

RESEARCH

Open Access

Assessing the efficacy of molecularly targeted agents on cell line-based platforms by using system identification

Xiangfang Li^{1*}, Lijun Qian², Jianping Hua³, Michael L Bittner³, Edward R Dougherty^{1,3,4}

From IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS) 2011 San Antonio, TX, USA. 4-6 December 2011

Abstract

Background: Molecularly targeted agents (MTAs) are increasingly used for cancer treatment, the goal being to improve the efficacy and selectivity of cancer treatment by developing agents that block the growth of cancer cells by interfering with specific targeted molecules needed for carcinogenesis and tumor growth. This approach differs from traditional cytotoxic anticancer drugs. The lack of specificity of cytotoxic drugs allows a relatively straightforward approach in preclinical and clinical studies, where the optimal dose has usually been defined as the “maximum tolerated dose” (MTD). This toxicity-based dosing approach is founded on the assumption that the therapeutic anticancer effect and toxic effects of the drug increase in parallel as the dose is escalated. On the contrary, most MTAs are expected to be more selective and less toxic than cytotoxic drugs. Consequently, the maximum therapeutic effect may be achieved at a “biologically effective dose” (BED) well below the MTD. Hence, dosing study for MTAs should be different from cytotoxic drugs. Enhanced efforts to molecularly characterize the drug efficacy for MTAs in preclinical models will be valuable for successfully designing dosing regimens for clinical trials.

Results: A novel preclinical model combining experimental methods and theoretical analysis is proposed to investigate the mechanism of action and identify pharmacodynamic characteristics of the drug. Instead of fixed time point analysis of the drug exposure to drug effect, the time course of drug effect for different doses is quantitatively studied on cell line-based platforms using system identification, where tumor cells’ responses to drugs through the use of fluorescent reporters are sampled over a time course. Results show that drug effect is time-varying and higher dosages induce faster and stronger responses as expected. However, the drug efficacy change along different dosages is not linear; on the contrary, there exist certain thresholds. This kind of preclinical study can provide valuable suggestions about dosing regimens for the *in vivo* experimental stage to increase productivity.

Introduction

Drug development is currently an expensive and prolonged process with high attrition rate. The rate of new drug approvals in the U. S. has remained essentially constant since 1950, while the costs of drug development have soared [1]. Industry analysts estimate that it takes \$1 billion to \$4 billion in R&D and 10-15 years

for every new drug brought to market [1-3]. In aggregate, the industrial average rate of attrition measured from first trials in humans to registration seems to be locked at ~85-90% [4,5]. The situation in oncology drug development is even worse [3,6,7]. By contrast, the overall clinical success rate for new anticancer agents (~5%) is much lower than other therapeutic areas (e.g. success rate for cardiovascular diseases is ~20%) [8]. As a result, the American Cancer Society’s 2005 statistical report shows that cancer is now the leading cause of death for Americans under age 85 [9]. One common explanation

* Correspondence: xiangfang@ece.tamu.edu

¹Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

Full list of author information is available at the end of the article

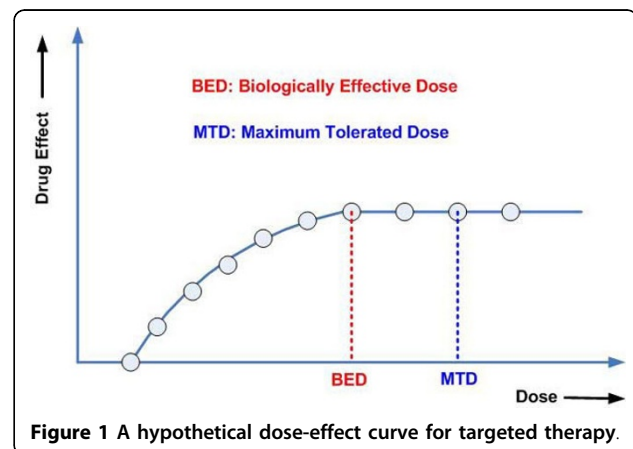
for the recent shrinking of oncology drug pipelines is that discovery is moving into more complex areas of human health [10,11], such as cancer, which is more likely to result from the interaction of several different genes/pathways [12,13]. The conundrum confronting the cancer research community is twofold: first, the pharmaceutical industry is facing difficult times owing to low productivity and spiraling cost [4]; second, on consumers front, patients await better treatments and cancer drugs are an unaffordable luxury for many consumers [14]. To move ahead, scientists realize that they need some fresh thinking in basic, translational and clinical research [15] to improve R&D productivity and reduce attrition rates, and such efforts calls for joint collaboration from different disciplines [5,16-20].

The focus of anticancer drug development in recent years has shifted from cytotoxic drugs to targeted therapy [16,19,21-23]. The goal of this target-based approach is to improve the efficacy and selectivity of cancer treatment by developing agents that block the growth of cancer cells by interfering with specific targeted molecules needed for carcinogenesis and tumor growth [21,22]. This approach is different from traditional cytotoxic anticancer drugs, where most compounds are targeted against molecules required for the maintenance of structural and genetic integrity of rapidly dividing cells. However, despite advances in understanding of the molecular mechanisms of cancer, the promise of targeted cancer therapy remains largely unfulfilled [8,24], with only a few well-known examples, such as imatinib [25] and trastuzumab [26], currently approved [27]. Many promising candidates prove ineffective or toxic owing to a poor understanding of the molecular mechanisms of biological systems they target. Different reasons have been proposed to explain this limited effectiveness of anticancer drug development, including insufficient translational research and lack of adequate preclinical models that recapitulate disease complexity and molecular heterogeneity [8,16,28,29]. Ideally, preclinical models should validate the target, provide information about the mechanism of action of the drug, and identify pharmacodynamic markers of activity. Once the target and mechanism of action have been identified using *in vitro* models, experiments should be undertaken to ensure that inhibition of the target can be achieved at tolerated doses *in vivo* and to identify possible biomarkers of response. Improved preclinical evaluation of compounds has the potential to augment the detection of activity and toxicity, and to reduce the high attrition rate.

While the lack of specificity of the traditional cytotoxic anticancer agents allows a relatively straightforward, well-established approach, developing a paradigm to better analyze the efficacy of molecularly targeted agents (MTAs) is substantially more complex [18,22,30-32]. Many targets are involved in cell signaling pathways,

which are most often not linear, but connected and redundant [33]. Control strategies typically involve a higher multiplicity of inputs and a multiple layer of feedback [34]. As a result, strategies traditionally applied to the development of cytotoxic drugs may not be appropriate for MTAs [32]. Current treatment plan and efficacy evaluations are usually designed empirically for MTAs, without adequate knowledge of the optimal dose and the appropriate schedule [32]. A novel preclinical model combining experimental methods and theoretical analysis is proposed in this study to investigate the mechanism of action and identify pharmacodynamic characteristics of the drug. It is expected that through such preclinical study, valuable suggestions about dosing regimens could be furnished for the *in vivo* experimental stage to increase productivity. We consider several challenges for MTA dosing.

Firstly, the optimal dose has usually been defined as the "maximum tolerated dose" (MTD) for conventional cytotoxic anticancer drugs rather than the dose that produces a quantifiable therapeutic effect. This toxicity-based dosing approach is founded on the assumption that the therapeutic anticancer effect and toxic effects of the drug increase in parallel as the dose is escalated [22]. Such an assumption is sound if the mechanisms of action of the toxic and therapeutic effects are the same, as is often the case with cytotoxic agents. However, most MTAs are expected to be more selective and less toxic than conventional cytotoxic drugs [23]. As a result, the maximum therapeutic effect may be achieved at a dose, defined as the "biologically effective dose" (BED), which could be substantially lower than the traditionally established MTD as discussed by Johnston [31]. A hypothetical dose-effect curve is shown in Figure 1. In addition, the toxic effect may not parallel the therapeutic effect and not be predictive of the therapeutic effect [22]. Hence, the dosing study for MTAs should be based on both drug efficacy and toxicity considerations. Enhanced efforts to



molecularly characterize the drug efficacy for MTAs in preclinical models will be valuable for successfully estimating the BED for clinical trials.

Secondly, the pharmacodynamics (PD) of drugs have been extensively investigated *in vitro* and *in vivo*; however, most analyses have reported the relationship of drug exposure to drug effect at a fixed time point. When drug effect is examined at a fixed time point, the drug concentration-effect relationship can be characterized through well established models, such as the Hill equation [35], also called the sigmoidal E_{max} model [36]. However, characterization of the entire time course of drug effect may provide additional information [37]. For example, it may help to design the optimal schedule for drug administration.

Thirdly, traditional design of the dosing regimen to achieve some desired target goal such as relatively constant serum concentration may not be optimal because MTA targets mostly sit in interacting complex dynamical regulatory networks and such complex target contexts pose significant challenges for assessing mechanisms of action for MTAs [30]. For example, Shah and co-workers [38] demonstrate that the BCR-ABL inhibitor dasatinib, which has greater potency and a short half-life, can achieve deep clinical remission in CML patients by achieving transient potent BCR-ABL inhibition, while traditional approved tyrosine kinase inhibitors usually have prolonged half lives resulting in continuous target inhibition. A similar study of whether short pulses of higher dose or persistent dosing with lower doses have the most favorable outcomes has been carried out by Amin and co-workers in the setup of inactivation of HER2-HER3 signaling [39].

In sum, it is difficult and expensive to optimize dosing regimens using strictly empirical methods for MTAs. A novel preclinical model combining experimental methods and theoretical analysis is proposed in this study to investigate the mechanisms of action and identify pharmacodynamic characteristic of MTAs. As a first step, the time courses of drug effect for different doses are quantitatively studied on cell line-based platforms using system identification, where a tumor cell's response to investigational drugs through the use of fluorescent reporters is sampled frequently over a time course. A dynamic model is proposed to study the time course of drug efficacy for MTAs and then the experimental data are analyzed by our proposed model using a Kalman filter. Through such preclinical study, valuable suggestions about dosing regimens may be furnished for the *in vivo* experimental stage to increase productivity.

Methods

The proposed approach is an integration of experiment and theory to investigate regulatory process dynamics by combining multiple complementary disciplines,

including: (i) using fluorescent reporters in molecular technology to study cells' transcriptional activities under drug perturbation; (ii) these being captured by an automatic epifluorescent microscope over a time course; and (iii) such data being processed by large-scale image processing for dynamic analysis. A truly multi-dimensional dynamics of tumor cell response to drugs can be characterized through systematic perturbations to test different combinations of cell types, reporters, and drugs/dosages, augmented by iterative systematic theoretical analysis. This methodology differs from high-throughput technique like RNA expression profiling with microarrays, which provide a snapshot of an aspect of the system at one time point.

Experimental methodology

Understanding cell response to a drug requires experimental designs that ask very specific questions about what is happening in a cell in the absence of a drug and how the cell activities change when the drug is present. The objective of the experimental protocol is to efficiently capture cell process dynamics in response to drugs and thereby obtain a deeper understanding of the genetic regulatory mechanisms, the point being to make preclinical research more predictive. Fluorescent reporters have long been used in molecular technology to study cells' transcriptional activities or the cellular localization of components, either in a population of cells or a single cell [40-42]. In this study, we track the transcriptional activities of particular genes. A fluorescent reporter to serve this purpose can be constructed by fusing the promoter region of a gene of interest with the coding sequence of a fluorescent protein, most commonly a green fluorescent protein (GFP). By delivering a single cassette bearing the promoter/GFP reporter into the genome of each cell in a population of cells, any change in the expression levels of the native coding sequence driven by that promoter will be reflected in the transcriptional activity of the cassette. This allows the estimation of the total fluorescence of the reporter in the cell, captured by imaging with an epifluorescent microscope, which is then used as a relative measure of the transcriptional activity of the native gene. Because this procedure is non-invasive to the cell, it allows tracking of the same cell population for an extended period of time by imaging the same site repeatedly. The recent introduction of automated digital microscopes allows researchers to use multi-well microtiter plates and sequentially capture the transcriptional activities in all wells. In our experimental protocol, a single assay is carried out by epifluorescent imaging of a site at the bottom of each well in a 384 well plate, producing an image of the cells in that region (~200-400 cells) bearing fluorescent reporters. The imaging speed of automated systems easily accommodates sampling an entire 384 well plate at hourly intervals. If

needed, the experiment can be extended to multiple plates to cover a wider range of cell types and reporters.

In this experimental set-up, using different wells to test different combinations of cell type, GFP reporter and experimental condition allows this approach to provide a multi-dimensional examination of the cells' responses to a variety of stimuli. Not only can it follow multiple genes simultaneously, but it can also compare cellular activities under various conditions. Furthermore, it captures the dynamics of transcriptional regulation. This produces data on ~200-400 individual cells per well that can be analyzed both individually, as a distribution, or in aggregate, as an average. Fluorescent intensity data can be extracted from these images using specialized image analysis tools developed for this application [43]. This image processing procedures include finding cells, identifying individual cells, and quantifying the fluorescence associated with each cell. The objective is to extract gene expression levels from the fluorescent image and track them over the time course. We approach this goal through morphology-based image processing methods.

Image processing

Typical fluorescent images are shown in Figure 2 (left panels), where nuclei are detected in the blue channel and promoter reporters to study cells' transcriptional activities are detected in the green channel. With a 384-well plate there will be at least 384 videos for evaluation and the number can be much higher if the experiment requires multiple plates to cover all experimental conditions. Visual evaluation is unreliable when one needs to quantitatively compare different conditions and the high-throughput nature of the green fluorescent protein reporter approach calls for a more automatic and quantitative solution to efficiently extract gene-expression levels from the fluorescent images and track them over the time course.

To facilitate automatic processing of the experiment results, the transcriptional levels of the fluorescent images need be properly extracted, quantized, and saved and the image processing algorithm should be fast with good balance between performance and robustness [43]. An algorithm based on morphological image processing [44], in particular, the watershed transformation [45] is currently adopted in our study. Overall, the image processing breaks down into three major components: (i) nuclei channel segmentation, (ii) reporter channel segmentation, and (iii) measurement of cell-by-cell promoter activity levels. Figure 3 shows the segmentation results of a typical fluorescent image pair, where only a portion of the full image is shown in order to show the segmentation details. Once the individual cells are identified, the transcriptional activity represented by the reporter is extracted for every cell by summing up the background subtracted pixel intensity of the whole cell area and taking a \log_2 transform before being exported.

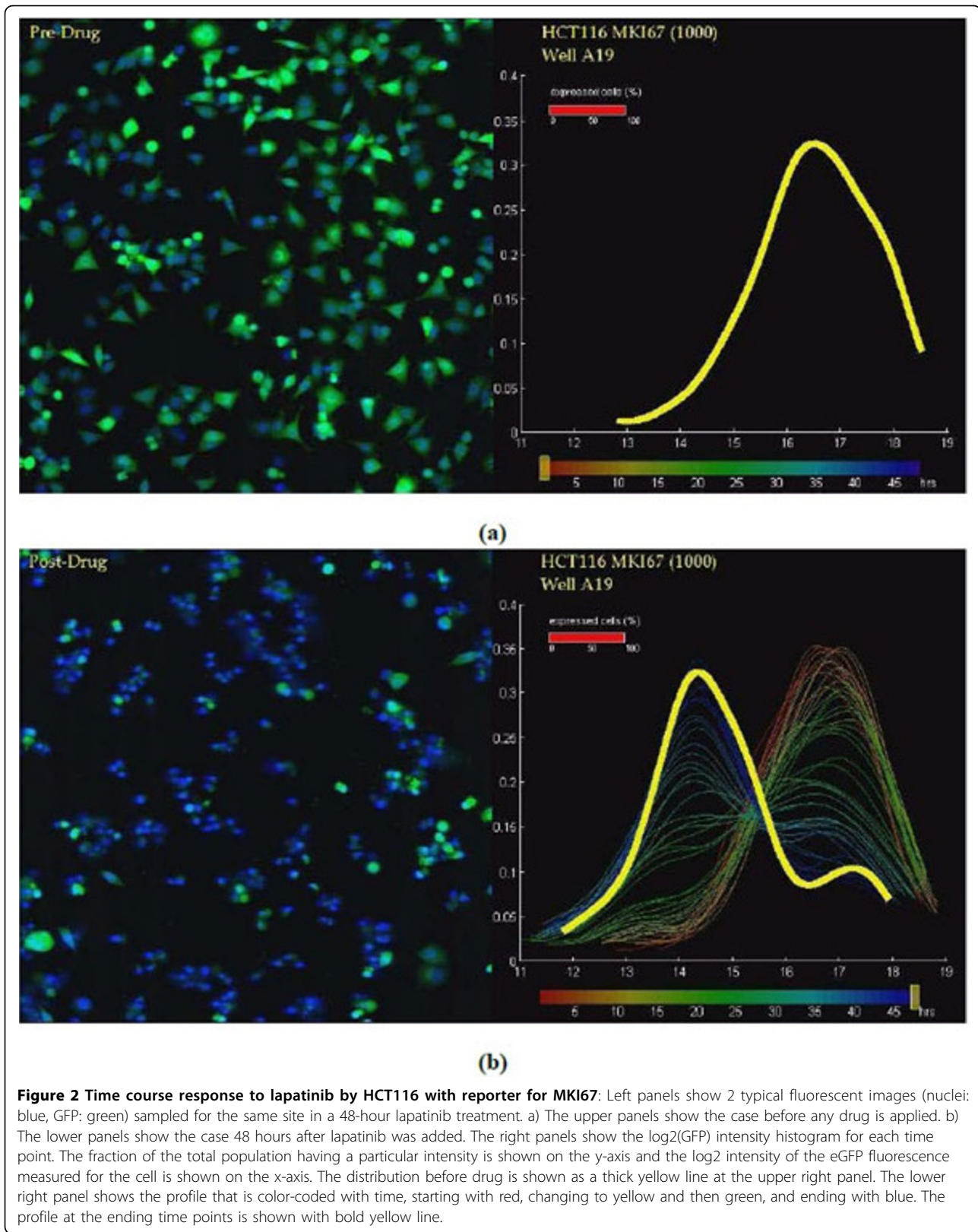
Experimental set-up for the dosing study

The dosing study is carried out on the colon cancer cell-line HCT116 with a reporter for the MKI67 gene, a nuclear antigen tightly correlated with proliferation [46,47], with responses to lapatinib treatment with 6 dosages (1 to 32 μM). First we infect the HCT116 cell lines with the desired packaged reporter (packaged as lentiviral particles). Then plate cells/reporter pair in a media containing a live-cell nuclear stain. The cells are allowed to attach to the plate and grow overnight. Drugs are added to the appropriate wells (we have 6 wells [biological duplicates] for each dosage). In order to remove environmental effects, such as growth factor depletion, there are 6 control wells for each dosage (no drug added, total 36 wells). We image the plate once an hour for 48 hours to characterize the response of each cell/reporter pair to the drug over time. Note that the fluorescence intensity of cells without a GFP reporter expressed is not zero, since cells have numerous small molecules which fluoresce in the same wavelengths as GFP when excited with 488 nm light. This defines the minimum fluorescence, which is approximately 2^{14} . One of the time courses from experiment (dosage = 8 μM) is shown in Figure 2. The left panels of Figure 2 show two fluorescent images sampled for the same site in a 48 hour lapatinib treatment for 8 μM dosage. The right panels of Figure 2 show the $\log_2(\text{GFP})$ intensity histogram for each time point.

Since MKI67 is turned on during proliferation and off when the cells are not cycling, it is expected to show a binary, switch-like histogram of cell intensities, rather than a graded transition. This behavior is observed in Figure 2. We have the readout of the GFP intensity level for each individual cell/dosage pair with 48 time points. These can be compared with a threshold value to determine whether that cell is shifted or not [37,43]. Such a reporter assay allows one to determine the dynamics of drug responses for different dosages. Consequently, we propose a time-varying model for the cell shifting process where the drug effect coefficient is assumed changing with time. This is in contrast to many existing approaches where the drug effect coefficient is treated as a constant and the experiment just provides one reading rather than time-series characterization.

Mathematical model formulation

The experimental results provide information on the percentage of cells shifted as a consequence of the drug activity. The measurements facilitate asking important questions in drug development. For instance, does dosing alter the extent of response, the timing of response, or both? In addition to qualitative questions, we are interested in modeling the drug effect *quantitatively*, which requires a novel mathematical model that is biologically



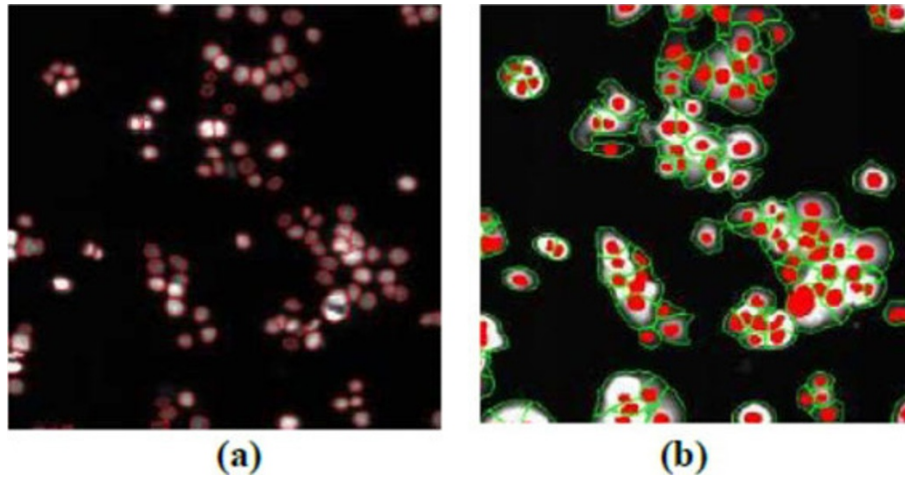


Figure 3 Segmentation Results: a) left panel: nuclear channel, where red lines are the identified nuclei boundaries; b) right panel: reporter channel, where green lines are identified cell boundaries, while the red objects are the nuclei used as markers.

sound and fits the experimental setup. Our experiments and the proposed modeling has two important features: (i) Our experiment is based on the readout of the intensity level of each individual cell, which is compared with a threshold value to determine whether that cell is shifted or not. Although we count the number of shifted cells at each sampling time point, the proposed model is *not* a population model merely giving the average readout of all the cells. (ii) Our experiment collects time-series data under drug perturbation for 48 hours, with one sample per hour. A time-varying model is proposed for the cell shifting process, where the drug effect coefficient is assumed changing with time.

Because there are different numbers of cells in different wells (the range is about ~200-400 cells per well), we perform normalization to calculate the percentage of cells shifted. Since there are many factors including drug effect that contribute to the cell shifting, calibration is performed by comparing to the control group to exclude other contributing factors. The notations used in this work are listed below

- N : total number of cells
- $N_1(t)$: number of shifted cells at time t after applying drug
- $\rho_1(t) = N_1(t)/N$: percentage of cells shifted at time t after applying drug
- N_c : total number of cells in the control group (no drug applied)
- $N_{1c}(t)$: number of shifted cells at time t in the control group
- $\rho_{1c}(t) = N_{1c}(t)/N_c$: percentage of cells shifted at time t in the control group
- $\rho(t) = \rho_1(t) - \rho_{1c}(t)$: calibrated percentage of cells shifted at time t after applying drug

- $\rho_{av}(t) = E[\rho(t)]$: mean of the calibrated percentage of cells shifted at time t after applying drug
- $X_i(t)$: state of cell i at time t after applying drug (either shift-ready or not)

We justify $N_1(t)$ being modeled as a Gaussian process when the number of cells per well is sufficiently large. Then a model is proposed for the cell shifting process, where the calibrated percentage of shifted cells follows a Gaussian process.

$N_1(t)$ is a Gaussian process when the number of cells per well is large enough

In general, N is a random variable since N may be different from well to well in the experiment; however, N can be treated as a known constant for each specific well, as can N_c . At any given time point t_j in the experiment, $X_i(t_j)$ can be considered as either shift-ready or not. Thus, the experiment of drug effect on each cell can be treated as a Bernoulli trial and $X_i(t_j)$ can be modeled as a Bernoulli random variable, i.e., the Probability Mass Function (PMF) of $X_i(t_j)$ is given by

$$P_{X_i}(x_i) = \begin{cases} p & x_i = 1 \\ 1 - p & x_i = 0 \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $0 \leq p \leq 1$ and t_j is dropped for simplicity of presentation. Under this definition, $N_1 = \sum_{i=1}^N X_i$. Assuming that all cell states are independent, N_1 has the binomial PMF given by

$$P_{N_1}(n_1) = \binom{N}{n_1} p^{n_1} (1 - p)^{N - n_1} \quad (2)$$

When the number of cells per well is large, say $N > 100$, the PMF of N_1 at any given time instant can be

accurately approximated by the Gaussian distribution due to the central limit theorem. Next we show that $N_1(t)$ is a Gaussian process.

Proposition 1. *The random process $N_1(t)$ is approximately Gaussian when the number of cells per well is large.*

Proof. At the beginning of the experiment, t_0 , $N_1(t_0)$ is a Gaussian random variable. For any sampling point, at time t_j , $N_1(t_j)$ can be expressed as

$$N_1(t_j) = N_1(t_{j-1}) + \Delta N_1(t_j) \quad (3)$$

where $N_1(t_{j-1})$ is the total number of shifted cells at time t_{j-1} , and the additional number of shifted cells in the time interval $[t_{j-1}, t_j]$ is given by

$$\Delta N_1(t_j) = \sum_{i=1}^{N-N_1(t_{j-1})} X_i \quad (4)$$

If $N - N_1(t_{j-1})$ is sufficiently large, $N - N_1(t_{j-1}) > 32$, then $\Delta N_1(t_j)$ is well approximated by a Gaussian random variable. Since $N_1(t_0)$ is Gaussian, $N_1(t_j)$ is Gaussian as well by mathematical induction. \square

Modeling the cell shifting process

From our previous experimental observation, the cell shifting process on colon cancer cell-line HCT116 with a reporter for the MKI67 gene under lapatinib treatment shows a binary shifting characteristic. It is assumed that the number of shifting cells is related to: (i) the drug effect corresponding to different dosages; and (ii) the number of proliferating cells (non-shifted cells, $N - N_1$). Since $N_1(t)$ is Gaussian process when the number of cells per well is large and N is a constant, the percentage of cells shifted at time t after applying drug, $\rho_1(t) = N_1(t)/N$, is a Gaussian process normalized to $0[1]$. Similarly, for the control group, $\rho_{1c}(t) = N_{1c}(t)/N_c$, is also a Gaussian process normalized to $0[1]$. Then $\rho(t) = \rho_1(t) - \rho_{1c}(t)$, the calibrated percentage of cells shifted at time t after applying drug, is a Gaussian process too. We are interested in the distribution of $\rho(t)$, specifically, how the mean value of $\rho(t)$, $\rho_{av}(t)$, changes along time under different dosage. Based on the above discussions, we propose the following model for cell shifting:

$$\frac{d\rho_{av}}{dt} = (\gamma_1^u + \mu_1)(1 - \rho_{av}) - (\beta + \mu_2)\rho_{av} + v \quad (5)$$

where γ_1^u is the drug effective coefficient depending on the dosage d , and $\beta > 0$ is a balancing factor. $\rho_{av}(t)$ changes along time since the corresponding random process $\rho(t)$ is non-stationary, thus its mean changes with time. Specifically, the change of $\rho_{av}(t)$ follows a linear differential equation (Eq.(5)) that reflects the fact that the change would be positively affected by the product of drug effectiveness and the percentage of cells not shifted

(1st term in Eq.(5)), and negatively affected by the percentage of cells already shifted (2nd term in Eq.(5)), thus the term ‘‘balancing factor’’ for β since more shifted cells mean less non-shifted cells that the drug may affect.

In this model, we assume that both γ_1^u and β change along time, thus the proposed model is a time-varying system. It is also assumed that the number of non-shifted cells, $N - N_1$, decreases exponentially with the factor γ_1^u . $\mu = [\mu_1 \mu_2]^T$ and v are independent Gaussian white noise processes. μ represents the process noise. Its covariance matrix is

$$E[\mu(n)\mu^T(k)] = \begin{cases} Q(n), & n = k \\ 0, & n \neq k \end{cases}$$

v is the measurement noise. Its covariance matrix is

$$E[v(n)v(k)] = \begin{cases} R(n), & n = k \\ 0, & n \neq k \end{cases}$$

The noise terms account for the various uncertainties introduced by the experiment. For instance, the cells may not be at the same cell cycle during the experiments, and thus may not be affected by the drug if some of the cells are actually dormant. This kind of uncertainties are modeled by process noise μ . There also exists another type of uncertainty due to measurement procedures, such as the imperfect photographic device and the image processing software. This type of uncertainty is modeled by measurement noise v .

To observe the relationship between the drug effect coefficient γ_1^u and the dosage d , we need to estimate γ_1^u for each dosage. Since this is a time-varying model, γ_1^u changes with time.

System identification from time-series data using Kalman filter

Kalman filtering [48] provides minimum-mean-square-error estimation of the state of a stochastic linear system disturbed by Gaussian white noise. In our proposed scheme, a Kalman filter is applied to estimate the coefficients, γ_1^u and β , of the proposed cell shifting model. The corresponding state and measurement equations are

$$w(n) = w(n - 1) + \mu(n - 1) \quad (6)$$

$$\delta(n) = C(n)w(n) + v(n) \quad (7)$$

where the 2-dimensional state vector (containing the parameters to be estimated) is $w = [\gamma_1^u \beta]^T$. δ can be calculated as $\delta(n) = \frac{\rho_{av}(n+1) - \rho_{av}(n)}{\Delta t}$. $C = [1 - \rho_{av} - \rho_{av}]$.

The implementation of the Kalman filter is given by the following equations [48]:

$$\hat{w}^-(n) = \hat{w}^+(n - 1) \quad (8)$$

$$P^-(n) = P^+(n - 1) + Q(n - 1) \quad (9)$$

$$\hat{w}^+(n) = \hat{w}^-(n) + K(n)[d(n) - C(n)\hat{w}^-(n)] \quad (10)$$

$$K(n) = P^-(n)C^T(n)[C(n)P^-(n)C^T(n) + R(n)]^{-1} \quad (11)$$

$$P^+(n) = P^-(n) - K(n)C(n)P^-(n) \quad (12)$$

where $K(n)$ is the Kalman filter gain and P is the covariance matrix of the error. The superscripts $-$ and $+$ indicate the *a priori* and *a posteriori* values of the variables, respectively. \hat{w}^- and \hat{w}^+ are the prior and posterior estimates, respectively. Q and R are the covariance matrices of the parameter noise and external noise, respectively. The initial conditions are $\hat{w}(0|\delta 0) = E[\hat{w}(0)]$ and $P_0 = E[w(0)w^T(0)]$.

In general, a Kalman filter may be interpreted as a one-step predictor with an appropriate gain calculator [49]. Specifically, Eq.(10) is the one-step predictor, Eq.(11) calculates the Kalman filter gain, and Eq.(12) solves the corresponding Riccati equation.

Convergence of the Kalman filter is an important issue [48]. The rate of convergence is defined as the number of iterations to obtain the optimum estimates. The convergence of the Kalman filter includes the convergence of the estimates $\hat{w}(n)$ and the convergence of the estimation error $e(n)$. Convergence will be studied in detail in the simulations.

In practice, noise statistics (such as the covariance matrices) may not be known and need to be estimated. The Kalman filter is sensitive to the estimation error of noise statistics. Poor estimates of the noise covariance can result in filter divergence. An alternative would be using an H_∞ filter [50,51].

Results

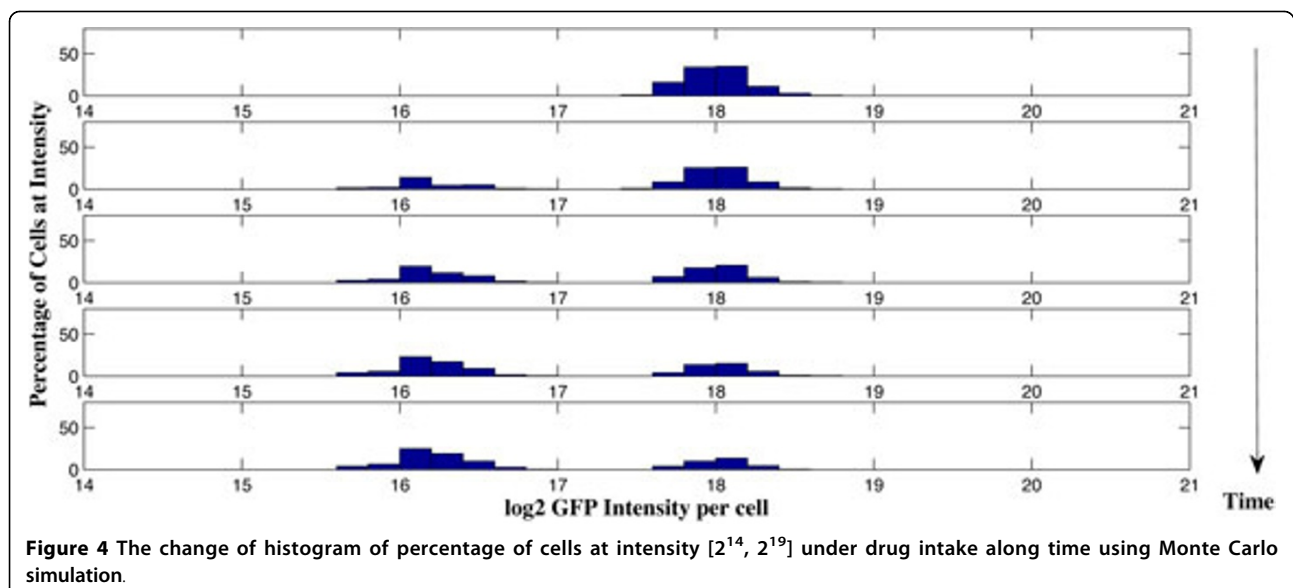
Two-step analysis is performed to evaluate the drug effect study for different dosages. Firstly, we performed a proof-of-concept experiment using Monte Carlo simulation to demonstrate that the proposed model can mimic experimental observation. Secondly, we analyzed the time-varying drug effect for different dosages based on real experimental data from Dr. Bittner's lab at Translational Genomics Research Institution (TGen).

Proof-of-concept experiment using Monte Carlo simulation

It is assumed that a group of 200 cells has mean GFP intensity at 2^{18} . When the drug is applied, each cell determines whether to shift to a lower intensity or not individually by flipping a coin (Bernoulli trial) at each time point, as we assumed in the theoretical model. The histograms of percentage of cells at intensity in the range of $[2^{14}, 2^{19}]$ along time are shown in Figure 4. It is observed that the resulting histograms from the Monte Carlo simulation of the theoretical model match the measurement results from the TGen experiments performed on the cell-line. This demonstrates that the cell shifting is probably a binary decision, which lays the ground for our proposed theoretical model where a group of cells' decision can be modeled as binomial and can be closely approximated by Gaussian distribution when the number of cells is large.

Drug effect analysis for the dosing study performed at TGen

For the experiments performed on the cell-line at TGen, there are 6 different dosages tested for the drug lapatinib, from $1\mu M$ to $32\mu M$. There are 6 biological duplicates



for each dosage and each biological duplicate contains 200 to 400 cells. The obtained experimental data set contains time-series data of the intensity readings for each cell per hour along a 48-hour period. There are also corresponding experimental data set of the control group (without drug) for the purpose of calibration. The calibrated percentage of shifted cells is used as measurement data in the proposed algorithm using Kalman filtering. The obtained estimates of the drug effect coefficient (γ_1^u) and the balancing factor (β) along 48 hours for 6 different dosages are shown in Figure 5 and Figure 6, respectively.

It is observed from Figure 5 that in general the drug effect coefficient (γ_1^u) increases with the applied dosage, as expected. It seems that there exist certain thresholds for γ_1^u . For instance, γ_1^u is much bigger with the dosages above $8\mu M$. It is also observed that γ_1^u increases with time as well. This reveals the time varying

nature of the drug effect. Furthermore, Figure 5 shows that higher dosage corresponds to faster response time, e.g., γ_1^u increases earlier and faster for higher dosage starting at ~ 10 hour. It is worth pointing out that, ideally, the percentage of shifted cells should be more than that in the control group without drug input, i.e., $0 \leq \rho(t) \leq 1$. However, due to uncertainties and noise in the experiments, we actually observe that $\rho(t)$ may be negative, especially during the first ~ 10 hours, before the drug is in effect.

Unlike γ_1^u , it is observed in Figure 6 that β remains roughly flat along time for a given dosage, because β is the balancing factor and should not change with time. However, β is different for different applied dosage, since higher dosage requires a higher balancing factor to maintain stability of the system. Again, the uncertainties and noise may dominate the system during the first ~ 10 hours (before the drug is in effect).

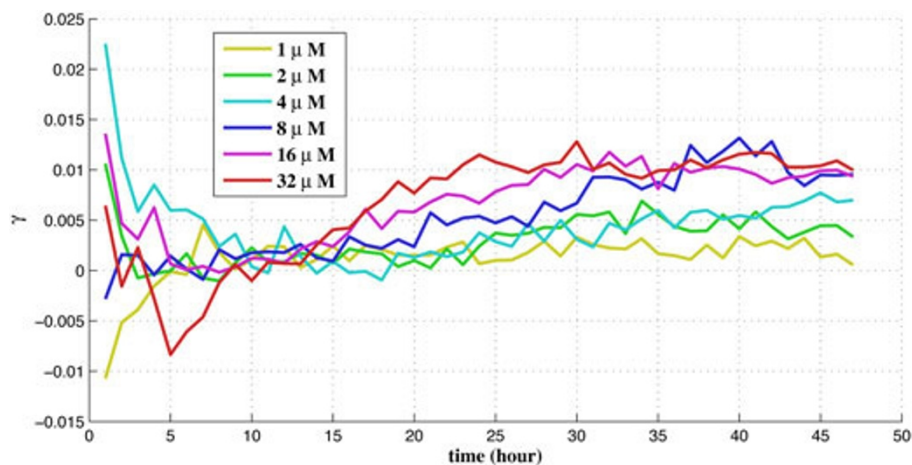


Figure 5 The estimate of the drug effect coefficient along time for 6 different dosages.

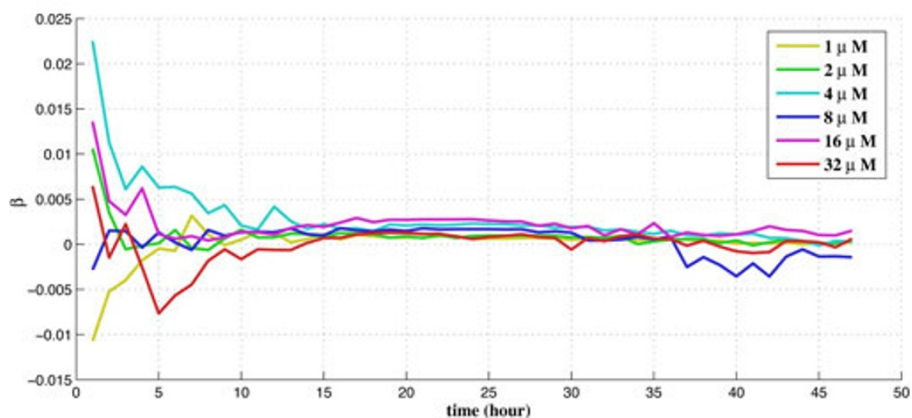


Figure 6 The estimate of the balancing factor along time for 6 different dosages.

Figure 7 shows the convergence of the Kalman filter. It converges in a few iterations in all cases.

Post data processing for the dosing study performed at TGen

From Figure 5 and Figure 6, it is observed that drug effect (γ_1^u) and the balancing factor (β) is very “jittery,” especially for the initial ~10 hours. Such a phenomenon may result from experimental noise, or that the cells may need certain “commitment time” after the drug is added. In order to better compare the drug effect for different dosages, we smooth the results and only take

into account data after the first 10 hours. We apply a moving-average filter with filter coefficients determined by an unweighted linear least-squares regression and a 2nd-degree polynomial model. The span for the moving average is 5. Figure 8 shows the smoothed drug effect coefficient (γ_1^u) along time for 6 individual dosages. It can be observed that the drug effect is more jittery for small dosages, such as $1\mu\text{M}$. The smoothed γ_1^u along time for 6 dosages are compared in Figure 9. It is observed that there exists a “plateau” ($\gamma_1^u \approx 0.01$) for higher dosages above $8\mu\text{M}$. The plateau is reached at 38 hours, 30 hours, and 24 hours, for dosages $8\mu\text{M}$, $16\mu\text{M}$,

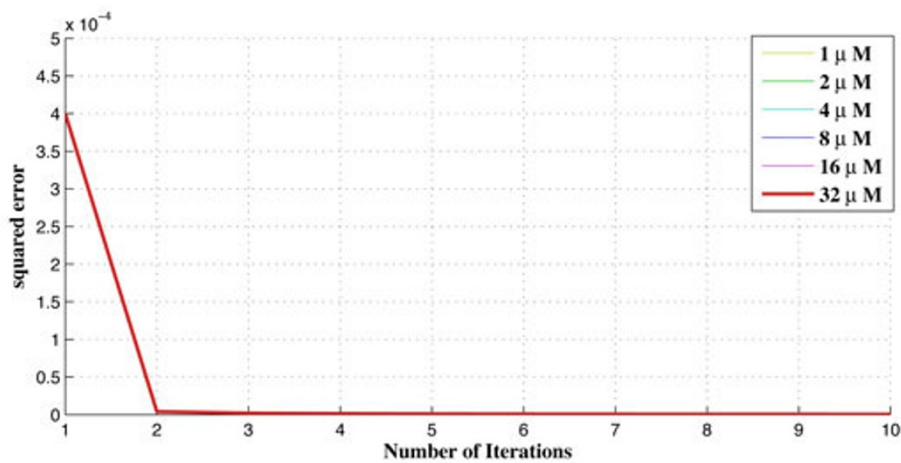


Figure 7 The Convergence result of the the proposed algorithm using Kalman filter.

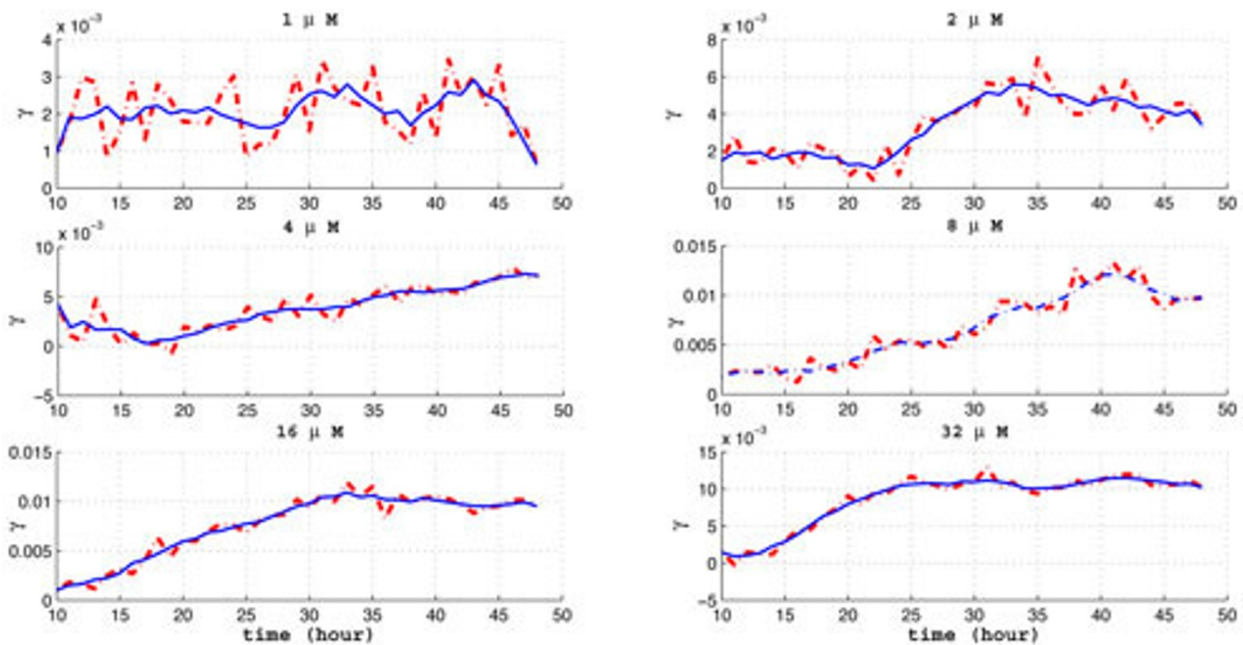


Figure 8 The smoothed drug effect coefficient along time for 6 individual dosage.

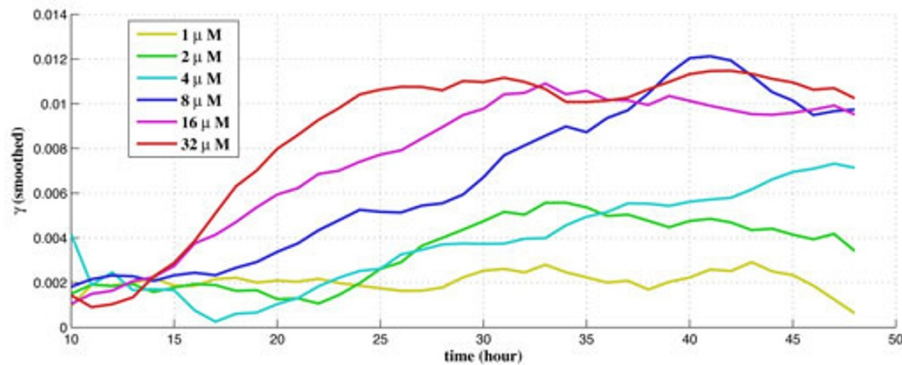


Figure 9 The smoothed drug effect coefficient along time for 6 different dosages.

and $32\mu\text{M}$, respectively. The smoothed balancing factor (β) for individual dosage can be found in Figure 10, and the smoothed β for 6 dosages are compared in Figure 11.

Conclusions and future work

The ultimate goal of target-based cancer drug development is to improve the efficacy and selectivity of cancer treatment by exploiting the differences between cancer cells and normal cells. The current cancer drug development process is confronting huge challenges, such as how to better understand the target in context and develop predictive preclinical models to better understand the molecular mechanisms of the biological systems they target and hence reduce the attrition rate. An integrated experimental and theoretical approach is proposed to assess the efficacy of molecularly targeted agents based on cell-line platforms. As a first step, drug efficacies for different dosages are

characterized along time. Specifically, tumor cell's responses are analyzed through the use of fluorescent reporters sampled frequently over a time course; quantification is done by microscopic scanning of cells in culture in multi-well plates using the automated epifluorescent imager; fluorescent intensity data are extracted from these images using specialized large-scale image analysis tools developed for this application; the dynamics of drug efficacy for different dosages are studied using dynamic modeling; and time-varying parameters are estimated using system identification techniques. It is observed that the drug efficacy is time and dosage dependent. The objectives are two-fold: (i) The dosing study for MTAs should be based on both efficacy and toxicity consideration to find the biologically effective dose (BED) instead of the maximum tolerated dose (MTD) for cytotoxic agents. The time course of drug effect for different dosages can provide

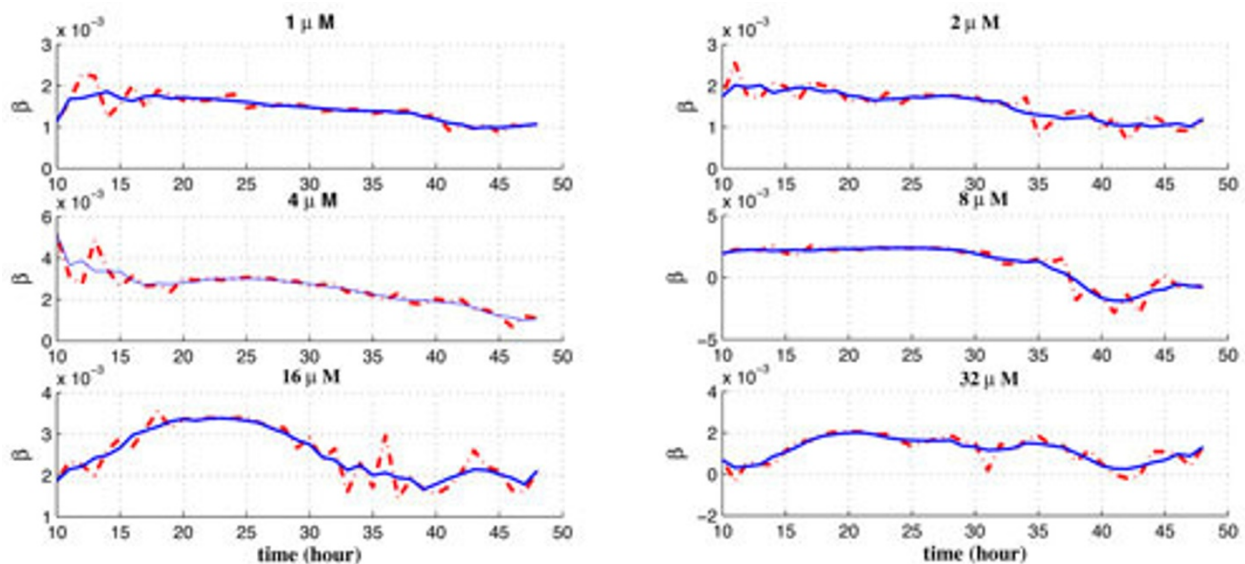


Figure 10 The smoothed balancing factor coefficient along time for 6 individual dosage.

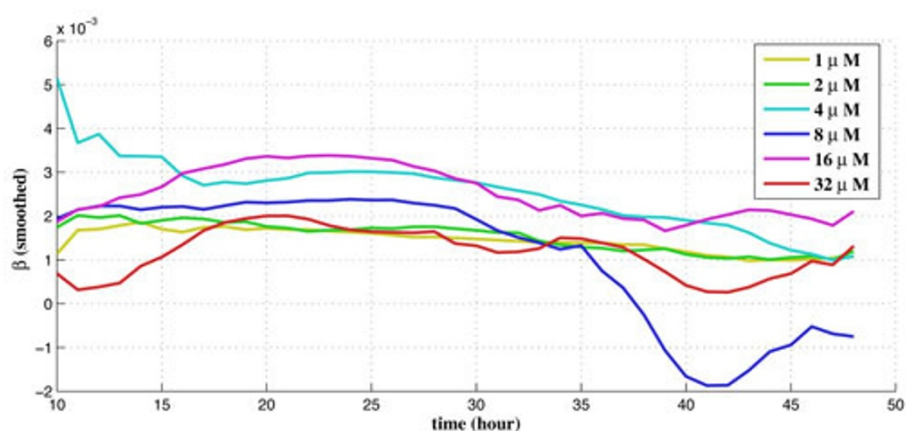


Figure 11 The smoothed balancing factor coefficient along time for 6 different dosages.

information on the gradient of drug effect vs. dosage, and thus on the BED. (ii) Instead of a fixed time point pharmacodynamics study of MTA, characterization of the entire time course of drug effect provides insight into designing an optimal schedule for drug administration.

Based on a similar experimental set-up and measurements to follow the cell/drug (dosages) dynamics, a truly multi-dimensional dynamics of tumor cell responses to drugs can be characterized through systematic perturbations to test different combinations of cell types, reporters, and drugs/dosages, augmented by iterative systematic theoretical analysis. Such an approach would facilitate the study of optimal dose and schedule, such as whether short pulses of higher dose, persistent dosing with lower dose, or some other regimen would have the most favorable outcomes. Moreover, the complex target context can be inferred with multi-dimensional cell response dynamics with the help of advanced system identification methods. In sum, better intervention strategies can be designed. Such topics are either currently being pursued or will be in future projects.

Acknowledgements

Based on "Assessing the efficacy of molecularly targeted agents by using Kalman filter", by Xiangfang Li, Lijun Qian, Michael L Bittner and Edward R Dougherty which appeared in *Genomic Signal Processing and Statistics (GENSIPS), 2011 IEEE International Workshop on*. © 2011 IEEE [37]. Xiangfang Li has been supported by the National Cancer Institute (2 R25CA090301-06). The experimental and image analysis work was supported in part by the W. M. Keck Foundation and Predictive Biomarker Sciences. This article has been published as part of *BMC Genomics* Volume 13 Supplement 6, 2012: Selected articles from the IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS) 2011. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/13/S6>.

Author details

¹Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA. ²Department of Electrical and Computer Engineering, Prairie View A&M University, Prairie View, TX 77446, USA.

³Computational Biology Division, Translational Genomics Research Institution, Phoenix, AZ 85004, USA. ⁴Department of Bioinformatics and Computational Biology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA.

Authors' contributions

XL and LQ developed and implemented the algorithm, conducted all simulations and data processing and wrote the initial draft of the paper. JH performed the image analysis. MB performed the experiments. ED advised XL on algorithm development and revised the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Published: 26 October 2012

References

1. FitzGerald GA: Re-engineering Drug Discovery and Development. *LDI Issue Brief* 2011, **17**(2):1-4.
2. DiMasi E, et al: The price of innovation: new estimates of drug development costs. *J Health Econ* 2003, **22**:151-185.
3. DiMasi J, Grabowski H: Economics of new oncology drug development. *Journal of Clinical Oncology* 2007, **25**(2):209-216.
4. Federsel HJ: In search of sustainability: process R&D in light of current pharmaceutical industry challenges. *Drug Discovery Today* 2006, **11**:966-974.
5. Kola I, Landis J: Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004, **3**(8):711-715.
6. DiMasi E, et al: The price of innovation: new estimates of drug development costs. *J Health Econ* 2003, **22**:151-185.
7. Hait WN: Anticancer drug development: the grand challenges. *Nature Reviews Drug Discovery* 2010, **9**:253-254.
8. Ocana A, Pandiella A, Siu LL, Tannock IF: Preclinical development of molecular-targeted agents for cancer. *Nat Rev Clin Oncols* 2010, **8**(4):200-209.
9. Twombly R: Cancer Surpasses Heart Disease as Leading Cause of Death for All But the Very Elderly. *Journal of the National Cancer Institute* 2005, **97**(5):330-331.
10. Horrobin D: Modern biomedical research: an internally self-consistent universe with little contact with medical reality? *Nat Rev Drug Discov* 2003, **2**(2):151-154.
11. Stoffels P: Collaborative Innovation for the Post-Crisis World. *The Boston Globe* 2009.
12. Vogelstein B, Kinzler K: Cancer genes and the pathways they control. *Nature Medicine* 2004, **10**:789-799.
13. Hanahan D, Weinberg R: The hallmarks of cancer. *Cell* 2000, **100**:57-70.

14. Communications S: **The sustainability of the current drug development process: barriers and new orientations.** *Ti Pharma Escher Workshop: Barriers to Pharmaceutical Innovation* Leiden, The Netherlands; 2010, 1-6.
15. Abbott A: **The drug deadlock.** *Nature* 2010, **468**:158-159.
16. Kummar S, Chen HX, Wright J, Holbeck S, Millin MD, Tomaszewski J, Zweibel J, Collins J, Doroshow JH: **Utilizing targeted cancer therapeutic agents in combination: novel approaches and urgent requirements.** *Nature Reviews Drug Discovery* 2010, **9**:843-856.
17. Li X, Qian L, Bittner M, Dougherty E: **Characterization of Drug Efficacy regions based on dosage and frequency schedules.** *IEEE Trans Biomed Eng* 2011, **58**(3):488-498.
18. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL: **How to improve R&D productivity: the pharmaceutical industry's grand challenge.** *Nature Reviews Drug Discovery* 2010, **9**:203-214.
19. Collins I, Workman P: **New approaches to molecular cancer therapeutics.** *Nature Chemical Biology* 2006, **2**:689-700.
20. Davidov E, Holland J, Marple E, Naylor S: **Advancing drug discovery through systems biology?** *Drug Discov Today* 2003, **8**(8):175-183.
21. Balis F: **Evolution of anticancer drug discovery and the role of cell-based screening.** *J Natl Cancer Inst* 2002, **94**(2):78-79.
22. Fox E, Curt G, Balis F: **Clinical trial design for target based therapy.** *The Oncologist* 2002, **7**(5):401-409.
23. Hait WN, Hambley T: **Targeted cancer therapeutics.** *Cancer Res* 2009, **69**:1263-1267.
24. Hambley T: **Is anticancer drug development heading in the right direction?** *Cancer Res* 2009, **69**(4):1259-1262.
25. Druker B: **Perspectives on the development of imatinib and the future of cancer research.** *Nature Med* 2009, **15**:1149-1152.
26. Vogel C, et al: **Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer.** *J Clin Oncol* 2002, **20**:719-726.
27. McClellan M, Benner J, Schilsky R, Epstein D, Woosley R, Friend S, Sidransky D, Geoghegan C, Kessler D: **An accelerated pathway for targeted cancer therapies.** *Nature Reviews Drug Discovery* 2011, **10**:79-80.
28. Dougherty E, Bittner M: *Epistemology of the Cell: A Systems Perspective on Biological Knowledge* Wiley-IEEE Press; 2011.
29. Sawyers C: **Translational research: are we on the right track?** *J Clin Invest* 2008, **118**(11):3798-3801.
30. Millar A, Lynch K: **Rethinking clinical trials for cytostatic drugs.** *Nature Reviews Cancer* 2003, **3**:540-545.
31. Johnston S: **Farnesyl transferase inhibitors: a novel targeted therapy for cancer.** *The Lancet Oncology* 2001, **2**:18-26.
32. Kummar S, Gutierrez M, Doroshow J, Murgu A: **Drug development in oncology: classical cytotoxics and molecularly targeted agents.** *British Journal of Clinical Pharmacology* 2006, **62**:15-26.
33. Kholodenko BN: **Cell-signalling dynamics in time and space.** *Nat Rev Mol Cell Biol* 2006, **7**:165-176.
34. Dougherty E, Brun M, Trent J, Bittner M: **Conditioning-Based Modeling of Contextual Genomic Regulation.** *IEEE/ACM Trans Comput Biol Bioinform* 2009, **6**(2):310-320.
35. Hill A: **The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves.** *J Physiol* 1910, **40**:iv-vii.
36. Holford N, Sheiner L: **Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models.** *Clin Pharmacokinet* 1981, **6**(6):429-53.
37. Li X, Qian L, Bittner ML, Dougherty ER: **Assessing the efficacy of molecularly targeted agents by using Kalman filter.** *Genomic Signal Processing and Statistics (GENSIPS), 2011 IEEE International Workshop on: 4-6 December 2011* 2011, 50-51.
38. Shah N, Kasap C, Weier C, Balbas M, Nicoll J, Bleickardt E, Nicaise C, Sawyers C: **Transient Potent BCR-ABL Inhibition Is Sufficient to Commit Chronic Myeloid Leukemia Cells Irreversibly to Apoptosis.** *Cancer cell* 2008, **14**(6):485-493.
39. Amin D, Sergina N, Ahuja D, McMahon M, Blair J, Wang D, Hann B, Koch K, Shokat K, Moasser M: **Resiliency and Vulnerability in the HER2-HER3 Tumorigenic Driver.** *Science Translational Medicine* 2010, **2**(16):16ra7.
40. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC: **Green fluorescent protein as a marker for gene expression.** *Science* 1994, **263**:802-805.
41. Hao L, Johnsen R, Lauter G, Baillie D, Brglin TR: **Comprehensive analysis of gene expression patterns of hedgehog-related genes.** *BMC Genomics* 2006, **7**:280.
42. Kanda T, Sullivan KF, Wahl GM: **Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells.** *Curr Biol* 1998, **8**:377-385.
43. Hua J, Chao S, Cypert M, Gooden C, Shack S, Alla L, Smith E, Trent JM, Dougherty ER, Bittner ML: **Tracking Transcriptional Activities with High-throughput Epifluorescent Imaging.** *Journal of Biomedical Optics* 2012, **17**(4):046008.
44. Dougherty E, Lotufo R: *Hands-on morphological image processing* SPIE Optical Engineering Press; 2003.
45. Vincent L, Soille P: **Watersheds in digital spaces: an efficient algorithm based on immersion simulations.** *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1991, **13**(6):583-598.
46. Walker R, Camplejohn R: **Comparison of monoclonal antibody Ki-67 reactivity with grade and DNA flow cytometry of breast carcinomas.** *Br J Cancer* 1988, **57**(3):281-283.
47. Spyrtatos F, et al: **Correlation between MIB-1 and other proliferation markers: clinical implications of the MIB-1 cutoff value.** *Cancer* 2002, **94**(8):2151-2159.
48. Grewal M, Andrews A: *Kalman Filtering: Theory and Practice* Englewood Cliffs, NJ: Prentice Hall; 1993.
49. Haykin S: *Adaptive Filter Theory (4th Ed)* Prentice Hall; 2001.
50. Shaked U, Theodor Y: **H_{∞} optimal estimation: A tutorial.** *IEEE CDC* 1992, 2278-2286.
51. Qian L, Wang H, Li X: **Genetic Regulatory Networks Inference: Combining a genetic programming and H_{∞} Filtering Approach.** *Applied Statistics for Network Biology: Methods in Systems Biology* Wiley; 2011, 133-153.

doi:10.1186/1471-2164-13-S6-S11

Cite this article as: Li et al: Assessing the efficacy of molecularly targeted agents on cell line-based platforms by using system identification. *BMC Genomics* 2012 **13**(Suppl 6):S11.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

