the ON state.

After a cell is instantiated, we then determine how long it will take for the cell to divide by taking a random draw from some probability distribution, and save this in the local variable **tau**. Importantly, we can choose to draw this number from any distribution, and are not limited to the assumption of exponential cell cycle time. In the example code, we are drawing from a gamma distribution with a mean of 1, using the NumPy library, which we have imported at the top of the code.

Once we have chosen the cell cycle time of this cell, we must check to see if it switches states before dividing. To do this, we have derived formulas for the probability of being ON after some time t if a cell starts in the OFF state, and vice versa. Please note that a cell could theoretically switch back many times before dividing, so in order to calculate these probability formulas it was necessary to use the moment analysis techniques described in Section III of this paper.

Next, we check to see if the time of division of this cell exceeds the end time of the experiment **tend**. If it doesn't, then we recursively call the function to instantiate two new cells, representing the daughter cells of the original cell, with their birth timepoints being the division timepoint of the original cell, and inheriting the state of the original cell. However, if the division timepoint of the original cell is greater than the **tend** endpoint of the experiment, then we do not call the function again, and instead append the state of the cell to the global list **end_states**, which will eventually record the state of every cell in the colony at the end of the experiment.

So, we can simulate a colony of cells by calling the **cell()** function with a start-time of 0, an end time of whatever we set as our **tend**, and a randomly-selected initial state, which will be 1 (ON) with probability f, and 0 (OFF) with probability $1 - f$. After all recursive iterations of the functions are finished running, it will leave us with our **end_states** full of the end states of each cell in the colony at the end of the experiment. In order to do a full simulation of the modified fluctuation test experiment, we can simply run this colony simulation many times, and record the fraction of cells in the ON state each time, which will allow us to measure the CV^2 in these fractions across colonies.

ACKNOWLEDGMENTS

REFERENCES

[1] H. Rubin, "The significance of biological heterogeneity," *Cancer And Metastasis Review*, vol. 9, no. 1, p. 1–20, 1990.

[2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, p. 1183–1186, 2002.

[3] A. Raj and A. V. Oudenaarden, "Nature, nurture, or chance: Stochastic gene expression and its consequences," *Cell*, vol. 135, no. 2, p. 216–226, 2008.

[4] J. M. Raser and E. K. Oshea, "Noise in gene expression: Origins, consequences, and control," *Science*, vol. 309, no. 5743, p. 2010–2013, 2005.

[5] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: from theories to phenotypes," *Nature Reviews Genetics*, vol. 6, no. 6, p. 451–464, 2005.

[6] A. Eldar and M. B. Elowitz, "Functional roles for noise in genetic circuits," *Nature*, vol. 467, no. 7312, p. 167–173, 2010.

[7] G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. V. Oudenaarden, "Systematic identification of signal-activated stochastic gene regulation," *Science*, vol. 339, no. 6119, p. 584–587, 2013.

[8] G. Chalancon, C. N. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, and M. M. Babu, "Interplay between gene expression noise and regulatory network architecture," *Trends in Genetics*, vol. 28, no. 5, p. 221–232, 2012.

[9] A. Magklara and S. Lomvardas, "Stochastic gene expression in mammals: lessons from olfaction," *Trends in Cell Biology*, vol. 23, no. 9, p. 449–456, 2013.

[10] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, "An excitable gene regulatory circuit induces transient cellular differentiation," *Nature*, vol. 440, no. 7083, p. 545–550, 2006.

[11] H. Maamar, A. Raj, and D. Dubnau, "Noise in gene expression determines cell fate in Bacillus subtilis," *Science*, vol. 317, no. 5837, p. 526–529, 2007.

[12] R. D. Dar, N. N. Hosmane, M. R. Arkin, R. F. Siliciano, and L. S. Weinberger, "Screening for noise in gene expression identifies drug synergies," *Science*, vol. 344, no. 6190, p. 1392–1396, 2014.

[13] N. Battich, T. Stoeger, and L. Pelkmans, "Control of transcript variability in single mammalian cells," *Cell*, vol. 163, no. 7, p. 1596–1610, 2015.

[14] I. G. Johnston, B. Gaal, R. P. D. Neves, T. Enver, F. J. Iborra, and N. S. Jones, "Mitochondrial variability as a source of extrinsic cellular noise," *PLoS Computational Biology*, vol. 8, no. 3, 2012.

[15] J. Cha and I. Lee, "Single-cell network biology for resolving cellular heterogeneity in human diseases," *Experimental and Molecular Medicine*, vol. 52, no. 11, p. 1798–1808, 2020.

[16] A. A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, and et al., "Dynamic proteomics of individual cancer cells in response to a drug," *Science*, vol. 322, no. 5907, p. 1511–1516, 2008.

[17] K. E. Gascoigne and S. S. Taylor, "Cancer cells display profound intra- and interline variation following prolonged exposure to antimitotic drugs," *Cancer Cell*, vol. 14, no. 2, p. 111–122, 2008.

[18] C. A. Chang, J. Jen, S. Jiang, A. Sayad, A. S. Mer, K. R. Brown, A. M. Nixon, A. Dhabaria, K. H. Tang, D. Venet, and et al., "Ontogeny and vulnerabilities of drug-tolerant persisters in HER2 breast cancer," *Cancer Discovery*, 2021.

[19] Y. Lu, H. Singh, A. Singh, and R. D. Dar, "A transient heritable memory regulates HIV reactivation from latency," *iScience*, vol. 24, no. 4, p. 102291, 2021.

[20] A. Singh and L. S. Weinberger, "Stochastic gene expression as a molecular switch for viral latency," *Current Opinion in Microbiology*, vol. 12, no. 4, p. 460–466, 2009.

[21] H. R. Clark, C. Mckenney, N. M. Livingston, A. Gershman, S. Sajjan, I. S. Chan, A. J. Ewald, W. Timp, B. Wu, A. Singh, and et al., "Epigenetically regulated digital signaling defines epithelial innate immunity at the tissue level," *Nature Communications*, vol. 12, no. 1, 2021.

[22] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells," *Genetics*, vol. 149, no. 4, p. 1633–1648, 1998.

[23] R. Losick and C. Desplan, "Stochasticity and cell fate," *Science*, vol. 320, no. 5872, p. 65–68, 2008.

[24] G. Balázsi, A. Van Oudenaarden, and J. J. Collins, "Cellular decision making and biological noise: From microbes to mammals," *Cell*, vol. 144, no. 6, p. 910–925, 2011.

[25] T. M. Norman, N. D. Lord, J. Paulsson, and R. Losick, "Memory and modularity in cell-fate decision making," *Nature*, vol. 503, no. 7477, p. 481–486, 2013.

[26] F. St-Pierre and D. Endy, "Determination of cell fate selection during phage lambda infection," *Proceedings of the National Academy of Sciences*, vol. 105, no. 52, p. 20705–20710, 2008.

[27] K. H. Kim and H. M. Sauro, "Adjusting phenotypes by noise control," *PLoS Computational Biology*, vol. 8, no. 1, 2012.

[28] H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, and S. Huang, "Transcriptome-wide noise controls lineage choice in mammalian progenitor cells," *Nature*, vol. 453, no. 7194, p. 544–547, 2008.

[29] E. Abranches, A. M. V. Guedes, M. Moravec, H. Maamar, P. Svoboda, A. Raj, and D. Henrique, "Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency," *Development*, vol. 141, no. 14, p. 2770–2779, 2014.

[30] M.-E. Torres-Padilla and I. Chambers, "Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage," *Development*, vol. 141, no. 11, p. 2173–2181, 2014.

[31] R. L. Thompson, C. M. Preston, and N. M. Sawtell, "De novo synthesis of VP16 coordinates the exit from HSV latency in vivo," *PLoS Pathogens*, vol. 5, no. 3, 2009.

[32] J.-W. Veening, E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen, "Bet-hedging and epigenetic inheritance in bacterial cell development," *Proceedings of the National Academy of Sciences*, vol. 105, no. 11, p. 4393–4398, 2008.

[33] J.-W. Veening, W. K. Smits, and O. P. Kuipers, "Bistability, epigenetics, and bet-hedging in bacteria," *Annual Review of Microbiology*, vol. 62, no. 1, p. 193–210, 2008.

[34] S. J. Altschuler and L. F. Wu, "Cellular heterogeneity: Do differences make a difference?" *Cell*, vol. 141, no. 4, p. 559–563, 2010.

[35] S. M. Shaffer, M. C. Dunagin, S. R. Torborg, E. A. Torre, B. Emert, C. Krepler, M. Beqiri, K. Sproesser, P. A. Brafford, M. Xiao, and et al., "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance," *Nature*, vol. 546, no. 7658, p. 431–435, 2017.

[36] L. Schuh, M. Saint-Antoine, E. M. Sanford, B. L. Emert, A. Singh, C. Marr, A. Raj, and Y. Goyal, "Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer," *Cell Systems*, vol. 10, no. 4, 2020.

[37] S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. Mcdermott, N. Azizian, L. Zou, M. A. Fischbach, and et al., "A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations," *Cell*, vol. 141, no. 1, p. 69–80, 2010.

[38] P. Mu, Z. Zhang, M. Benelli, W. R. Karthaus, E. Hoover, C.-C. Chen, J. Wongvipat, S.-Y. Ku, D. Gao, Z. Cao, and et al., "SOX2 promotes lineage plasticity and antiandrogen resistance in TP53 - and RB1 - deficient prostate cancer," *Science*, vol. 355, no. 6320, p. 84–88, 2017.

[39] M. Ramirez, S. Rajaram, R. J. Steininger, D. Osipchuk, M. A. Roth, L. S. Morinishi, L. Evans, W. Ji, C.-H. Hsu, K. Thurley, and et al., "Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells," *Nature Communications*, vol. 7, no. 1, 2016.

[40] D. Raha, T. R. Wilson, J. Peng, D. Peterson, P. Yue, M. Evangelista, C. Wilson, M. Merchant, and J. Settleman, "The cancer stem cell marker aldehyde dehydrogenase is required to maintain a drug-tolerant tumor cell subpopulation," *Cancer Research*, vol. 74, no. 13, p. 3579–3590, 2014.

[41] S. Paryad-Zanjani, M. M. Saint-Antoine, and A. Singh, "Optimal scheduling of therapy to delay cancer drug resistance," *IFAC-PapersOnLine*, vol. 54, no. 15, p. 239–244, 2021.

[42] T. Cassidy, D. Nichol, M. Robertson-Tessi, M. Craig, and A. R. A. Anderson, "The role of memory in non-genetic inheritance and its impact on cancer treatment resistance," *PLOS Computational Biology*, vol. 17, no. 8, 2021.

[43] S. E. Luria and M. Delbrück, "Mutations of bacteria from virus sensitivity to virus resistance," *Genetics*, vol. 28, no. 6, p. 491–511, 1943.

[44] S. Sarkar, "Haldanes solution of the Luria-Delbrück distribution." *Genetics*, vol. 127, no. 2, p. 257–261, 1991.

[45] B. Houchmandzadeh, "General formulation of Luria-Delbrück distribution of the number of mutants," *Physical Review E*, vol. 92, no. 1, 2015.

[46] C. M. Holmes, M. Ghafari, A. Abbas, V. Saravanan, and I. Nemenman, "Luria–Delbrück, revisited: the classic experiment does not rule out Lamarckian evolution," *Physical Biology*, vol. 14, no. 5, p. 055004, 2017.

[47] Q. Zheng, "Progress of a half century in the study of the Luria–Delbrück distribution," *Mathematical Biosciences*, vol. 162, no. 1-2, p. 1–32, 1999.

[48] S. M. Shaffer, B. L. Emert, R. A. R. Hueros, C. Cote, G. Harmange, D. L. Schaff, A. E. Sizemore, R. Gupte, E. Torre, A. Singh, and et al., "Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors," *Cell*, vol. 182, no. 4, 2020.

[49] Z. Meir, Z. Mukamel, E. Chomsky, A. Lifshitz, and A. Tanay, "Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells," *Nature Genetics*, vol. 52, no. 7, p. 709–718, 2020.

[50] P. Bokes and A. Singh, "A modified fluctuation test for elucidating drug resistance in microbial and cancer cells," *European Journal of Control*, vol. 62, p. 130–135, 2021.

[51] D. J. Wilkinson, *Stochastic modelling for systems biology*. Taylor Francis, 2012.

[52] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *The Journal of Chemical Physics*, vol. 115, no. 4, p. 1716–1733, 2001.

[53] A. Singh and J. P. Hespanha, "Approximate moment dynamics for chemically reacting systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, p. 414–418, 2011.

[54] J. P. Hespanha and A. Singh, "Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems," *International Journal of Robust and Nonlinear Control*, vol. 15, no. 15, p. 669–689, 2005.

[55] A. Singh and J. P. Hespanha, "Stochastic hybrid systems for studying biochemical processes," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1930, p. 4995–5011, 2010.

[56] A. Singh and J. P. Hespanha, "Models for multi-specie chemical reactions using polynomial stochastic hybrid systems," *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005.

[57] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, p. 403–434, 1976.