

SCIENTIFIC REPORTS



OPEN

A multi-state model of chemoresistance to characterize phenotypic dynamics in breast cancer

Grant R. Howard^{1,2}, Kaitlyn E. Johnson¹, Areli Rodriguez Ayala ¹, Thomas E. Yankeelov^{1,3,4,5,6} & Amy Brock^{1,3,4}

The development of resistance to chemotherapy is a major cause of treatment failure in breast cancer. While mathematical models describing the dynamics of resistant cancer cell subpopulations have been proposed, experimental validation has been difficult due to the complex nature of resistance that limits the ability of a single phenotypic marker to sufficiently identify the drug resistant subpopulations. We address this problem with a coupled experimental/modeling approach to reveal the composition of drug resistant subpopulations changing in time following drug exposure. We calibrate time-resolved drug sensitivity assays to three mathematical models to interrogate the models' ability to capture drug response dynamics. The Akaike information criterion was employed to evaluate the three models, and it identified a multi-state model incorporating the role of population heterogeneity and cellular plasticity as the optimal model. To validate the model's ability to identify subpopulation composition, we mixed different proportions of wild-type MCF-7 and MCF-7/ADR resistant cells and evaluated the corresponding model output. Our blinded two-state model was able to estimate the proportions of cell types with an R-squared value of 0.857. To the best of our knowledge, this is the first work to combine experimental time-resolved drug sensitivity data with a mathematical model of resistance development.

We aim to investigate how the therapeutic sensitivity of a breast cancer cell population changes over time following exposure to a pulse of chemotherapy. We hypothesize that intratumoral heterogeneity and cellular plasticity play a direct role in the progression of resistance. This hypothesis is based on previous work demonstrating that exposure to chemotherapy induces gene expression changes, metabolic state transitions, and increased drug resistance in subsets of cancer cells^{1–10}. We test this hypothesis of the direct role of the changing composition of subpopulations of differing drug resistance in the observed resistance response using mathematical modeling to estimate the relative frequencies of cells in different drug sensitivity states over time.

Approximately 30 percent of women diagnosed with early-stage breast cancer develop resistance and ultimately progress to metastatic breast cancer¹¹. Doxorubicin is a standard-of-care cytotoxic agent indicated for the treatment of breast cancer; however, the average time to develop resistance to doxorubicin is only 6 to 10 months¹¹. Thus, it is critical to develop a mathematical-experimental approach to describe and predict the conditions and dynamics associated with the onset of resistance *in vitro*, ultimately to improve the efficacy of clinical treatment regimens. We and others have demonstrated evidence of cellular plasticity and adaptability in response to treatment with chemotherapy^{1–6}. For example, it has recently been revealed that melanoma cells exhibit heterogeneity in their metabolic state, with cells utilizing different amounts of oxidative phosphorylation and aerobic glycolysis⁹. In this study of the role of metabolic usage in drug response, functional heterogeneity played a direct

¹Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas, 78712, USA. ²Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas, 78712, USA. ³Institute for Computational Engineering Sciences, The University of Texas at Austin, Austin, Texas, 78712, USA. ⁴Livestrong Cancer Institutes, Dell Medical School, The University of Texas at Austin, Austin, Texas, USA. ⁵Diagnostic Medicine, Dell Medical School, The University of Texas at Austin, Austin, Texas, USA. ⁶Oncology, Dell Medical School, The University of Texas at Austin, Austin, Texas, USA. Grant R. Howard and Kaitlyn E. Johnson contributed equally to this work. Correspondence and requests for materials should be addressed to A.B. (email: amy.brock@utexas.edu)

role in drug resistance as treating with a drug that inhibited aerobic glycolysis led to an increase in sensitivity to treatment⁹. The ability of individual cells to transition from a drug-sensitive to drug-resistant state has been observed in HL60 leukemia cells following chemotherapy exposure. Pisco *et al.* demonstrated that a subpopulation of cells increases expression of the ABC-transporter protein MDR1 in response to a chemotherapeutic pulse, leading to increased drug efflux and increased chemoresistance in those cells¹. These experimental results focus on specific drug resistance phenotypes that emerge in cell subpopulations following treatment. However, because of the vast complexity of resistance mechanisms, it is difficult to identify a single molecular marker of drug resistance that encompasses all drug resistant cells^{12,13}.

Mathematical descriptions of the dynamics of drug resistance may play a critical role in the development of strategies to combat drug resistance^{12,14–18}. Theoretical models have been proposed that incorporate heterogeneous subpopulations in predicting and optimizing treatment response^{19–28} however, these models have not been fully validated with experimental cell population data *in vitro* or *in vivo*. While approaches that incorporate the heterogeneity of resistant and sensitive subpopulations are promising, they remain largely theoretical in nature¹⁶. Strategies such as optimal control theory^{21,22} (treatment aimed at maintaining the optimal composition of cell subpopulations), adaptive therapy²⁹, and alternate metronomic dosing schemes^{21,30} have rarely been implemented in patient care because of lack of experimental validation. Validation of the presence of the predicted subpopulations proposed in these models is essential for progressing from theoretical predictions to implementation.

Although resistance to chemotherapy is a major cause of failure in breast cancer treatment, we do not currently have a mathematical model describing the development of resistance in the context of a dynamic heterogeneous cancer cell population. Conversely, experimental evidence concerning the variety of biological mechanisms of drug resistance is largely derived from static biological observations^{31,32}. Many studies have relied on chemoresistant cell lines established by long-term exposure of cells to escalating doses of chemotherapeutic agent. In some cases, the chemotherapeutic is a required component of the cell culture media, to maintain resistant cell lines with a median lethal dose (LD50) up to 14 times higher than the original cell line³². Resistance observed in these cell lines may not be physiologically relevant to the clinical onset of chemoresistance, in which transient drug resistance may be induced in response to periodic treatment.

In this contribution, we calibrated experimental drug sensitivity data to multiple dynamic population models to test the hypothesis that there is a time-dependent population response to a chemotherapy treatment, and that this response is best described by models that incorporate heterogeneity and cellular plasticity. We combine the functional relevance of experimentally observed drug resistance data with various mathematical models to reveal the dynamic proportions of cells in subpopulations defined by their degree of drug resistance. To validate that our modeling approach was able to identify the composition of a cell population, we applied the model to known mixtures of reference cell populations with different resistance. To the best of our knowledge, this is the first effort to temporally resolve the proportion of drug sensitive and resistant cells in an experimental population in response to transient drug exposure.

Materials and Methods

Data acquisition. *Cell culture.* MCF-7 human breast cancer cells were obtained from ATCC and maintained in MEM (Minimum Essential Media, Thermo Fischer) supplemented with 10% fetal bovine serum (Gibco) and 1% Penicillin-Streptomycin (Gibco). MCF-7/ADR human breast cancer cells were obtained from Robert Clarke³³ and maintained in MEM (Gibco) supplemented with 10% fetal bovine serum (Gibco), 1% Penicillin-Streptomycin (Gibco), and 500 nM doxorubicin (Sigma-Aldrich). A subline of the MCF-7 breast cancer cell line was engineered to constitutively express EGFP (enhanced green fluorescent protein) with a nuclear localization signal (EGFP-NLS). Genomic integration of the EGFP expression cassette was accomplished utilizing the Sleeping Beauty transposon system³⁴. The EGFP-NLS sequence was ordered as a gBlock from IDT and cloned into the optimized sleeping beauty transfer vector pSBbi-Neo. pSBbi-Neo was a gift from Eric Kowarz (Addgene plasmid #60525)³⁴. To mediate genomic integration, this two-plasmid system consisting of the transfer vector containing the EGFP-NLS sequence and the pCMV(CAT)T7-SB100 plasmid containing the Sleeping Beauty transposase was co-transfected into the MCF-7 population utilizing Lipofectamine 2000. mCMV(CAT)T7-SB100 was a gift from Zsuzsanna Izsvak (Addgene plasmid # 34879)³⁵. GFP⁺ cells were collected by fluorescence activated cell sorting. This MCF-7-EGFPNLS1 cell line is maintained in MEM (Gibco) supplemented with 10% fetal bovine serum and 200 µg/mL G418 (Caisson Labs).

Time resolved resistance measurement. MCF-7 cells were plated at 6600 cells/cm³ and cultured for two days in growth media. The media was then exchanged for growth media containing 500 nM doxorubicin. After 24 hours, the doxorubicin media was removed and replaced with growth media to end the drug pulse. Cells were passaged and counted weekly and drug sensitivity assays were performed weekly, as described below. Cell number counts at each week were used to determine the average per capita growth rate per day of the recovering cell population (Fig. 1A). The pulsed dosing of doxorubicin, followed by weekly drug sensitivity assays for 8 weeks, was repeated for a total of five independent MCF-7 cell populations in order to obtain multiple replicates at each time point assayed.

Weekly drug sensitivity assay. Each week, a subset (300,000) of the cells that were exposed to doxorubicin at the start of the experiment were plated into a 12-well plate in growth media. After two days of culture, media was exchanged for growth media containing doxorubicin at a range of concentrations (0, 4, 14, 24, 36, 48, 60, 72, 84, 96, 120, and 144 µM). Twenty-four hours after this dosing, the cells (including supernatant media) were collected via trypsinization, pelleted, and resuspended in 20 µL of media. Live and dead cells were identified with acridine orange and propidium iodide (ViaStain AOPI Staining Solution, Nexcelom Bioscience) and quantified with a

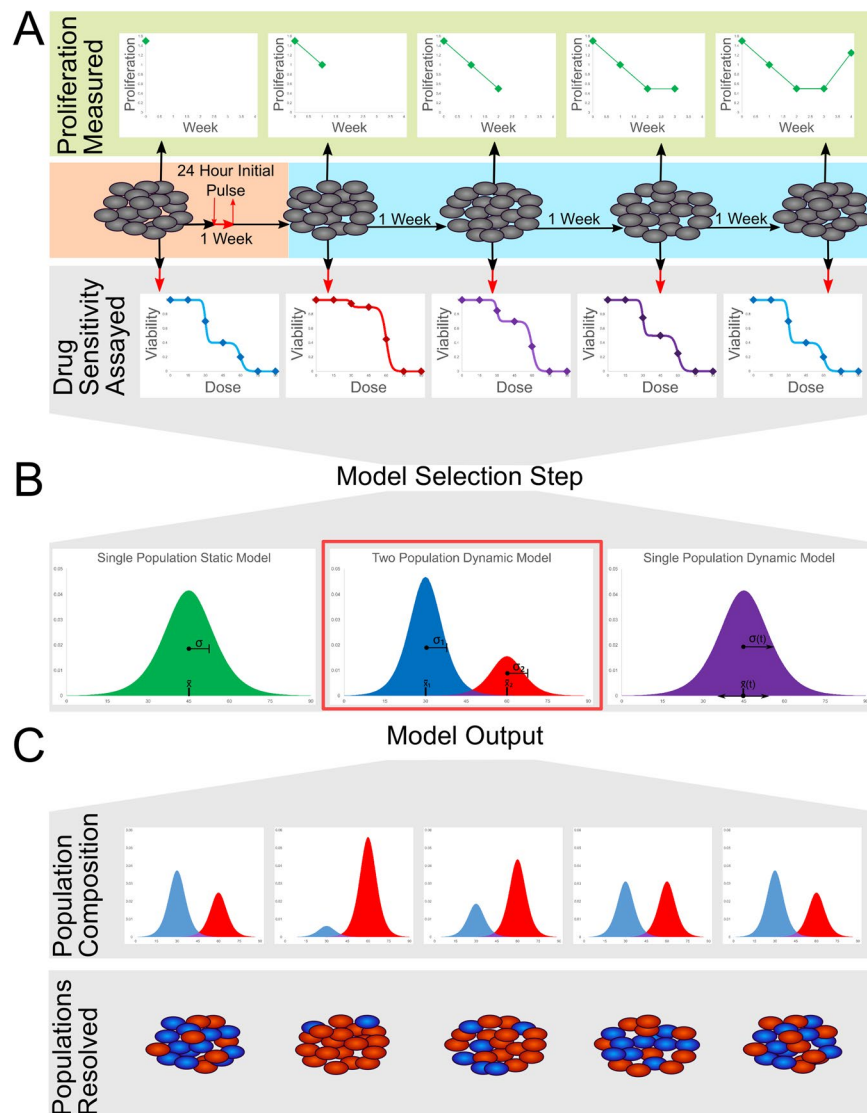


Figure 1. Experimental and modeling workflow: (a) MCF-7 breast cancer cells are treated with an initial pulse of doxorubicin (500 nM) for 24 hours. After treatment, the instantaneous growth rate is measured at each week. However, the subpopulation composition of drug sensitive and resistant cells is not easily identifiable from any single biomarker, as is indicated by the gray cells. To quantify the changes in drug resistance as the population responds to treatment, a subset of cells are extracted each week and a drug sensitivity assay is performed. (b) Using the combined data set containing a drug sensitivity assay at each time point, multiple mathematical models are tested to determine the optimal method for capturing the dynamic response of the cell population. Model selection statistics indicate that a multi-population model of at least two subpopulations is the optimal model. (c) The dynamic two population model estimates the presence of two subpopulations with distinct LD50s and variances corresponding to a sensitive and resistant subpopulation. The model mandates that these states remain constant throughout drug response, with the changes in drug sensitivity of the whole population resulting from changes in the proportions of the areas under the curve of the sensitive versus the resistant population. The model reveals the composition of resistant and sensitive subpopulations at each time point, as is indicated schematically by the ability to identify the proportions of red and blue cells in the population at each week.

Nexcelom Cellometer VBA. The ratios of live to dead cells were used to determine the viability at each concentration of doxorubicin (Fig. 1A).

Cell mixtures for model validation. MCF-7-EGFP-NLS1(wild-type) and MCF-7/ADR (resistant) cells were counted, mixed at desired ratios (1:0, 3:1, 1:1, 1:3, and 0:1), and plated in 12-well plates as described above (*Weekly drug sensitivity assay*). For each defined mixture, a sample of the untreated sample was counted in the Nexcelom Cellometer VBA to determine the measured percent of resistant cells, using the EGFP fluorescence of the MCF-7-EGFP-NLS1 cell line as a marker for the number of MCF-7-EGFP-NLS1 cells which are wild-type with respect

Description	Model equation	Variables and parameters
Single static model	$V_{\text{sin_gstat}}(d) = \frac{V_{\text{max}}}{1 + \exp(m_{\text{ss}}(d - c_{\text{ss}}))}$	V = fraction of cells viable in population V_{max} = maximum viability of cell population (baseline viability at dose = 0 μM) d = dose of doxorubicin (μM) m_{ss} = slope of loss of viability as dose increases, $m_{\text{ss}} = \frac{1}{\sigma}$ c_{ss} = LD50 to describe all data assuming no change in time d = dose of doxorubicin (μM)
Single dynamic model	$V_{\text{indyn}}(d, t) = \frac{V_{\text{max}}}{1 + \exp(m_{\text{sd}}(t)(d - c_{\text{sd}}(t)))}$	$m_{\text{sd}}(t)$ = slope of loss of viability as dose increases $m_{\text{sd}}(t) = \frac{1}{\sigma(t)}$ for each population $c_{\text{sd}}(t)$ = LD50 to describe data at each time point
Two population dynamic model	$V_{\text{twopop}}(d, t) = V_{\text{max}} \left(\frac{f_{\text{sens}}(t)}{1 + \exp(m_{\text{sens}}(d - c_{\text{sens}}))} + \frac{1 - f_{\text{sens}}(t)}{1 + \exp(m_{\text{res}}(d - c_{\text{res}}))} \right)$	$m_{\text{sens}}(t)$ = slope of loss of viability as dose increases, $m_{\text{sens}}(t) = \frac{1}{\sigma(\text{sens})}$ $m_{\text{res}}(t)$ = slope of loss of viability as dose increases, $m_{\text{res}}(t) = \frac{1}{\sigma(\text{res})}$ c_{sens} = LD50 to describe sensitive population c_{res} = LD50 to describe resistant population

Table 1. Mathematical models to describe dynamic drug sensitivity data: We present the equations used for each of the three different structural models that were fit to the time-resolved drug sensitivity assays. The column labeled, “Model equation” provides the functional form of the equation, with t representing a parameter that was fit to the data set at each time point measured. The column labeled, “Variables and parameters” describes the variables used in terms of their physical meaning and their relation to the time-resolved drug sensitivity assays.

to drug sensitivity. The measured percent of cells of each type was calculated by normalizing based on measurement of fluorescence in pure wild type (MCF-7-EGFP-NLS1) and MCF-7/ADR samples.

Data Analysis. *Calibration of experimental data to multiple structural models of dose-response.* The combined data set containing the drug sensitivity assays from each week were fit to three different models (Table 1). The combined data set consists of three variables, the time (in weeks) post-initial doxorubicin exposure, the concentration of doxorubicin applied at that time, and the corresponding cell viability. To perform an estimation of the parameters for all three models, a nonlinear, least-squares approach was implemented in MATLAB (Mathworks). Sigmoidal viability curves are often used to describe chemotherapy dose-response curves³⁶ because they correspond to a population of cells that die at different doses that can be described with a unimodal distribution of cells versus dose as shown in Fig. 1B. The parameters of the sigmoidal dose response function are physically identifiable as the dose at which half the cells die (LD50) is the center parameter (c_{ss}) and the standard deviation of the unimodal distribution is inversely related to the slope of the sigmoidal dose-response curve (m_{ss}). The single static population model is the simplest of the models and ignores the temporal dependency of response. Here the drug response of the combined data set is described by a single homogenous cell population with a single (static) LD50 and slope. The single static population model equation is:

$$V_{\text{sin_gstat}}(d) = \frac{V_{\text{max}}}{1 + \exp(m_{\text{ss}}(d - c_{\text{ss}}))} \quad (1)$$

where V is the proportion of cells viable at the dose, d , of doxorubicin in μM applied, c_{ss} is the LD50 of the population, m_{ss} is the slope at which the cells die due to increases in concentration, and V_{max} is the maximum viability of the cell population (as measured by the assay in absence of drug). The V_{max} parameter is included to normalize for naturally occurring cell death independent of the effects of doxorubicin. The single static model represents the null hypothesis that the initial pulsed dose has no time-dependency in its effect on the cancer cell population.

The single dynamic population model incorporates a temporal dependency when fitting the combined data set. For each time point that drug sensitivity was assessed, the data is fit to an individual dose-response curve to generate LD50 and slope parameters. The model describes the drug response as a single homogenous population whose drug tolerance can change in time. The single dynamic population model equation is:

$$V_{\text{indyn}}(d, t) = \frac{V_{\text{max}}}{1 + \exp(m_{\text{sd}}(t)(d - c_{\text{sd}}(t)))} \quad (2)$$

where the c_{sd} and m_{sd} (LD50 and slope, respectively) parameters pertaining to each week, leading to a 16-parameter model (slope and LD50 at each of the 8 weeks). This model is akin to individually fitting a dose response curve to each week that the drug sensitivity assays were performed.

Finally, the two-population dynamic model describes a cell population with two cell states that differ in drug sensitivity. The dynamics of the drug response are captured by the relative frequency of cells in each state at each time point. The two-state dynamic population model equation is:

$$V_{\text{twopop}}(d, t) = V_{\text{max}} \left(\frac{f_{\text{sens}}(t)}{1 + \exp(m_{\text{sens}}(d - c_{\text{sens}}))} + \frac{1 - f_{\text{sens}}(t)}{1 + \exp(m_{\text{res}}(d - c_{\text{res}}))} \right) \quad (3)$$

	Single static model	Single dynamic model	Two population model
Number of parameters	2	16 2 per time point	12 4 + 1 per time point
AIC value	−2004.4	−2015.6	−2126.4
Mean-squared error	0.057	0.040	0.024

Table 2. Model fit and model selection statistics indicate that the two population model is the optimal model: The two population dynamic model has the lowest AIC value and the lowest mean-squared error, indicating that this model is superior to the single dynamic population model and the single static population model. The number of parameters for the single dynamic model and the two population model vary by the number of time-resolved drug sensitivity assays examined. In this case, for the single dynamic model there are two model parameters of LD50 and slope at each time point examined. In the two population model, there are four model parameters of LD50 and slope for the sensitive and resistant populations respectively, which remain constant, and one additional parameter per time point is used to describe the proportion of cells in each subpopulation at the time the cells were assayed.

where each cell state is modeled as a subpopulation of cells whose LD50 is centered about a mean and slope (c_{sens} , m_{sens} and c_{res} , m_{res} , respectively) which remain constant over time and whose f_{sens} and f_{res} (shown here as $1 - f_{sens}$) parameters can vary to best capture the drug sensitivity assay at each week. To fit the two-state model to data from multiple time points, the parameters of the sensitive and resistant slope and LD50 were forced to be constant at all time points, and the sensitive and resistant fraction parameters were allowed to float at each week, leading to a 12-parameter model (4 fixed parameters, 8 time-dependent fraction parameters for each of the 8 weeks). Equation 3 describes two cell states with distinct LD50 values whose relative frequencies are able to change in time after initial chemotherapy exposure, but whose LD50s remain constant. The overall cell population viability (measured) is modeled as a direct sum of the viability response in each subpopulation.

Statistical analysis and model selection. For all three models, the confidence intervals on the parameter estimates were constructed using the bootstrapping method of replacement³⁷, with 500 bootstrapped simulated data sets. For each model, the mean-squared error and the Akaike Information Criterion³⁸ were calculated for stand-alone model statistics. The Akaike Information Criterion estimator (i.e., the AIC value) is used for direct model comparison. The AIC value evaluates a model based on goodness of fit and penalizes for the complexity of the model using the number of parameters, with a lower AIC value indicating a better model. These evaluation criteria are used to determine the most appropriate model to describe the dynamic dose response data (Fig. 1b, Table 2).

Model validation. To validate the modeling approach, we tested each model's ability to identify known mixtures of wild-type MCF-7 cells with MCF-7/ADR resistant cells. The same two population model and fitting algorithm described above (in the section labeled, *Calibration of experimental data to multiple structural models of dose-response*) were used to fit the combined data set containing a mixture identifier, concentration of doxorubicin, and corresponding cell viability. Therefore, in this validation step, instead of grouping the data by week post drug treatment, we grouped by mixture composition. We allowed each group of mixture replicates to be fit to their own fractional parameter and maintained that the LD50 and centers of the two populations remain constant as described previously. We evaluated the model output of relative frequencies against our measured relative frequencies of wild-type MCF-7 and MCF-7/ADR cells (see *Cell mixtures for model validation*). The bootstrapping method of replacement, again with 500 simulated data sets, was used to construct the confidence intervals for each parameter estimate.

Results

Cancer cell population exhibits time-dependent response to pulse treatment. To determine whether the resistance of the MCF-7 population changes in time after the pulse treatment of doxorubicin, we fit the combined data set to both the static and dynamic single population models, as shown in Fig. 2a,b. In Fig. 2a, the experimentally measured viability is shown alongside the single static population model curve, for both the untreated controls (black) and the pulse-treated cell populations (purple). The LD50 to describe the resistance of these populations is $37.0 \pm 3.5 \mu\text{M}$ and $50.4 \pm 2.4 \mu\text{M}$ for the untreated and treated populations, respectively. Both slope and center parameters for the single static population model can be found in Table S1 of the Supplementary Materials. In Fig. 2b, we show the LD50 estimates, with the 95% confidence intervals, from the single dynamic model over all 8 weeks. The lower AIC value of the single dynamic model (−2015.6) compared to the single static model (−2004.4) indicates the time-resolved dose-response data are better described by the single dynamic model that allows the slope and LD50 value of the population to change at each time point (Table 2). The pattern in estimates of the LD50 values in the single dynamic model (Fig. 2b) corroborate this statistical analysis, showing a significant peak in the drug resistance at intermediate time points, with an LD50 at week 2 at $67.2 \pm 10.0 \mu\text{M}$, followed by a slow return towards baseline, reaching $46.4 \pm 5.1 \mu\text{M}$ at 8 weeks. The LD50 and slope parameter values for the single static population model for all time points can be found in Table S2 of the Supplementary Materials.

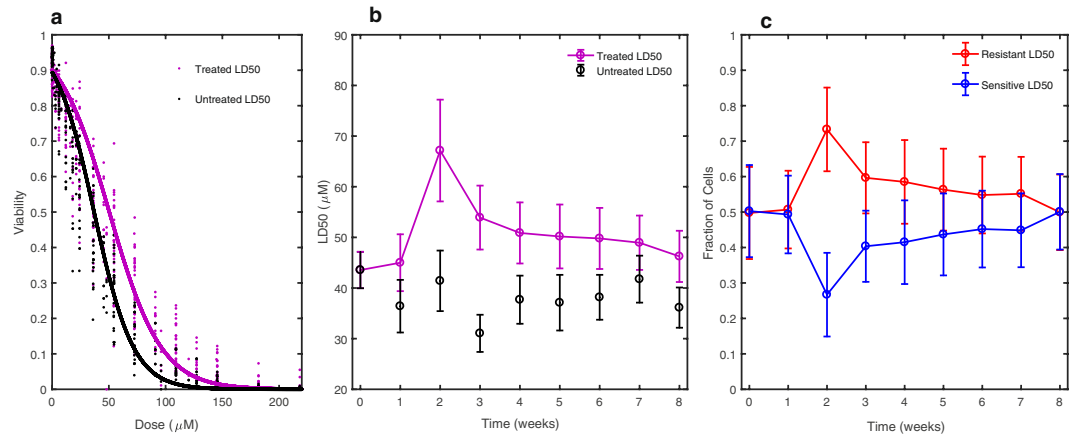


Figure 2. Time-resolved drug sensitivity assays fit to multiple models: (a) The static single population model (solid lines in panel a) demonstrates that the average resistance of the 8 weeks of compounded drug sensitivity data significantly increases from an $LD_{50} = 37.0 \pm 3.5 \mu\text{M}$ for the untreated to an $LD_{50} = 50.4 \pm 2.4 \mu\text{M}$ following exposure to pulse-treatment of doxorubicin. (b) The LD_{50} estimates resulting from analysis with the dynamic single population model indicate that the increase in resistance is time-dependent, peaking at 2 weeks after treatment with an $LD_{50} = 67.2 \pm 10.0 \mu\text{M}$, followed by a slow return towards baseline resistance levels, reaching an $LD_{50} = 46.4 \pm 5.1 \mu\text{M}$ at 8 weeks. (c) The two population dynamic model displays the model-estimated proportions of a population with $LD_{50} = 79.7 \pm 6.5 \mu\text{M}$ (resistant) and a population with $LD_{50} = 22.4 \pm 2.0 \mu\text{M}$ (sensitive). The relative proportions of these two populations change over time, yielding the observed LD_{50} for the overall population. In panels b and c, parameter fits at each time point are connected by a line for visual aid and the error bars represent the 95% confidence intervals on the parameter values using the bootstrapping method of replacement with $n = 500$.

Incorporating heterogeneity via drug sensitivity states improves description of response. To determine whether the dynamic drug response could be explained by a model of two subpopulations, we fit the data to the two population dynamic model to determine the degree of drug sensitivity of the two subpopulations and the resulting sensitive and resistant fraction parameter estimates at each week, with their 95% confidence intervals (Fig. 2c). The two-state dynamic model estimates the presence of a resistant subpopulation with an LD_{50} of $79.7 \mu\text{M} \pm 6.5 \mu\text{M}$ and a sensitive subpopulation with an LD_{50} of $22.4 \pm 2.0 \mu\text{M}$ (Fig. 2c). The parameter values for the sensitive and resistant slope and center, and the fractional parameters, can be found in Table S3 of the Supplementary Materials. We again use model selection to demonstrate that the two population model is an improvement over the single dynamic population model for describing the dynamic drug response of the cell population. Support for selection of this model is indicated by the lower AIC value (-2126.4) of the two population dynamic model compared with the single dynamic model (-2015.6) (Table 2). The results of the model selection analysis for describing dose response curves with multiple conditions (in this case time points) reveals that the differences in dose-response can not be modeled well by allowing each condition to be fit to a single sigmoidal curve allowed to shift, but rather is improved by modeling a sum of sigmoids corresponding to a multi-modal distribution of cells versus lethal dose whose proportions of cells in each state can change for each condition. To illustrate the improvement in fit to the data, Fig. 3a displays the two-state model curve overlaid on the dose-response data at 2 and 8 weeks, demonstrating the ability of the “shoulders” in the curve to fit the initially steep loss of cell viability at lower doxorubicin concentrations, as well as the persisting cell viability at higher doxorubicin concentrations. Figure 3b indicates the model output of relative frequencies of resistant and sensitive subpopulations for the model fit and data shown in Fig. 3a. The improvement in fit of the two-state model is additionally supported by the lower mean-squared error value for the two-state model than the single dynamic model (mean-squared errors of 0.024 and 0.040, respectively), despite it having less parameters than the single dynamic population model (Table 2).

Subpopulation levels of resistant cells transiently increase. The key result for the treatment of doxorubicin on the MCF-7 cell line is that the proportion of cells in the resistant state is consistently higher than baseline from weeks 2 to 5 after the initial drug pulse, followed by return towards the initial resistant and sensitive subpopulation levels (Fig. 2c). The measured per capita growth rate per day (births per cell per day minus deaths per cell per day) (Fig. 4a) was used to estimate the total number of cells at each week. The number of total cells was combined with the fractional parameter estimates at each week (Fig. 2c) to obtain estimates of the number of resistant and sensitive cells at each week in the treated population (Fig. 4b,c). These estimates are purely empirical and make no assumptions about the mechanism at which the cells in each state reached the estimated number of cells at each time, allowing for the possibility of differential growth rates, drug sensitivities, and cell state transitions to determine the corresponding subpopulation levels. This differs from other published work which has assumed sensitive and resistant cells have different growth rates²⁹ and distinct transition rates; here we do not attempt to define the mechanism by which the number of cells in each state was obtained.

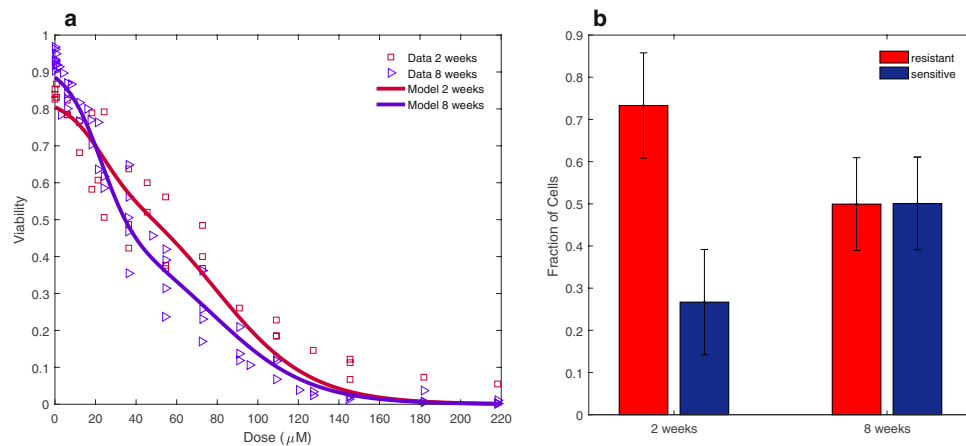


Figure 3. Example fit of drug sensitivity to the two-population model: (a) Best fit of the two-population model to the dose response data at two and eight weeks post-treatment demonstrates the ability of the model to capture the differences in subpopulation levels at the different time points. (b) Two population model output of parameter values of resistant and sensitive fractions at 2 and 8 weeks. The error bars represent the 95% confidence intervals on the parameter values using the bootstrapping method of replacement with $n = 500$.

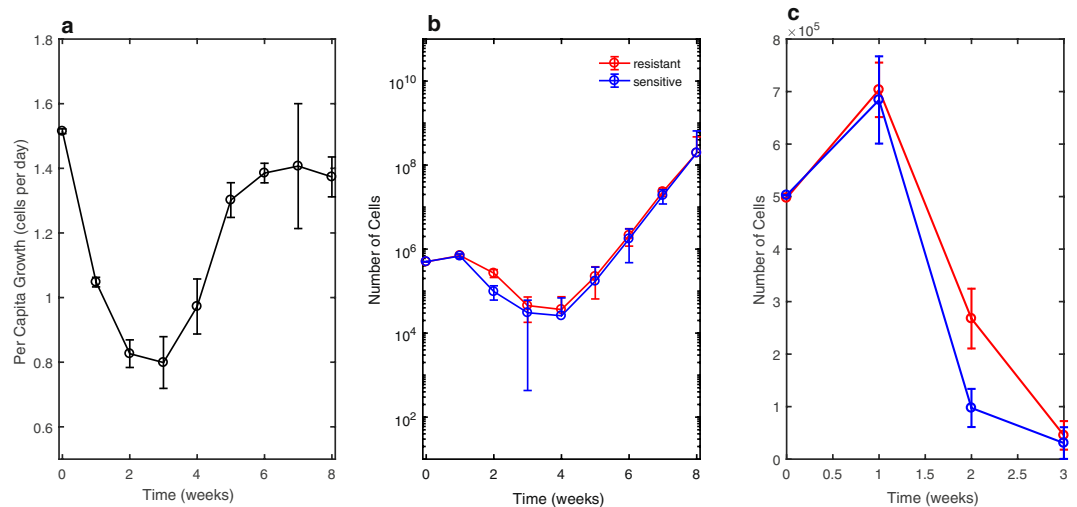


Figure 4. Data driven estimates of phenotypic dynamics: (a) Per capita growth rate (number of births per cell per day minus number of deaths per cell per day) of the entire cell population following pulse treatment of doxorubicin. Error bars represent the 95% confidence intervals from three replicates with per capita growth rate measured at each week. (b) Estimates of the number of resistant and sensitive cells over time are obtained by combining bulk population growth rate to estimate the total number of cells, with the estimates of subpopulation fractions from the two population model output. The estimates of the number of resistance and sensitive cells are purely empirical—cell numbers in each state can be obtained by a combination of differential growth rates, drug sensitivities, or cell state transitions and we do not attempt to identify the means by which the cell numbers are obtained. Error bars are the compounded 95% confidence interval of the per capita growth rate measurement and the parameter estimation of resistant fraction from the two population model output using the bootstrapping method of replacement with an $n = 500$. The number of cells is plotted in logarithmic scale. (c) A closer look in numeric scale of weeks 1–3 displaying the higher number of resistant cells over this time interval.

Model validation confirms ability to reveal subpopulation composition defined by drug sensitivity.

Without a molecular marker of drug resistance, the estimated changes in drug resistant and drug sensitive subpopulations from our two-state model are difficult to validate; that is, we did not know for certain that our model estimated parameters of resistant and sensitive fractions reflect the true subpopulation compositions. We validated the modeling approach by generating experimental reference standards consisting of mixtures of the wild-type MCF-7 cell line with its corresponding doxorubicin resistant cell line, the MCF-7/ADR. We evaluated the ability of the model to estimate the subpopulation composition of each reference mixture. In Fig. 5a, the two population dynamic model fit for each mixture is overlaid on the dose-response data for the corresponding

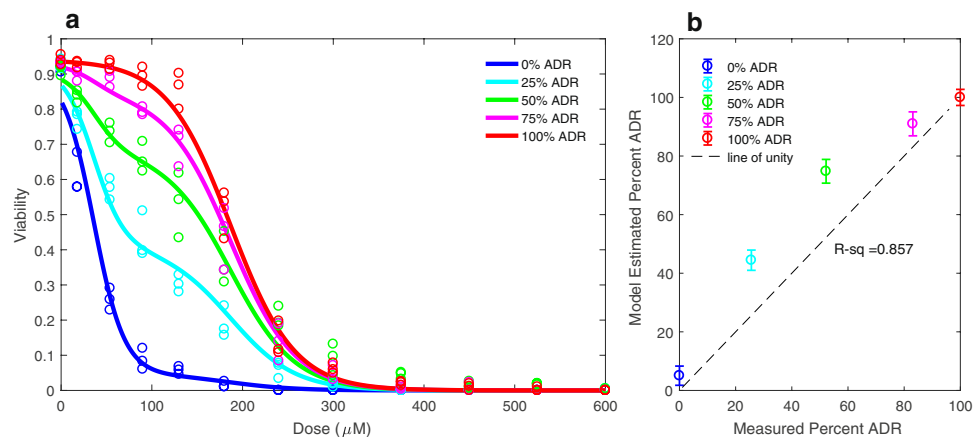


Figure 5. Model validation via identification of mixed populations: **(a)** The two-population model fit for each mixture of MCF-7/ADR resistant cell line and wild-type MCF-7 cell line overlaid on the experimentally measured drug sensitivity for the corresponding mixture. **(b)** Model-estimated percent of MCF-7/ADR cell line in mixture versus the measured MCF-7/ADR cell percentage using fluorescence cell counting. The coefficient of determination (R-squared) value of 0.857 indicates the efficacy of the fractional estimates from the two population model output.

	Measured	Model
LD50 _{ADR}	187.5 +/- 3.8	186.7 +/- 4.6
slope _{ADR}	0.034 +/- 0.0039	0.028 +/- 0.0028
LD50 _{WT}	37.1 +/- 3.3	35.95 +/- 2.8
slope _{WT}	0.055 +/- 0.0077	0.060 +/- 0.0076
frac _{0%ADR}	0.07 +/- 0.01	0.05 +/- 0.03
frac _{25%ADR}	0.30 +/- 0.04	0.44 +/- 0.04
frac _{50%ADR}	0.54 +/- 0.05	0.075 +/- 0.04
frac _{75%ADR}	0.83 +/- 0.05	0.91 +/- 0.04
frac _{100%ADR}	1.0 +/- 0.02	1 +/- 0.0272

Table 3. Model validation demonstrates identifiability of subpopulation compositions: Measured versus two population model estimated parameters indicate the model's ability to reveal the known proportions of the MCF-7/ADR resistant cell line mixed with the wild-type MCF-7 cell line. "Measured" LD50 values and slopes indicate the LD50 and slope of the isolated pure MCF-7/ADR cells and pure wild-type MCF-7 cells fit to the single static population model. Measured resistant fractions are measured precisely using the Nexcelom fluorescence imaging counter of the number of green fluorescent wild-type MCF-7 compared to the total number of brightfield cells counted.

mixture. In Fig. 5b, the measured percent of wild-type and MCF-7/ADR cells are plotted as the line of unity against the parameter estimations of the percent of resistant cells from the model output (Table 3). The measured proportions of wild type and ADR cells versus the model output have a coefficient of determination (R-squared) of 0.857 (Fig. 5b).

Discussion

The key results of this combined experimental and modeling approach demonstrate time-dependency in the response to a clinically relevant pulse-dose chemotherapy treatment, and a modeling system that reveals estimates of the composition of a cancer cell population with functional heterogeneity in its chemoresistance. We have identified that a transient increase in drug resistance is observed following drug exposure and have shown that this can be best described through a model that captures the changing compositions of distinct subpopulations defined by drug sensitivity. These subpopulations were not previously identified due to the complex nature of drug resistance mechanisms. The novelty in this modeling framework is that it is driven from drug sensitivity assays only, without imposing any assumed characteristics or isolated parameter measurements. A data-driven approach allows the model system to be applied to a variety of cell types and drug conditions; for example, in the model validation step, a mutant cell line with high resistance to doxorubicin is assessed. The key finding of this paper is to establish a foundation for describing observable resistance progression throughout time by revealing subpopulations that are not identifiable by molecular markers alone. Future work will need to systematically investigate the molecular and cellular mechanisms of these observed dynamics.

We acknowledge the many limitations of this study. The granularity of the modeling system was limited by technical constraints of the experiment dosing scheme. Bottlenecking of the population in the initial pulse-treatment limited the number of cells available at subsequent weeks for the corresponding drug sensitivity assay. While we were able to measure the response to 12 distinct doxorubicin concentrations each week, this was not sufficient to implement a multi-population model with more than two states, due to a lack of statistical power to significantly resolve differences in subpopulation compositions in time. We acknowledge that a model of only two subpopulations does not capture all likely relevant cell types but believe that this represents a useful simplification of resistance development. In the model validation phase, we were able to estimate the fractions of cells in each state with an R-squared value of 0.857 (Fig. 5b). We were also able to estimate the LD50s of the two populations in mixture using the model within the 95% confidence intervals of the isolated LD50s when fit to the static single population model alone (Table 3). It is possible that discrepancies in the fractional parameter estimates compared to the measured mixtures of the wild-type and resistant cell types may arise from biological interactions between the two that may increase the effective resistance of the cells when in contact with one another. A goal of future studies is to investigate the role of cell-cell interactions in drug resistance.

To ensure that the model parameters are uniquely identifiable and are not overfit to the experimental data sets, we tested the parameter identifiability of the model using simulated data with noise generated from the measured variability as a function of dose in the experimental data (Figure S1). We randomly generated 100 simulated data sets containing 5 different mixtures of resistant and sensitive cells to obtain a distribution for each of the parameter values, with the 95% confidence intervals around the true model parameter values shown in (Fig. S2), demonstrating that the variability in the estimated LD50 (Fig. S2a), slope (Fig. S2b), and fraction parameters (Fig. S2c) was reasonably small. We then addressed the question of the proximity at which the fractional parameters could be distinguished from one another by generating simulated data sets with experimental noise as described, this time containing either mixtures of mostly low levels of resistant cells (Fig. S3a) or of mostly high levels of resistant cells (Fig. S3b). In Fig. S3, the distribution of fractional parameter estimates is displayed as histograms, with each mixture labeled by color. We performed pairwise t-tests between each of the distributions of fractional parameter estimates to confirm that the distributions are statistically significantly different.

Experimental *in vitro* models of resistance to cytotoxic chemotherapy typically utilize resistant cell lines developed via continuous exposure to increased drug concentration^{32,39}. In most cases, drug resistant phenotypes are characterized by an end-point analysis following the stabilization of a resistant cell population. Previous work in the field has indicated that the resistant phenotype of doxorubicin-resistant MCF-7 arises by a multi-factorial process because of observable differences in morphology, gene expression, and DNA content between MCF-7 and MCF-7 resistant cell lines³². The MCF-7/ADR resistant cell line used in our model validation step has an LD50 value of $187.5 \pm 3.8 \mu\text{M}$ (Table 3) and is thus more than 5 times more resistant than the wild-type MCF-7 with an LD50 value of $37.5 \pm 3.8 \mu\text{M}$ (Table 3). The MCF-7/ADR we used typically show a larger proportion of spindle-shaped cells that grow in a more dispersed manner than the wild-type cells³⁹. Other groups have developed resistant MCF-7 cell lines that are 14-fold more resistant to doxorubicin than the original MCF-7 cell line³². They report that the MCF-7 resistant cells are on average larger, contain multiple nuclei, and upregulate genes involving metabolism, drug efflux, and down regulate genes involving DNA repair³². While these experimental observations provide us with key observables to identify as markers of resistance, they do not address the dynamic changes associated with resistance as it develops, nor are they all encompassing. To our knowledge, previous studies involving resistant cell lines have not reported time-resolved measurements of drug resistance following a clinically relevant pulsed dose chemotherapy treatment.

Mathematical modeling of heterogeneity in cancer cell populations has been investigated *via* multiple structural models^{15–25}. In particular, many models have provided predictive capabilities of cell-line specific drug sensitivity⁴⁰, as well as insightful metrics for capturing the growth inhibitory capacity of different drugs^{23,41}. Explorations into different therapy strategies such as optimal control theory have utilized the concept of resistant and sensitive cells within a tumor or cancer cell population^{17–25}. Many models of *in vitro* and *in vivo* cancer progression utilize compartmental ordinary differential equations and partial differential equations. In these models, oftentimes a number of key assumptions are made. For instance, in one model of a heterogeneous tumor, it is assumed that drug resistance is inversely related to proliferation rate²⁹. Other models assume that all sensitive cells are susceptible to the chemotherapy, and do not account for the ability of initially sensitive cells to acquire drug resistance¹⁶. While these models can be extremely useful in capturing drug response and demonstrating the theoretical response to alternate treatment strategies under these sets of conditions, some of their predictions have yet to be fully validated experimentally due to technical limitations in identifying drug resistant subpopulation levels over time.

In this work, we reveal the dynamic changes in subpopulation composition in response to a pulse treatment of drug. In the future, time-resolved subpopulation relative frequencies can be used to develop a model that describes the relative stability of drug sensitivity states and how they change in response to chemotherapy exposure. Ultimately, the results of these experimentally guided models can be used to predict the effect, in terms of composition of resistant cells, of a specific dosing regimen on a cancer cell population over time. The goal of future studies is to use the proposed modeling framework to develop and experimentally validate optimal dosing regimens to be used to combat chemoresistance.

Conclusion

We present this work as one demonstration of the role of heterogeneity in the development of drug resistance. Our analysis indicates that the response to pulsed chemotherapy is time-dependent and that the two-population model identifies subpopulation compositions that change over time. The approach we describe here uncovers chemoresistant subpopulations in breast cancer cell lines and is generalizable to any system in which subpopulations may play a role in a dynamic measurable outcome.

Data availability. The complete datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Pisco, A. O. *et al.* Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat. Commun.* **4**, 1–11 (2013).
- Brock, A., Chang, H. & Huang, S. Non-genetic heterogeneity—a mutation independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* **10**, 336–342 (2009).
- Gupta, P. B. *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
- Brock, A. & Huang, S. Precision Oncology: Between Vaguely Right and Precisely Wrong. *Cancer Res.* (in press) <https://doi.org/10.1158/0008-5472.CAN-17-0448> (2017).
- Brock, A., Krause, S. & Ingber, D. E. Control of cancer formation by intrinsic genetic noise and microenvironmental cues. *Nat. Rev. Cancer.* **15**, 499 (2015).
- Saunders, N. A. *et al.* Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. *EMBO Mol. Med.* **4**, 675–684 (2012).
- Lavi, O., Greene, J. M., Levy, D. & Gottesman, M. M. The role of cell density and intratumoral heterogeneity in multidrug resistance. *Cancer Res.* **73**, 24 (2013).
- Basanta, D., Gatenby, R. A. & Anderson, A. R. A. Exploiting evolution to treat drug resistance: combination therapy and the double bind. *Molec. Pharma.* **9**, 914–921 (2012).
- Hardeman, K. N. *et al.* Dependence on glycolysis sensitizes BRAF-mutated melanomas for increased response to targeted BRAF inhibition. *Sci. Rep.* **7**, 42604 (2017).
- Shajahan-Haq, A. N., Cheema, M. S. & Clarke, R. Application of Metabolomics in Drug Resistant Breast Cancer Research. *Metabolites* **5**, 100–118 (2015).
- Rivera, E. & Gomes, H. Chemotherapy resistance in metastatic breast cancer: the evolving role of ixabepilone. *Breast Canc. Res.* **12**, (2010).
- Wooten, D. J. & Quaranta, V. Mathematical models of cell phenotype regulation and reprogramming: make cancer cells sensitive again! *Biochim. Biophys. Acta.* **1867**, 167–175 (2017).
- Ibrahim-Hashim, A. *et al.* Defining cancer subpopulations by adaptive strategies rather than molecular properties provides novel insights into intratumoral evolution. *Cancer Res.* **77**, 9 (2017).
- Yankeelov, T. E. *et al.* Multi-scale modeling in clinical oncology: opportunities and barriers to success. *Ann. Biomed. Eng.* **44**, 2626–41 (2016).
- Chisholm, R. H., Lorenzi, T. & Clairambault, J. Cell population heterogeneity and evolution towards drug resistance in cancer: biological and mathematical assessment, theoretical treatment optimisation. *Biochim. Biophys. Acta.* **1860**, 2627–2645 (2016).
- Panetta, J. A mathematical model of breast and ovarian cancer treated with paclitaxel. *Math. Biosc.* **146**, 89–113 (1997).
- Enriquez-navas, P. M. *et al.* Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Sci. Transl. Med. Preclinical Models of Breast Cancer* **8**, 1–20, <https://doi.org/10.1126/scitranslmed.aad7842> (2016).
- Mumenthaler, S. M., D'Antonio, G., Preziosi, L. & Macklin, P. The need for integrative computational oncology: an illustrated example through MMP-mediated tissue degradation. *Front. in Oncol.* **3**, 9–12, <https://doi.org/10.3389/fonc.2013.00194> (2013).
- Hansen, E., Woods, R. J. & Read, A. F. How to use a chemotherapeutic agent when resistance to it threatens the patient. *PLoS Biol.* **15**, 2 (2017).
- Badri, H., Pitter, K., Holland, E. C., Michor, F. & Leder, K. Optimization of radiation dosing schedules for proneural glioblastoma. *J. Math. Biol.* **72**, 1301–1336 (2016).
- Foo, J. & Michor, F. Evolution of resistance to targeted anti-cancer therapies during continuous and pulsed administration strategies. *PLoS Comput. Biol.* **5**, 11 (2009).
- Greene, J. M. *et al.* Modeling intrinsic heterogeneity and growth of cancer cells. *J. Theor. Bio.* **367**, 262–277 (2015).
- Harris, L. A. *et al.* An unbiased metric of antiproliferative drug effect *in vivo*. *Nat. Meth.* **13**, 497–500 (2016).
- Mumenthaler, S. M. *et al.* The impact of microenvironmental heterogeneity on the evolution of drug resistance in cancer cells. *Cancer Inform.* **14**, 19–31 (2015).
- Silva, A. S. & Gatenby, R. A. A theoretical quantitative model for evolution of cancer chemotherapy resistance. *Bio. Dir.* **5**, 25 (2010).
- Poleszczuk, J. & Enderling, H. Cancer stem cell plasticity as tumor growth promoter and catalyst of population collapse. *Stem Cells Int.*, 1–12 <https://doi.org/10.1155/2016/3923527> (2016).
- Wilkinson, D. J. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Gen.* **10**, 122–133 (2009).
- McKenna, M. T., Weis, J. A., Brock, A., Quaranta, V. & Yankeelov, T. E. Precision medicine with imprecise therapy: computational modeling for chemotherapy in breast cancer. *Trans. Oncol.* **11**(3), 732–742 (2018).
- Gatenby, R. A., Silva, A. S., Gillies, R. J. & Frieden, B. R. Adaptive therapy. *Canc. Res.* **69**, 4894–4903 (2009).
- Montagna, E., Canello, G., Dellapasqua, S., Munzone, E. & Colleoni, M. Metronomic therapy and breast cancer: A systematic review. *Canc. Treat. Rev.* **40**, 942–950, <https://doi.org/10.1016/j.ctrv.2014.06.002> (2014).
- Gottesman, M. Mechanisms of cancer drug resistance. *Ann. Rev. Med.* **53**, 615–627 (2002).
- AbuHammad, S. & Zihlif, M. Gene expression alterations in doxorubicin resistant MCF7 breast cancer cell line. *Genomics* **101**, 213–220 (2013).
- Vickers P.J., Dickson R.B., Shoemaker R. & Cowan K.H. A multidrug-resistant MCF-7 human breast cancer cell line which exhibits cross-resistance to antiestrogens and hormone-independent tumor growth *in vivo*. *Mol. Endocrinol.* **2**, 886–92 (1988).
- Kowarz, E., Loescher D., & Marschalek R. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol. J.* **10**, 647–53 (2015).
- Mates, L. *et al.* Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* **41**, 753–61 (2009).
- Gardner, S. N. A mechanistic, predictive model of dose-response curves for cell-cycle phase-specific and -nonspecific drugs. *Cancer Res.* **60**, 1417–1425 (2000).
- Efron, B. Better bootstrap confidence intervals. *J. Am. Statist. Assoc.* **82**, 397 (1987)
- S. Konishi & G. Kitagawa, Information Criteria and Statistical Modeling, 1st edn. Springer-Verlag, (2008).
- Ke, W. *et al.* MCF-7/ADR cells (re-designated NCI/ADR-RES) are not derived from MCF-7 breast cancer cells: a loss for breast cancer multi-drug resistant research. *Med. Oncol.* **28**, 135 (2011).

40. McKenna, M. T., *et al.* A predictive mathematical modeling approach for the study of doxorubicin treatment in triple negative breast cancer. *Sci. Rep.*, 7, 1–14 (2017).
41. Hafner, M., Niepel, M., Chung, M., & Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Met.*, 13, 521–527. (2016).

Acknowledgements

The authors are grateful for support from the American Association for Cancer Research, and the Breast Cancer Research Foundation through a Translational Breast Cancer Research (#14-60-26-BROC to A. Brock). We thank the National Institutes of Health for funding through NIBIB T32 EB007507, NCI 1R01CA129961, NCI U01CA174706, and CPRIT RR160005. T.E.Y. is a CPRIT Scholar of Cancer Research.

Author Contributions

Conceptualization: A.B., G.R.H., K.E.J. Investigation: G.R.H., K.E.J., A.R.A. Data curation: G.R.H., K.E.J., A.R.A. Formal analysis: G.R.H., K.E.J., A.B., T.E.Y. Writing- original draft: G.R.H., K.E.J., A.B. Writing- review and editing: G.R.H., K.E.J., A.B., T.E.Y. Supervision: A.B. Project Administration: A.B. Funding acquisition: A.B.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30467-w>.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018