SCIENTIFIC REPORTS

OPEN

Stochastic modeling of phenotypic switching and chemoresistance in cancer cell populations

Niraj Kumar[1], Gwendolyn M. Cramer[1,2], Seyed Alireza Zamani Dahaj[1,3], Bala Sundaram[1], Jonathan P. Celli[1] & Rahul V. Kulkarni[1]

Phenotypic heterogeneity in cancer cells is widely observed and is often linked to drug resistance. In several cases, such heterogeneity in drug sensitivity of tumors is driven by stochastic and reversible acquisition of a drug tolerant phenotype by individual cells even in an isogenic population. Accumulating evidence further suggests that cell-fate transitions such as the epithelial to mesenchymal transition (EMT) are associated with drug resistance. In this study, we analyze stochastic models of phenotypic switching to provide a framework for analyzing cell-fate transitions such as EMT as a source of phenotypic variability in drug sensitivity. Motivated by our cell-culture based experimental observations connecting phenotypic switching in EMT and drug resistance, we analyze a coarse-grained model of phenotypic switching between two states in the presence of cytotoxic stress from chemotherapy. We derive analytical results for time-dependent probability distributions that provide insights into the rates of phenotypic switching and characterize initial phenotypic heterogeneity of cancer cells. The results obtained can also shed light on fundamental questions relating to adaptation and selection scenarios in tumor response to cytotoxic therapy.

Acquisition of drug resistance constitutes a major challenge in cancer therapy[1–13]. Therapeutic agents (with widely varying biochemical mechanisms) often exhibit a common pattern of providing an initial reduction in tumor burden followed by recurrence of therapeutically resistant disease with more aggressive progression[6,11,12,14]. Tumor recurrence, which is a major obstacle for cancer cure, is primarily associated with the survival and growth of cell phenotypes that are resistant to chemotherapy[15–17]. Therefore, in order to develop new strategies for the effective treatment of human cancers, a quantitative understanding of the underlying processes leading to drug resistance is essential.

Cellular phenotypic heterogeneity is widely observed in many cancers[3,5,8,18–20] as a tumor is often composed of multiple subpopulations[21,22] that show different responses to chemotherapy[9]. In particular, cellular phenotypes that are not sensitive to drugs survive the treatment and can drive drug resistance. As the underlying processes that can lead to the emergence of resistant cells are often stochastic, tumors may locally contain varying numbers of resistant cells. Therefore, quantifying the statistics of drug resistant cells in a tumor is important for effective therapy. Specifically, we are interested in studying population heterogeneity at the start of therapy and aim to address an important issue of therapeutic importance, namely, how to quantify randomness in the numbers of resistant cells prior to drug treatment.

In analyzing population heterogeneity in tumors, a basic question that arises is: How are cell phenotypes that confer survival advantage in the presence of chemotherapeutic drugs generated? A common explanation for the emergence of such phenotypes revolves around Darwinian selection of pre-existing cellular heterogeneity that arises due to random genetic mutations[23–25]. However, the fact that resistant cells switch reversibly to sensitive cells, and that resistant cells often appear on short time intervals (hours to few days), starting from clonal populations, suggests that non-genetic factors play a major role in the generation of phenotypic heterogeneity[5,7,13,26–31]. Such non-genetic phenotypic heterogeneity can arise due to multistability in the underlying gene expression dynamics[32,33] and noise in gene expression[34–36]. It is interesting to note that non-genetic factors are known to generate phenotypic variation and provide fitness benefits in diverse systems, e.g. in the evolution of microbial

[1]Department of Physics, University of Massachusetts Boston, Boston, MA, 02125, USA. [2]Present address: Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Present address: School of Physics, Georgia Institute of Technology, Atlanta, GA, 30332, USA. Correspondence and requests for materials should be addressed to R.V.K. (email: rahul.kulkarni@umb.edu)

colonies under stress[36]. Similar advantages have been reported in the context of *Saccharomyces cerevisiae* wherein phenotypic transitions due to non-genetic factors are known to enhance fitness[37]. Furthermore, theoretical studies have shown how factors such as fluctuations in gene expression can lead to the long-term survival of drug-resistant populations[38] and how stochastic switching between distinct cellular phenotypes under fluctuating environments can lead to optimal population growth[39,40].

These observations suggest that there are two distinct, though not mutually exclusive, mechanisms for the onset of drug resistance in cancer cells: (1) cell phenotypes that are resistant to chemotherapy pre-exist in the tumor prior to treatment and are selected for during the treatment, and (2) cells are induced to develop or acquire resistance due to treatment. Thus, it is important to distinguish between selection of pre-existing populations which are inherently less chemosensitive versus adaptive changes in cancer cells that are activated by exposure to chemotherapy. In the cancer biology literature, the latter option is often referred to as adaptation[41]. This adaptation-selection scenario was first explored in the famous Luria-Delbrück experiments[42] to understand the mechanism of bacterial resistance to bacteriophage infections. The corresponding analysis gave rise to the celebrated fluctuation test which is also used to estimate mutation rates in bacteria. It is important to note that, while in the Luria-Delbrück case phenotypic changes are driven by genetic mutations and thus an irreversible process, in our study, we are considering phenotypic changes that are reversible. Besides reversible phenotypic switching, it is important to consider intrinsic stochasticity in the underlying processes and to characterize cellular heterogeneity as highlighted by previous studies focusing on modeling drug resistance in cancer[43,44].

In consideration of cellular mechanisms likely to be associated with drug resistance, the epithelial-mesenchymal transition (EMT) emerges as a logical candidate. EMT is a conserved cellular program that enables cells of epithelial lineage to transiently acquire traits of mesenchymal cells, including reversible loss of adherens junctions and gain of proteins associated with enhanced motility, adhesion to extracellular substrates and remodeling of the extracellular matrix[45,46]. In cancer cells, this ability to reversibly adopt a more motile phenotype has been linked to tumor invasion and metastasis[47–49] but, more importantly for this study, EMT is also directly linked to chemotherapy resistance and cancer stem cell (CSC) properties[50,51]. The mechanisms through which cancer cells having undergone EMT become resistant to cancer therapeutics have been investigated in a number of studies and recently reviewed by Shibue and Weinberg[52]. EMT populations have been shown to be resistant to classical chemotherapy drugs via a combination of decreased apoptotic signal transduction, increased drug efflux (increased ATP binding cassette transporter expression), and decreased cell proliferation. Although other therapeutics are not explored in the present study it is also worth noting that EMT populations also exhibit resistance to molecular targeted and immunotherapy agents through other mechanisms which have also been studied. In the context of this background, experimental studies described herein focus on established markers of epithelial and mesenchymal phenotype in relation to chemotherapy response, which in this report involves pancreatic ductal adenocarcinoma (PDAC) cells. While recognizing that EMT is more likely a spectrum of intermediate states[53–57], the strong correlation in phenotype and drug response reported here motivates the adoption of two coarse-grained states to be used in the model development. Specifically, a relatively drug-sensitive state with more pronounced epithelial characteristics (E); and a drug-resistant state with increased mesenchymal characteristics (M). In the following sections, we will consider these phenotypes to form the basis of a two-state model of the dynamics of phenotypic switching and associated survival of cancer cells under cytotoxic stress[22,26].

The paper is organized as follows. We begin with a brief description and results from a set of motivating experiments, in which drug resistance is evaluated as a determinant of phenotype in pancreatic cancer cells *in vitro*, and conversely, phenotype as a determinant of drug response in the same cells. The two phenotype switching model inspired by the experimental data is then described. We then develop an analytic approach to quantify population heterogeneity at the start of therapy followed by a protocol for estimating phenotypic switching parameters. We then construct an approximate approach for characterizing the probability distribution of the fraction of resistant cells in a population. We conclude with remarks on possible extensions and future directions.

Results

Evaluation of drug resistance and phenotype in cell culture studies.
We first sought to compare phenotypic traits in naive and drug-resistant pancreatic ductal adenocarcinoma (PDAC) cells. PANC1 cells (a quasimesnchymal PDAC cell line[58]) were exposed to increasing doses of oxaliplatin chemotherapy over successive passages until resistant cells were stable through multiple passages and cryopreservation. As shown in Fig. 1 (upper panels), acquisition of chemoresistance leads to a marked change in phenotype from naive cells displaying characteristically epithelial adherens junctions, to drug-resistant cells with highly branched morphology, no evident E-cadherin, and marked increase in cytoskeletal vimentin (IF quantification, upper right). This pattern of changes in E-cadherin and vimentin expression are classic and well-established markers of EMT. We further examined the reciprocal scenario, in which the same parental cells were directly induced[45] to undergo EMT via administration of exogenous TGF-β (Fig. 1, lower panels). The resultant phenotype is strikingly similar to that of our drug-resistant cells and importantly, exhibits resistance to chemotherapy similar to when resistance was acquired directly through drug exposure. Collectively these results display a symmetry in that acquisition of drug resistance in epithelial cancer cells leads to increase in mesenchymal characteristics, while direct transition from epithelial to mesenchymal phenotype leads to drug resistance.

Coarse-grained Model.
Motivated by the preceding observations and by previous work[22,26], we now consider a simple coarse-grained model (Fig. 2) for phenotypic heterogeneity in tumor cells. We consider that the population of cancer cells consists of two distinct subpopulations; drug-sensitive or drug-tolerant. Based on our experimental results, we denote the drug-sensitive population by E (for epithelial phenotype) and the drug-tolerant population by M (for mesenchymal phenotype). The processes that control the evolution of tumor heterogeneity are as follows: (1) *birth*: each E-type or M-type cell gives rise to birth of new cells of the same type
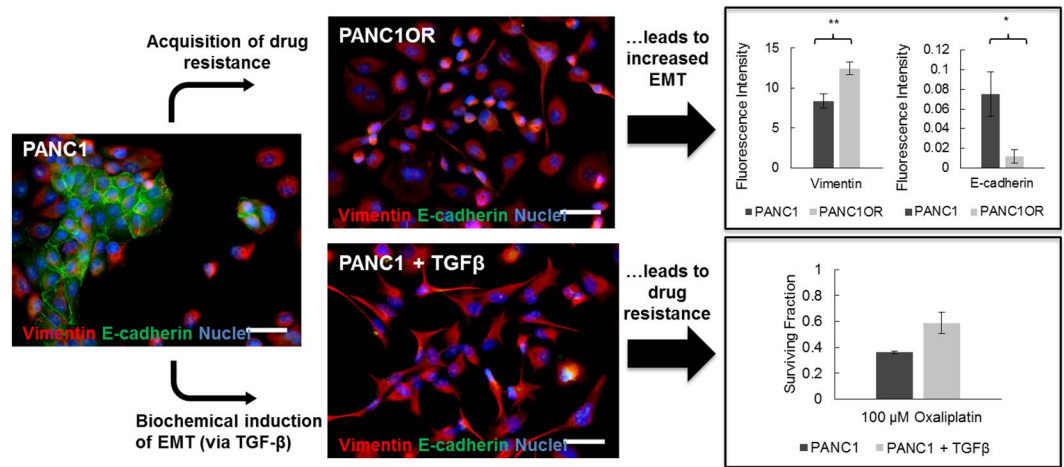
Figure 1. Equivalence in the acquisition of chemotherapy resistance and epithelial-mesenchymal transition in pancreatic cancer cells.
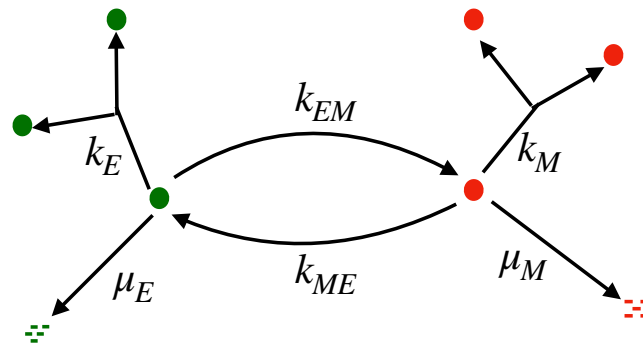
Figure 2. Schematic representation of the two phenotype EMT model of tumor growth: Sensitive and resistant phenotypes are shown as green and red circles respectively. The phenotypic switching rates are represented by K_{EM} and K_{ME}, birth rates by k_E and k_M, and death rates by μ_E and μ_M.

with rates k_E and k_M, respectively; (2) *death*: each E-type (M-type) cell degrades with rates μ_E (μ_M); (3) *phenotypic switching*: an E cell can switch to a M cell with rate k_{EM}, and M cell can switch back to E cell with rate k_{ME}. We assume that intrinsic rates of phenotypic switching and decay of each type of cancer cell are the same throughout the sample.

At any time t, the state of the system is defined by the number of E and M cells. The temporal evolution of the corresponding probability distribution is given by the master equation:

$$
\begin{aligned}
\frac{\partial P(E, M, t)}{\partial t} =\ & k_E(E - 1)P(E - 1, M, t) + k_M(M - 1)P(E, M - 1, t) \\
& + \mu_E(E + 1)P(E + 1, M, t) + \mu_M(M + 1)P(E, M + 1, t) \\
& + k_{ME}(M + 1)P(E - 1, M + 1, t) + k_{EM}(E + 1)P(E + 1, M - 1, t) \\
& - [k_E E + k_M M + \mu_E E + \mu_M M + k_{ME}M + k_{EM}E]P(E, M, t),
\end{aligned}
\tag{1}
$$

where $P(E, M, t)$ denotes the probability that there are E and M numbers of epithelial and messenchymal cells present at time t.

Within the framework of this model (Fig. 2), we now address a key issue: How to characterize the *initial* heterogeneity (i.e. prior to the start of drug exposure) in tumor cells. Consider a snapshot of the cell population in a tumor composed of distinct spatial regions. Each local neighborhood of the tumor can, in general, have widely different fractions of M cells in the local population; this is the heterogeneity we wish to quantify using our model. The proposed protocol for quantifying such heterogeneity involves drawing samples from different spatially distinct regions of the tumor, with each sample corresponding to a fixed number (N_0) of cells, see Fig. 3. Let $p_0 = M/N_0$ denote the local fraction of M-cells in the sample population. Thus, the probability (p_0) that a randomly chosen cell in the sample is a M cell can itself be considered to be a random variable. We denote the corresponding probability density function by $\rho(p_0)$ and characterize it by its mean $\langle p_0 \rangle$ and variance $\sigma_{p_0}^2 = \langle p_0^2 \rangle - \langle p_0 \rangle^2$. Thus, initial tumor heterogeneity is characterized not just by the presence of drug-tolerant M cells in the sample but also by variations in the number of M cells from sample to sample, characterized by the distribution $\rho(p_0)$. Note that
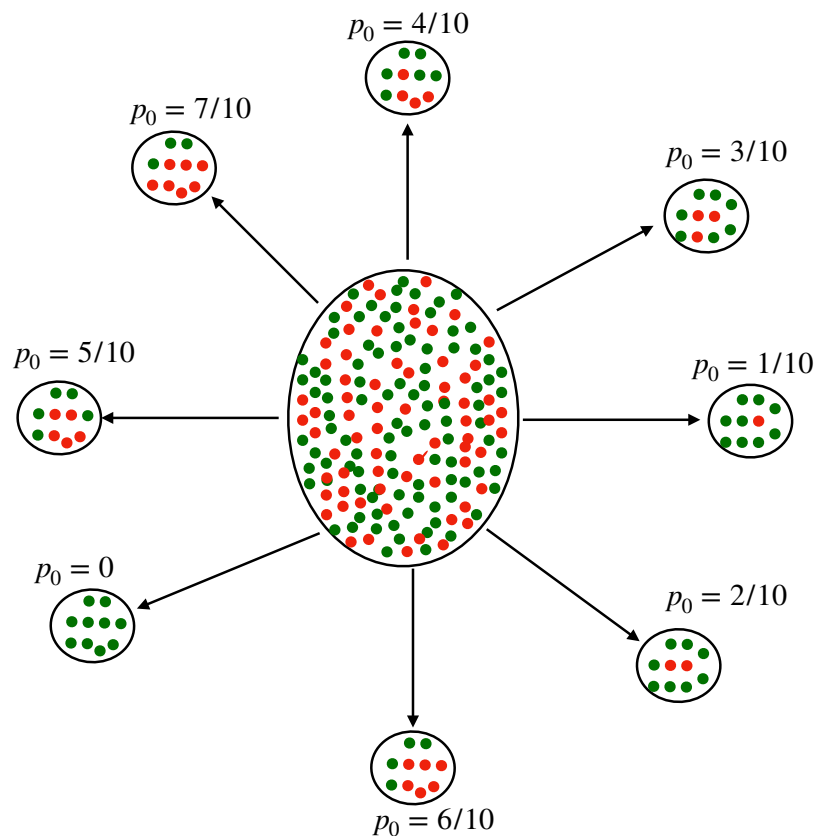
Figure 3. Schematic representation of tumor containing drug-sensitive (green circles) and drug-resistant cells (red circles) is shown at the center. For the sake of conceptual visualization, we have shown different samples taken from the tumor, each characterized by the same number of total cells (here $N_0 = 10$) but different number of M-cells, and thus different values for $p_0 = M/N_0$.

the sample size N_0 should be large enough such that the sampling noise can be taken to be negligible. Thus the heterogeneity in the local fraction of M cells is reflective of the underlying biological heterogeneity rather than sampling noise.

Analytical results for parameter estimation.   We now consider the stochastic process governing evolution of the tumor population upon treatment with drugs. Upon exposure to drugs, it is a reasonable assumption that growth is inhibited, so accordingly we set $k_E = k_M = 0$. As discussed in the preceding section, we consider the evolution of different sample populations, each of which has a fixed initial size N_0 such that the fraction of M cells is drawn from a distribution $\rho(p_0)$. In this limit, the key parameters of the model are: k_{EM}, k_{ME}, μ_E, μ_M, $\langle p_0 \rangle$, $\sigma_{p_0}^2$. In what follows, we derive analytical results that can be used to estimate model parameters by analyzing the distribution of surviving cells upon drug exposure.

We note that, within our model, the evolution of each cell in the population is independent of the state of the remaining cells. Correspondingly, we first focus on the time evolution of a *single* tumor cell, which is initially either E-type or M-type with corresponding probabilities as $1 - p_0$ and p_0. Let us denote by $P_E(P_M)$ the probability that the cell is $E(M)$-type at time t, conditional on the initial probability p_0 for it to be M-type. The corresponding probability generating function for the single cell, conditional on the value of p_0 ($g(z_1, z_2, t|p_0) = \sum_{\eta_E} \sum_{\eta_M} z_1^{\eta_E} z_2^{\eta_M} P(\eta_E, \eta_M, t|p_0)$), can be expressed as

$$g(z_1, z_2, t|p_0) = 1 - (P_E + P_M) + P_E z_1 + P_M z_2. \tag{2}$$

It is straightforward to derive analytic expressions for $P_E(P_M)$ and to thereby obtain an expression for $g(z_1, z_2, t|p_0)$ (Supplementary Material A). Now, let $G(z_1, z_2, t)$ denote the probability generating function corresponding to $P(E, M, t)$, the probability that we have E and M number of sensitive (E-type) and resistant (M-type) cells in the entire population at time t. Since each cell in the population (initial size N_0) evolves independently, the probability generating function for the joint distribution at time t (averaging over the initial choice of p_0) is given by

$$G(z_1, z_2, t) = \int_{p_0=0}^{p_0=1} dp_0 \rho(p_0) g(z_1, z_2, t|p_0)^{N_0}, \tag{3}$$

where $\rho(p_0)$ is the probability distribution function for the initial fraction of M-type cells (p_0).

The expression derived for the generating function, Eq. (3), can be used to derive analytic expressions for all the moments of the marginal distributions corresponding to E-type and M-type cells at time t. For example, expressions for mean number of E and M cells can be obtained using $\langle E \rangle = dG/dz_1|_{z_1=1,z_2=1}$ and $\langle M \rangle = dG/dz_2|_{z_1=1,z_2=1}$, respectively (see Supplementary Material A). This leads to the following expression for mean number of surviving cells at time t, $\langle N \rangle = \langle E+M \rangle$:

$$
\langle N \rangle / N_0 = \left(\frac{\gamma_0 + \alpha_0 - 2(\gamma_0 - \mu_E + \mu_E p_0 - \mu_M p_0)}{2\alpha_0} \right) \exp\left(-\frac{t}{2}(\gamma_0 + \alpha_0) \right)
$$
$$
- \left(\frac{\gamma_0 - \alpha_0 - 2(\gamma_0 - \mu_E + \mu_E p_0 - \mu_M p_0)}{2\alpha_0} \right) \exp\left(-\frac{t}{2}(\gamma_0 - \alpha_0) \right),
$$

(4)

where

$$
\gamma_0 = k_{EM} + k_{ME} + \mu_E + \mu_M, \quad \alpha_0 = \sqrt{\gamma_0^2 - 4(k_{ME}(\mu_E - \mu_M) + (\gamma_0 - \mu_M)\mu_M)}.
$$

(5)

It is clear from the above expression that by fitting the curve corresponding to the mean number of surviving cells as a function of time, the three parameter combinations: α_0, γ_0, and $\mu_E - \mu_E p_0 + \mu_M p_0$ can be determined.

To extract the remaining model parameters based on time-course data, we have to turn to analytic results for the higher moments. For example, we can use expressions for the Fano factor (F) associated with total number of surviving cells, which is given by $F = \sigma_N^2/\langle N \rangle$ with $\sigma_N^2 = \langle N^2 \rangle - \langle N \rangle^2$ denoting the variance in the number of surviving cells. The expression for σ_N^2 can be obtained using

$$
\sigma_N^2 = \sigma_E^2 + \sigma_M^2 + 2C_{EM},
$$

(6)

where $\sigma_E^2 = \langle E^2 \rangle - \langle E \rangle^2$ and $\sigma_M^2 = \langle M^2 \rangle - \langle M \rangle^2$ are variances associated with the marginal distributions for the E and M cells respectively, and $C_{EM} = \langle EM \rangle - \langle E \rangle \langle M \rangle$ is the correlation between numbers of E and M cells. We obtain an explicit expression for the Fano factor given by (see Supplementary Material A):

$$
F = 1 - \frac{\langle N \rangle}{N_0} + \frac{N_0(N_0 - 1)}{\langle N \rangle} \left[\left(\frac{(\mu_E - \mu_M)}{\alpha_0} \right) \left(\exp\left(-\frac{t}{2}(\gamma_0 - \alpha_0) \right) - \exp\left(-\frac{t}{2}(\gamma_0 + \alpha_0) \right) \right) \right]^2 \sigma_{p_0}^2.
$$

(7)

The Fano factor F is a measure of deviations from the Poisson distribution, for which $F = 1$. If $F < 1$ or $F > 1$, the distribution is sub-Poissonian or super-Poissonian, respectively. Before turning our attention to approaches for parameter estimation, let us first examine the expression derived for the Fano factor. We note that in the absence of initial variability in the fraction of M cells (i.e. $\sigma_{p_0}^2 = 0$), the Fano factor of surviving population is simply given by $F = 1 - \langle N \rangle / N_0$. As the cells are not dividing due to the exposure to drugs, the mean number of surviving cells ($\langle N \rangle$) is less than the initial population of cells (N_0) for $t > 0$. Thus, in this case, the Fano factor is *always* less than one and the distribution of cell population follows a sub-Poissonian distribution. However, given variability in the fraction of M-cells in the initial population, the Fano factor can potentially exceed one making the distribution super-Poissonian. This result implies that the observation of a Fano factor in excess of 1 in the distribution of surviving cells is an indicator of variance in the fraction of M-cells in the initial population. Thus the measurements of the moments of surviving cell populations can provide evidence for phenotypic heterogeneity in tumor populations prior to drug treatment.

To gain more quantitative insight into the initial heterogeneity, we need to estimate the parameters characterizing the mean and variance of $\rho(p_0)$. Let us rewrite Eq. (7) in a more compact form by regrouping terms in the expression to yield the following form

$$
\mathcal{F} = \left(\frac{\sigma_{p_0}(\mu_E - \mu_M)}{\alpha_0} \right) \left(\exp\left(\frac{t}{2}(\gamma_0 - \alpha_0) \right) - \exp\left(\frac{t}{2}(\gamma_0 + \alpha_0) \right) \right),
$$

(8)

where the quantity $\mathcal{F}$ can now be expressed entirely in terms of experimentally measurable quantities such as F and $\langle N \rangle$,

$$
\mathcal{F} = \exp\left[\frac{\ln\left(\left(F - 1 + \frac{\langle N \rangle}{N_0} \right) \frac{\langle N \rangle}{N_0(N_0 - 1)} \right)}{2} \right].
$$

(9)

Using the expressions for the mean and Fano factor of the surviving population as functions of time, Eqs (4) and (8), we can estimate four of the parameter combinations, namely, α_0, γ_0, $\mu_E - \mu_E p_0 + \mu_M p_0$, and $(\sigma_{p_0}(\mu_E - \mu_M))/\alpha_0$. Correspondingly, we need additional experiments to determine the entire set of 6 model parameters. As we now show, a set of measurements that accomplish this can be obtained by starting from different initial conditions.

The proposed protocol is motivated by that fact that experimental techniques such as fluorescence-activated cell sorting (FACS) can be used to prepare the samples in specified initial states. With this in mind, we begin from an initial condition where all cells are E-type i.e. $p_0 = 0$, $\sigma_{p_0}^2 = 0$. Using the derived results, we can determine the

parameters μ_E, α_0 and γ_0. We next consider the initial condition to be all M-type cells i.e. $p_0 = 1$, $\sigma_{p_0}^2 = 0$. Analysis of the corresponding time-course measurements of the number of surviving cells can now be used to estimate the parameter μ_M. Having estimated values of α_0, γ_0, μ_E, and μ_M, we can now find the switching rates, k_{EM} and k_{ME}, from Eq. (5). Finally, using the expressions for the mean number of surviving cells and corresponding Fano factors for arbitrary p_0, we can get explicit expressions for the probability p_0 and variance $\sigma_{p_0}^2$ (Supplementary Material A):

$$p_0 = \frac{\langle N \rangle - \langle N \rangle_0}{\langle N \rangle_1 - \langle N \rangle_0},$$

$$\sigma_{p_0}^2 = \frac{N_0 \langle N \rangle}{(\langle N \rangle_1 - \langle N \rangle_0)^2 (N_0 - 1)} \left[F - 1 + \frac{\langle N \rangle}{N_0} \right].$$
(10)

The above results are expressed in terms of experimentally measurable quantities, involving mean values $\langle N \rangle_0$ (for $p_0 = 0$) and $\langle N \rangle_1$ (for $p_0 = 1$) and the Fano-factor of the total surviving population, and thus can be used to estimate the population heterogeneity at the start of drug exposure based on time-course measurements of the surviving population size.

Modeling generation of tumor heterogeneity. The analysis in the preceding section holds regardless of the source of initial heterogeneity in tumor populations. In this section, we explore how the model introduced for tumor cell dynamics can also be used to analyze a potential mechanism for generation of tumor heterogeneity. We note that the proposed model in Fig. 2 can be seen as a generalized version of the celebrated Luria-Delbrück (LD) model with the important addition that, in the present case, the transition between the two phenotypes is reversible (as opposed to the Luria-Delbrück case). However, while the LD model can be solved exactly[59], the exact analytical solution of the reversible model in Fig. 2 is not known, to the best of our knowledge. Nevertheless, as we show below, exact expressions for the mean and variance of the number of E-type and M-type cells can be obtained and used to characterize heterogeneity in tumor cell populations. We can use the master equation, Eq. (1), to derive expressions for the mean number of E and M cells at any time t (Supplementary Material B). Using these expressions, the mean number of surviving cells $\langle N \rangle = \langle E \rangle + \langle M \rangle$ is given by:

$$\langle N \rangle = \left(\frac{E_0(\alpha - \gamma - 2k_E^f) + M_0(\alpha - \gamma - 2k_M^f)}{2\alpha} \right) \exp\left(-\frac{t}{2}(\gamma + \alpha) \right)$$
$$+ \left(\frac{E_0(\alpha + \gamma + 2k_E^f) + M_0(\alpha + \gamma + 2k_M^f)}{2\alpha} \right) \exp\left(-\frac{t}{2}(\gamma - \alpha) \right),$$
(11)

where

$$\gamma = k_{EM} + k_{ME} - k_E^f - k_M^f, \quad \alpha = \sqrt{\gamma^2 + 4(k_{ME}(k_E^f - k_M^f) + (\gamma + k_M^f)k_M^f)},$$
(12)

with $k_E^f = k_E - \mu_E$ and $k_M^f = k_M - \mu_M$ representing the effective birth rates for E-type and M-type cells respectively, while E_0 and M_0 are the initial numbers of E and M cells at $t = 0$.

The results show that the mean number of surviving cells at any time t is characterized by six parameters: initial number of E and M cells (E_0, M_0), two effective birth rates (k_E^f, k_M^f) and two switching rates (k_{EM}, k_{ME}). Given that the initial population can be chosen in a controlled manner, we can use the results for the mean population size to determine some of the model parameters. Specifically, we can set $M_0 = 0$ as the initial condition and fitting the data to obtain the coefficient of exponential terms in Eq. (11) will yield k_E^f. Next, we can set $E_0 = 0$ and Eq. (11) will allow us to extract k_M^f. Once we estimate k_E^f and k_M^f, we can extract the switching rates using the estimated values of γ and α, using Eq. (12). That is, the proposed procedure allows us to estimate the parameter combinations, k_E^f and k_M^f as well as the parameters k_{EM} and k_{ME}.

In order to estimate the remaining model parameters, we need to consider the higher moments. While obtaining analytical expressions for the full probability distribution is still an open problem, higher moments can be calculated in a straightforward manner. For example, using Eq. (1) the evolution equation for $\langle E^2 \rangle = \sum E^2 P(E, M, t)$, $\langle M^2 \rangle = \sum M^2 P(E, M, t)$ and $\langle EM \rangle = \sum EM P(E, M, t)$ is given by

$$\frac{\partial \langle E^2 \rangle}{\partial t} = (k_E + \mu_E + k_{EM})\langle E \rangle + k_{ME}\langle M \rangle + 2(k_E - \mu_E - k_{EM})\langle E^2 \rangle + 2k_{ME}\langle ME \rangle,$$

$$\frac{\partial \langle M^2 \rangle}{\partial t} = (k_M + \mu_M + k_{ME})\langle M \rangle + k_{EM}\langle E \rangle + 2(k_M - \mu_M - k_{ME})\langle M^2 \rangle + 2k_{EM}\langle ME \rangle,$$

$$\frac{\partial \langle ME \rangle}{\partial t} = -k_{ME}\langle M \rangle - k_{EM}\langle E \rangle + (k_E + k_M - \mu_E - \mu_M - k_{ME} - k_{EM})\langle ME \rangle + k_{ME}\langle M^2 \rangle + k_{EM}\langle E^2 \rangle$$
.
(13)

The above set of equations can be solved to get explicit expressions for $\langle E^2 \rangle$, $\langle M^2 \rangle$ and $\langle EM \rangle$ at any time t (Supplementary Material B), which can be used to get the variance (σ_N^2) in the total number of surviving cells by using Eq. (6). The analytic expression for the variance, in combination with the expression for mean number of surviving cells, can be used to extract all model parameters. Furthermore, the extracted parameters can then be compared with the parameters derived based on tumor cell dynamics *after* exposure to drugs. The comparisons

can provide insight into the relative roles of adaptation and selection in driving tumor heterogeneity. The scenario wherein the switching parameters k_{EM} and k_{ME} are effectively unchanged upon exposure to drugs, whereas μ_E and μ_M increase favors selection as the dominant driver of tumor heterogeneity. However, significant changes in the switching rates k_{EM} and k_{ME} are indicative of a role for adaptation as well in the generation of tumor heterogeneity.

Characterizing the distribution of the fraction of resistant cells. The results derived for the moments can also be used to characterize the probability distribution $\rho(p_0)$ for the fraction of M-type cells. Recall that we must have $0 \leq p_0 \leq 1$ and furthermore the first two moments of $\rho(p_0)$ can be obtained using the procedure outlined in the previous section. Thus a natural choice to characterize the distribution $\rho(p_0)$ is to take it to be the Beta distribution with the mean and variance as determined by measurements. The Beta distribution is a natural choice because the domain of a Beta distribution can be viewed as a probability. Another advantage of using a Beta distribution in describing the distribution (or likelihood) of a probability value is that it is a conjugate prior to binomial distribution. This means that, in carrying out Bayesian inference, if the likelihood function is binomial, then a Beta distribution prior will lead to a posterior that is also a Beta distribution (with renormalized parameters). Such a distribution is expressed in terms of two exponents (α and β) as

$$\rho(p_0) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_0^{\alpha-1} (1 - p_0)^{\beta-1}$$

(14)

with mean $\langle p_0 \rangle = \frac{\alpha}{\alpha + \beta}$ and variance $\sigma_{p_0}^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$. The two parameters of the Beta distribution can be estimated using the experimentally determined mean and variance. Explicitly, these are given by

$$\alpha = \frac{\langle p_0 \rangle (\langle p_0 \rangle (1 - \langle p_0 \rangle) - \sigma_{p_0}^2)}{\sigma_{p_0}^2}, \; \beta = \frac{(\langle p_0 \rangle - 1)[\langle p_0 \rangle^2 - \langle p_0 \rangle + \sigma_{p_0}^2]}{\sigma_{p_0}^2}.$$

(15)

To test this approach for characterizing the initial heterogeneity, we compare the Beta distribution with the results obtained from stochastic simulations of the model. Specifically, we carried out stochastic simulations using the Gillespie algorithm[60] for the model in Fig. 2 starting with 200 sensitive E-type cells and no resistant M-type cells. The empirically determined distribution for the fraction of M-type cells ($\rho(p_0)$) is then compared to the Beta distribution with the same mean and variance as the empirical distribution. The results obtained are shown in Fig. 4, which indicate that the Beta distribution is an excellent approximation for the range of parameters explored.

Discussion

To summarize, we have studied a coarse-grained stochastic model to quantify phenotypic heterogeneity in a population of cancer cells. Motivated by the experimental observation that both chemoresistance and TGF-β induced EMT lead to similar outcomes, the model assumes that a cell has two phenotypes corresponding to whether it is drug-sensitive or drug-resistant. Importantly, the model is also consistent with epigenetic mechanisms for generating phenotypic heterogeneity in cancer, given that it allows reversible phenotypic switching between sensitive and resistant cells.

For the model considered, we have derived analytic results, both in the presence and absence of chemotherapeutic agents, which provide insights into the role of phenotypic switching in generating population heterogeneity. One of the issues that we address through these results focuses on quantifying initial heterogeneity in the local fraction of resistant cells. This heterogeneity is characterized by mean fraction of resistant cells $\langle p_0 \rangle$ in a sample and its variance $\sigma_{p_0}^2$. We propose a protocol that can be used to estimate the model parameters based on measurements of mean and variance of the surviving population of tumor cells. Furthermore, our analysis also leads to a condition, in terms of experimentally measurable quantities, whose value serves as an indicator for the presence of initial heterogeneity in the fraction of resistant cells.

While the proposed method allows us to estimate the mean and variance of the fraction of resistant cells prior to therapy, obtaining an exact analytical form for the entire distribution appears to be challenging. However, our simulation results suggest that this distribution is well approximated by the Beta distribution, which can be characterized by using the mean and variance of the surviving population. Besides characterizing initial heterogeneity in the cancer cell population, the estimated model parameters can also be useful in analyzing the complex roles of adaptation and selection in the acquisition of chemoresistance. Furthermore, the results obtained provide exact analytical expressions characterizing the distribution of of tumor cell population under treatment by drugs. The simple model considered in this work can serve as a building block for studying models with more explicit spatial dependence. Going forward we envision further model development in dialog with experiments that longitudinally monitor phenotypic changes in time lapse microscopy studies, either by quantitative analysis of morphometric parameters or implementing fluorescent reporters of EMT which have been recently developed[61]. These results can serve as important inputs to future work focusing on evaluation of the hypothesis that model-informed design of treatment schedule and dose parameters may reduce the emergence of chemoresistance.

Methods

Cell culture and reagents. PANC1 cells were obtained from the American Type Culture Collection (Manassas, VA), and grown in T-75 cell culture flasks according to ATCC guidelines. DMEM medium (HyClone; Logan, UT) was supplemented with 10% FBS (HyClone; Logan, UT), 100 IU/mL penicillin and 1% streptomycin (HyClone; Logan, UT), and 0.5 ug/mL Amphotericin B (Corning; Corning, NY). The drug-resistant subline,
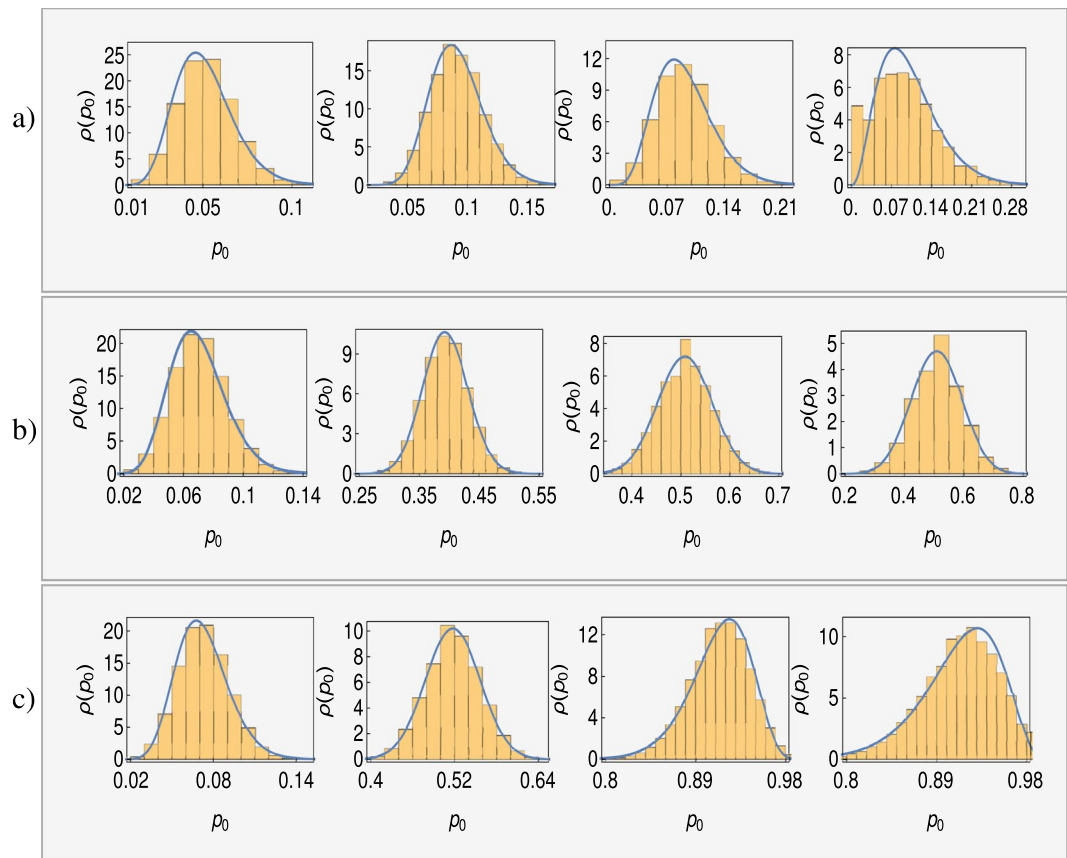
Figure 4. Simulation results for the distributions of fraction of M-cells $\rho(p_0)$ are shown as histograms and the continuous solid lines represent fits by Beta distributions. Top (**a**), middle (**b**) and lower (**c**) panels corresponds to $k_{EM}/k_{ME} = 0.1$, 1 and 10 respectively. In each panel, distributions are shown at various time points, $t = 0.1$, 1, 10, 20 from left to right. Other parameters are: $k_E = 0.2$, $k_M = 0.1$, $\mu_E = 0.3$, $\mu_M = 0.15$.

PANC1-OR was generated as described previously[62]. Briefly, increasing concentrations of oxaliplatin were added to each cell type in regular media over successive passages until a stable proliferative phenotype without chemotherapy was observed and maintained following cryopreservation and confirmed by comparative dose response and measurement of a statistically significant increase in IC50.

Immunofluorescence sample preparation and imaging. Formaldehyde-fixed cells in optical-bottom multiwell plates were incubated overnight at 4 °C with primary antibodies against e-cadherin and vimentin (Cell Signaling EMT Duplex; Danvers, MA). After washing with PBS, cells were incubated for 1 hour with mouse or rabbit Alexa Fluor secondary antibodies (Cell Signaling; Danvers, MA). Cells were mounted with ProLong Gold Antifade reagent containing DAPI (ThermoFisher Scientific Molecular Probes; Waltham, MA) and imaged after 24 hours using a Zeiss LSM 880 confocal microscope with the same detector settings and excitation laser power settings across groups. Images were analyzed using custom Matlab scripts where fluorescent signal for each protein was normalized to the number of cells based on DAPI-stained nuclei.

Therapeutic response assessment. In sample wells receiving chemotherapy treatment, oxaliplatin (Selleck Chemical; Houston, TX) was added to the media at doses ranging from 0.1 to 500 μM for 48 hours. In experiments where EMT was induced via TGF-beta, 10 ng/mL human recombinant TGF-beta (Gibco, Thermo Fisher Scientific) in 1% FBS DMEM was added to designated wells for 48 hours and respective comparison groups were also grown in 1% FBS for the same duration. In all therapeutic studies treatment conditions were prepared in at least triplicate within each batch including internal controls with sham manipulations. Therapeutic response was assessed via the CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega; Madison, WI) at 490 nm absorbance in a BioTek® Epoch Microplate Spectrophotometer.

References

1. Housman, G. *et al.* Drug resistance in cancer: an overview. *Cancers* **6**, 1769–1792 (2014).
2. Gottesman, M. M. Mechanisms of cancer drug resistance. *Annu. review medicine* **53**, 615–627 (2002).
3. Chisholm, R. H., Lorenzi, T. & Clairambault, J. Cell population heterogeneity and evolution towards drug resistance in cancer: Biological and mathematical assessment, theoretical treatment optimisation. *Biochimica et Biophys. Acta (BBA)-General Subj.* **1860**, 2627–2645 (2016).
4. Bozic, I. & Nowak, M. A. Resisting resistance. *Annu. Rev. Cancer Biol.* **1**, 203–221 (2017).

5. Pogrebniak, K. L. & Curtis, C. N. Harnessing tumor evolution to circumvent resistance. *Trends Genet.* (2018).
6. Nikolaou, M., Pavlopoulou, A., Georgakilas, A. G. & Kyrodimos, E. The challenge of drug resistance in cancer treatment: a current overview. *Clin. & Exp. Metastasis* 1–10 (2018).
7. Salgia, R. & Kulkarni, P. The genetic/non-genetic duality of drug 'resistance' in cancer. *Trends cancer* (2018).
8. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. reviews Clin. oncology* **15**, 81 (2018).
9. Zhou, B.-B. S. *et al.* Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nat. reviews Drug discovery* **8**, 806 (2009).
10. Zahreddine, H. & Borden, K. Mechanisms and insights into drug resistance in cancer. *Front. pharmacology* **4**, 28 (2013).
11. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **13**, 714 (2013).
12. Garraway, L. A. & Jänne, P. A. Circumventing cancer drug resistance in the era of personalized medicine. *Cancer discovery* **2**, 214–226 (2012).
13. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nat.* **546**, 431 (2017).
14. Chabner, B. A. & Roberts, T. G. Chemotherapy and the war on cancer. *Nat. Rev. Cancer* **5**, 65–72 (2005).
15. Gatenby, R. & Brown, J. The evolution and ecology of resistance in cancer therapy. *Cold Spring Harb. perspectives medicine* **8**, a033415 (2018).
16. Gallaher, J., Enriquez-Navas, P. M., Luddy, K. A., Gatenby, R. A. & Anderson, A. R. Spatial heterogeneity and evolutionary dynamics modulate time to recurrence in continuous and adaptive cancer therapies. *Cancer research* canres–2649 (2018).
17. Castorina, P., Carcò, D., Guiot, C. & Deisboeck, T. S. Tumor growth instability and its implications for chemotherapy. *Cancer research* **69**, 8507–8515 (2009).
18. Pardal, R., Clarke, M. F. & Morrison, S. J. Applying the principles of stem-cell biology to cancer. *Nat. Rev. Cancer* **3**, 895 (2003).
19. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nat.* **501**, 328 (2013).
20. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323 (2012).
21. Gupta, P. B., Chaffer, C. L. & Weinberg, R. A. Cancer stem cells: mirage or reality? *Nat. medicine* **15**, 1010–1012 (2009).
22. Zhou, J. X., Pisco, A. O., Qian, H. & Huang, S. Nonequilibrium population dynamics of phenotype conversion of cancer cells. *PloS one* **9**, e110714 (2014).
23. Nowell, P. C. The clonal evolution of tumor cell populations. *Sci.* **194**, 23–28 (1976).
24. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci.* **110**, 4009–4014 (2013).
25. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nat.* **501**, 338 (2013).
26. Pisco, A. O. *et al.* Non-darwinian dynamics in therapy-induced cancer drug resistance. *Nat. communications* **4**, 2467 (2013).
27. Pisco, A. O. & Huang, S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'what does not kill me strengthens me'. *Br. journal cancer* **112**, 1725 (2015).
28. Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised epigenetic states and acquired drug resistance in cancer. *Nat. Rev. Cancer* **14**, 747 (2014).
29. Su, Y. *et al.* Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. *Proc. Natl. Acad. Sci.* 201712064 (2017).
30. Inde, Z. & Dixon, S. J. The impact of non-genetic heterogeneity on cancer cell death. *Critical reviews biochemistry molecular biology* **53**, 99–114 (2018).
31. Lee, J. *et al.* Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc. Natl. Acad. Sci.* **111**, E364–E373 (2014).
32. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nat.* **453**, 544–547 (2008).
33. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. review letters* **94**, 128701 (2005).
34. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
35. Zhuravel, D. *et al.* Phenotypic impact of regulatory noise in cellular stress-response pathways. *Syst. synthetic biology* **4**, 105–116 (2010).
36. Fraser, D. & Kaern, M. A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. microbiology* **71**, 1333–1340 (2009).
37. Acar, M., Mettetal, J. T. & Van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. genetics* **40**, 471 (2008).
38. Charlebois, D. A., Abdennur, N. & Kaern, M. Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Phys. review letters* **107**, 218101 (2011).
39. Belete, M. K. & Balázsi, G. Optimality and adaptation of phenotypically switching cells in fluctuating environments. *Phys. Rev. E* **92**, 062716 (2015).
40. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Sci.* **309**, 2075–2078 (2005).
41. Scheel, C., Onder, T., Karnoub, A. & Weinberg, R. A. Adaptation versus selection: the origins of metastatic behavior. *Cancer research* **67**, 11476–11480 (2007).
42. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genet.* **28**, 491 (1943).
43. Kessler, D. A., Austin, R. H. & Levine, H. Resistance to chemotherapy: patient variability and cellular heterogeneity. *Cancer research* **74**, 4663–4670 (2014).
44. Komarova, N. Stochastic modeling of drug resistance in cancer. *J. theoretical biology* **239**, 351–366 (2006).
45. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The J. clinical investigation* **119**, 1420–1428 (2009).
46. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial–mesenchymal transition. *Nat. reviews Mol. cell biology* **15**, 178 (2014).
47. Heerboth, S. *et al.* Emt and tumor metastasis. *Clin. translational medicine* **4**, 6 (2015).
48. Yang, J. & Weinberg, R. A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. cell* **14**, 818–829 (2008).
49. Zhang, Y. & Weinberg, R. A. Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. *Front. medicine* 1–13 (2018).
50. Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *cell* **139**, 871–890 (2009).
51. Singh, A. & Settleman, J. Emt, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29**, 4741 (2010).
52. Shibue, T. & Weinberg, R. A. Emt, cscs, and drug resistance: the mechanistic link and clinical implications. *Nat. reviews Clin. oncology* **14**, 611 (2017).

53. Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. Microrna-based regulation of epithelial–hybrid–mesenchymal fate determination. *Proc. Natl. Acad. Sci.* 201318192 (2013).
54. Jolly, M. K. *et al.* Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. oncology* **5**, 155 (2015).
55. Jolly, M. K. *et al.* Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7**, 27067 (2016).
56. Hong, T. *et al.* An ovol2-zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS computational biology* **11**, e1004569 (2015).
57. Li, C. & Balazsi, G. A landscape view on the interplay between emt and cancer metastasis. *NPJ systems biology applications* **4**, 34 (2018).
58. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. medicine* **17**, 500 (2011).
59. Antal, T. & Krapivsky, P. Exact solution of a two-type branching process: models of tumor progression. *J. Stat. Mech. Theory Exp.* **2011**, P08018 (2011).
60. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal physical chemistry* **81**, 2340–2361 (1977).
61. Toneff, M. *et al.* The z-cad dual fluorescent sensor detects dynamic changes between the epithelial and mesenchymal cellular states. *BMC biology* **14**, 47 (2016).
62. Cramer, G. M., Jones, D. P., El-Hamidi, H. & Celli, J. P. Ecm composition and rheology regulate growth, motility, and response to photodynamic therapy in 3d models of pancreatic ductal adenocarcinoma. *Mol. Cancer Res.* **15**, 15–25 (2017).

Acknowledgements

Author Contributions

N.K., B.S., J.C. and R.V.K. designed research; N.K., S.Z.D. and R.K. carried out theoretical/computational aspects of research, G.C. carried out experimental aspects of research; N.K., B.S., J.C. and R.V.K. wrote the paper and all authors reviewed the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-46926-x.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.