Predictive Modeling of Drug Effects on Signaling Pathways in Diverse Cancer Cell Lines

Shital Kumar Mishra[1], Sourav S. Bhowmick[1], Huey Eng Chua[1], and Jie Zheng[1,2,3,*]
[1]Biomedical Informatics Lab, School of Computer Science and Engineering,
Nanyang Technological University, Singapore 639798
[2]Genome Institute of Singapore (GIS), Biopolis, Singapore 138672
[3]Complexity Institute, Nanyang Technological University, Singapore 637723
s120018@e.ntu.edu.sg, assourav@ntu.edu.sg, chua0530@ntu.edu.sg, zhengjie@ntu.edu.sg
*To whom correspondence should be addressed : zhengjie@ntu.edu.sg

ABSTRACT

Apoptosis of a healthy cell is a process of programmed cell death regulated via well-established signaling pathways. However, in cancer cells apoptotic pathways are in atypical forms that lead to continuous survival, growth and proliferation of tumor cells. A major objective in cancer research is to investigate the dynamics of signaling pathways that influence the apoptosis of tumor cells. Inspired by the success of dynamical modeling and data analysis in cancer biology, in this study we propose a hybrid modeling approach combining computational models with experimental phosphoproteomics data. We construct a knowledge-based model of ordinary differential equations (ODEs) for the apoptotic signaling network and subsequently infer model parameters (e.g. reaction rates) from real phosphoproteomics data for three breast tumor cell lines, i.e., BT-20, MCF7 and MDA-MB-453 using a Bayesian framework of inference. The model is used to predict apoptosis in response to various perturbations such as caspase knockdown for each of the three cell lines which can be validated using the experimental literature. The inferred changes of the parameters reveal drug effects on diverse cell lines under the treatment with the drug of Erlotinib. Therefore, our hybrid modeling approach represents a novel method for understanding and predicting the impact of anti-cancer therapies on cancer cells at the systems level.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Time series analysis; G.2.2 [**Graph Theory**]: Network problems; I.2.6 [**Learning**]: Parameter learning; J.3 [**Life and Medical Sciences**]: Biology and genetics.

General Terms

Theory.

BCB '16, October 02 - 05, 2016, Seattle, WA, USA

ACM 978-1-4503-4225-4/16/10…$15.00

DOI: http://dx.doi.org/10.1145/2975167.2975188

Keywords

Cell fate; Apoptosis; Cancer; Signaling pathway; Ordinary differential equations; Context-specific modeling.

1. INTRODUCTION

Programmed cell death, also known as apoptosis, is a well-coordinated cellular process that is essential for the homeostatic growth and functioning of a multicellular organism as it removes damaged cells to avoid harmful effects such as toxicity [1]. Apoptosis is regulated by several signaling pathways and crosstalks among the pathways which are highly complex. In malignant cells these signaling pathways are dysregulated which allows the abnormal cells to grow and proliferate beyond their expected lifetime leading to diseases such as cancer [2, 3]. Therefore, in-depth understanding of the apoptotic signaling pathways is very important for finding effective treatments to selectively kill tumor cells. However, the complexity of the regulatory mechanisms of apoptosis (especially of cancer cells) remain elusive [1]. It is still difficult to model and simulate the dynamics of cellular signaling pathways partly due to insufficient understanding of the biochemical kinetics at the systems level. To address these challenges, researchers in computational systems biology have developed *in silico* systems modeling approaches to gain deeper insights into these pathways. These modeling approaches can be broadly classified into two basic types, i.e. knowledge-driven and data-driven modeling.

Numerous knowledge-driven dynamical models have been proposed to study apoptosis. For example, authors of [4] proposed a computational model consisting of 52 ordinary differential equations (ODEs), for estimating apoptosis based on activation of the mitochondrial signaling network. In [5], another model of apoptosis was proposed to analyze the effects of fast degradation of CASP8 and CASP3 proteins on apoptosis. The study suggests that a rapid increase in the activity level of CASP3 can be sufficient for initiating apoptosis. The authors of [6] proposed a computational model combining the strengths of ODEs and Boolean models for analyzing the behaviors of the NF-kB pathways. The study in [7] used boolean model to analyse Mitogen-Activated Protein Kinase (MAPK) pathways. The authors of [8] modeled signaling pathways of TNF and EGFR proteins, which were connected to the Gene Regulatory Network (GRN) inside nucleus and the computational model was encoded using ODEs.

On the other hand, data-driven statistical modeling ap-

proaches focus on the inference of model structures or parameters directly from experimental data. For example, the study in [2] used partial least squares regression (PLSR) to reveal anti-cancer drug effects on three breast cancer cell lines, i.e., BT-20, MCF7 and MDA-MB-453. In [9], a basal and signaling profiles for different treatments in the panel of "NCI-ICBP-43" cell lines were generated. The study identified several proteins that can be potential biomarkers for enhanced drug effect on tumor cells. The authors also used the PLSR method to predict the sensitivities of cancer cells to 23 targeted therapeutics. The significance of prediction was assessed using correlation between predicted and measured GI50 values. In [10], a comprehensive growth factor response dataset was analyzed to understand the roles of AKT and ERK proteins in the growth of breast tumor. The authors used unsupervised k-means clustering (where $k = 4$) to identify characteristics of signaling that varied across the dataset with respect to cell lines, ligands and receptors. The study in [11] provided a Bayesian approach to combining gene expression data with phosphorylation data, in response to 26 stimuli, to identify network components transmitting signals for a diverse set of stimuli in rat and human cells.

Although both the knowledge-driven and data-driven modeling approaches are promising to yield insights into cancer cellular signaling, it is desirable to combine the strengths of the two types of modeling approaches. In our previous study [12], we proposed a hybrid model that used ODE-based simulations and Genetic Algorithm for model selection to predict the drug-induced rewiring in the apoptotic signaling network in cancer cells. We changed the network topology while fixing the model parameters to find rewiring in response to drug-induced perturbations. Since the signaling networks inside the cells are complex, containing several interconnected pathways, there could be multiple causal effects responsible for drug sensitivity. For instance, changes in the kinetic rates may affect the drug sensitivity of tumor cells [13]. Thus, deriving model parameters from experimentally measured phosphorylation levels of signaling proteins can help in understanding the network dynamics and drug effects.

In this paper, we propose a new hybrid model of cancer signaling network, which is used to conduct new prediction and analysis of anticancer drug effects. Compared with [12], this paper presents significant extensions in the following aspects. First, the signaling network in [12] has been expanded to a larger and more comprehensive network that contains more proteins including kinases and phosphatase for phosphorylation and dephosphorylation of signaling proteins, respectively. Secondly, instead of using a genetic algorithm for heuristic optimization in model selection, here we use a well-studied software tool named ABC-SysBio [14]. ABC-SysBio is based on a framework of Bayesian learning that can infer the model parameters from real data. Thirdly, by integrative and comparative analysis of data from multiple cancer cell lines, we use our model to predict context-specific apoptosis. Moreover, we perform *in silico* caspase knockdown experiments for each of the aforementioned three cell lines and assess the regulatory impact of caspases on apoptosis. A major difference between this paper and [12] is that, while we focused on the inference of network rewiring in [12], here we infer parameters (i.e. reaction rate constants) for a fixed network topology. The rationale is that most events of network rewiring induced by drug treatment are edge deletions, which can be represented by the change of rate constants associated with network edges. In a sense, such network rewiring events can be considered as a special type of changes of parameters. However, we are aware that not all network rewiring can be represented by a change of parameter, such as adding a new edge. This issue will be discussed at the end of this paper. Our pipeline of studies is illustrated with the flowchart in Figure 1.

Given a knowledge-based ODE model derived from the literature, our method can infer parameters from experimental data by iterative optimization and also predict cell line specific changes of apoptosis in response to drug effects. Our results demonstrate that the hybrid modeling approach, which combines the knowledge-driven dynamical modeling and data-driven statistical modeling, can predict signaling activities and cell fates (apoptosis in this paper), more accurately than models without calibration to real data. The study, therefore, demonstrates the scalability of the proposed computational model to diverse cancer cell lines. In several studies, such as in [9], where experimental cell signaling data are available, but experimental apoptosis data are missing, our model can be used to predict the apoptosis and the drug effects. Considering the inherent complexity of cancer and the emergence of large amount of new "omics" data from cancer research, such a hybrid modeling strategy would be useful for understanding cancer and discovery of novel anti-cancer therapeutics. Our model which has been tested on the three different breast cancer cell lines can facilitate the development of "precision medicine" [15].

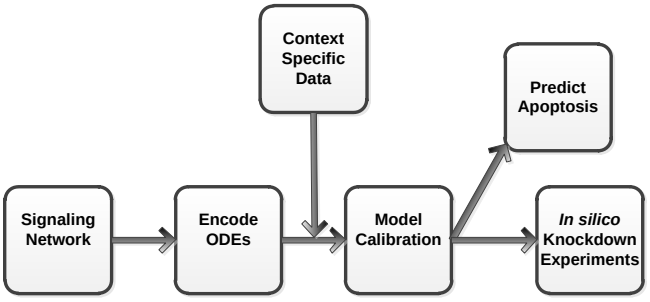


Figure 1: Flow diagram depicting the steps in our method. We derive the signaling network manually using GeneGO MetaCore database and encode the ODEs for the network using Michaelis-Menten kinetics. The ODE-based computational model is calibrated using cell line specific data. Subsequently, for each cell line, we predict the apoptosis and assess the impact of caspases on apoptosis using *in silico* knockdown experiments.

2. METHODS

2.1 Single-Cell Data

For our study, we used an experimental single-cell dataset, published by the group of Michael B. Yaffe [2], to calibrate and test our computational model of signaling network (Figure 1). This dataset contains time-course data for both cell signaling activities (measured by phosphorylation lev-

els) and its corresponding cell fates measured with flow cytometry. These data are available for three breast cancer cell lines, i.e., BT-20, MCF7 and MDA-MB-453. Phosphorylation levels of 35 proteins were recorded at 8 time points (numbered as 0, 1, 2,$\cdots$, 7). The indices from 0 to 10 correspond to the time points of 0, 0.1, 0.25, 0.5, 1, 2, 4, 6, 8, 12 and 24 hours, respectively, after the induction of the drug to the cancer cells. Out of the 35 proteins, however, data of only 32 proteins are available in the published dataset. The cell fate data were measured for 6 cellular phenotypes (i.e., Apoptosis, Proliferation, G1, S, G2 and M) in the three breast cancer cell lines over 24 hours at five time points corresponding to a subset of the aforementioned experimental time points (i.e. 0, 6, 7, 8 and 9). The measured signaling and cell fate data are synchronized with respect to the time points, e.g., the aforementioned time point 6 refers to the physical time of 4 hours after the drug induction for both signaling and cell fate measurements. Six types of drug treatments were applied to the cancer cells, i.e. Dimethyl sulfoxide (DMSO), Erlotinib (TAR), Doxorubicin (DOX), Doxorubicin and Erlotinib together (DT), Doxorubicin followed by Erlotinib (D-T), and Erlotinib followed by Doxorubicin (T-D). Then the phosphorylation levels of signaling proteins and cell phenotypes in response to the treatments were measured. The study in [2] reported that the level of apoptosis could be increased by as much as 500% in response to the T-D treatment compared with the other treatments. Henceforth, this dataset is referred to as the "Yaffe's dataset".

2.2 Derivation of Signaling Network Structure

Model construction of signaling network is a crucial step for subsequent simulation, prediction and data analysis. Our signaling network was derived using the GeneGO MetaCore database [16]. The diagram of the constructed network is shown in Figure 2. The network comprises a total of 71 nodes and 131 edges, in which there are 19 proteins with phosphorylation data available in the Yaffe's dataset [2]. The network includes kinases and phosphatases that catalyze the phosphorylation and dephosphorylation reactions, respectively. The types of reactions included in the network are phosphorylation, dephosphorylation and cleavage. Since our goal is to study the dynamics of phosphorylation of proteins in a signaling network, we did not represent protein complexes or regulation of gene expression (e.g. transcription and translation) explicitly. Our network includes several important proteins from different apoptotic pathways, such as BID, p53, CASP9, CASP3, CASP8, CASP6. Although in this paper we have focused on the study of responses of breast cancer cells to Erlotinib, this network model of signaling pathways could be adapted to study other aspects of apoptosis and the effects of other drugs.

2.3 Kinetic Modeling

For the signaling network manually constructed (Figure 2), we encode the kinetics in ODEs using the CellDesigner software [17]. The ODEs for biochemical reactions in the network were formulated based on the Michaelis-Menten kinetics [18]. The computational model comprises 71 ODEs and 163 reaction rate constants. The numerical solution of the ODEs in the form of time-series data was obtained using the MATLAB ODE solver. We denote the activity level of protein X as $[X]$, the rate constant for the ith catalysed reaction as k_i and Michaelis-Menten rate constant by m_i. For example, the ODE for $pSTAT3$ with respect to time t is given as:

$$\begin{aligned}\frac{d}{dt}[pSTAT3] = \; & k_1 \cdot [STAT3] \cdot [pEGFR]/(m_1 + [STAT3]) \\ & + \; k_2 \cdot [STAT3] \cdot [pCHK1]/(m_2 + [STAT3]) \\ & + \; k_3 \cdot [STAT3] \cdot [pP38]/(m_3 + [STAT3]) \\ & - \; k_4 \cdot [pSTAT3] \cdot [pPH10]/(m_4 + [pSTAT3]),\end{aligned}$$

where $[pSTAT3]$ and $[STAT3]$ are the concentrations of the phosphorylated and unphosphorylated forms of the $STAT3$ protein. Here we assume that the overall concentration of each protein (i.e. both phosphorylated and unphosphorylated) in a cell is stable, so that the phosphorylation levels (measured as the percentage of copies of each protein in the phosphorylated form) can represent the cellular state. The steady state constraint is mathematically encoded as :

$$\frac{d}{dt}[pSTAT3] + \frac{d}{dt}[STAT3] = 0.$$

2.4 Context-Specific Modeling

A long-standing challenge for dynamic modeling in systems biology is parameter estimation, for which various methods have been proposed [19, 20, 21]. Several software have been developed based on Bayesian inference for parameter estimation and model ranking using likelihood function [22]. However, for complex ODE models, deriving the likelihood function is not always possible. To address this issue, we use the software suite called ABC-SysBio [14]. A very important feature of ABC-SysBio is that it is likelihood-free by using the framework of Approximate Bayesian computation (ABC). Moreover, ABC-SysBio is able to handle missing data. A critical issue for parameter inference is overfitting of model with the experimental data. However, parameter estimation methods based on Bayesian inferential techniques are promising to address the issue of overfitting [14].

Using the Bayesian inference based parameter estimation, we are able to calibrate the knowledge-derived network model to real data, and thereby make the model context-specific. To this end, we incorporate cell line specific data for our model calibration. A breast cancer subtype can be caused by mutations in thousands of genes [23]. To understand context-specific drug sensitivity of breast cancer, data from diverse breast cancer cell lines can be analysed together [24]. In this study we use a generic network topology in Figure 2 to predict drug effects in the three breast cancer cell lines, i.e., BT-20, MCF7 and MDA-MB-453. The model is calibrated with signaling data (i.e., independent variables) from each of the three cell lines. Then we predict the cell line specific apoptotic response (i.e., dependent variable) and assess the correlation between *in silico* and experimental apoptosis data. To study the impact of caspases on apoptosis, we perform *in silico* caspase knockdowns experiment on each of the three cell lines. The knockdown simulations show that different cell lines have different signaling dynamics which can be caused by a combination of overactivation and suppression of the network edges in the cancer signaling pathways.

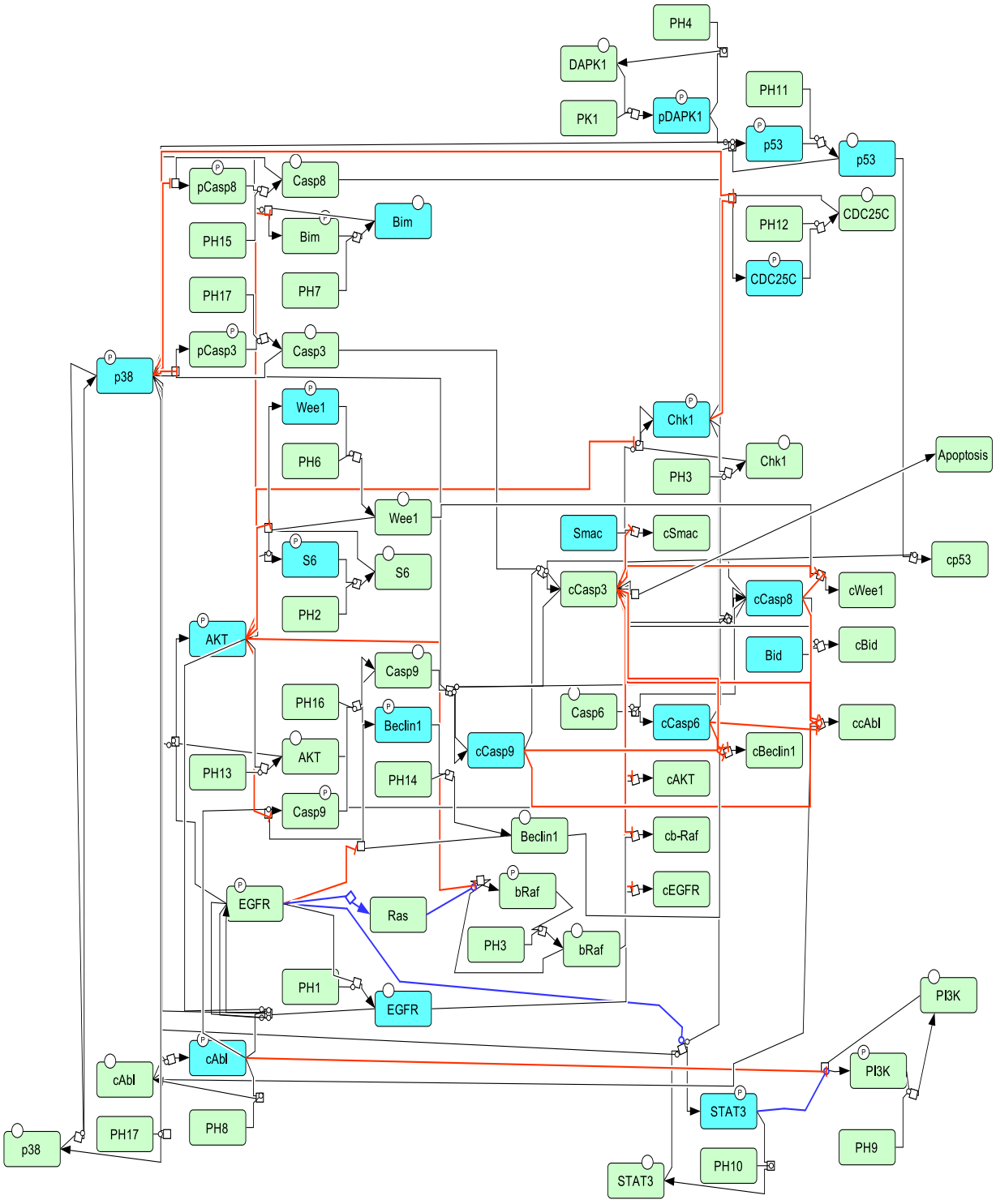


Figure 2: Apoptosis signaling network. The network kinetic parameters were inferred from the phosphoproteomics data from [2]. Nodes in cyan color (total 19) represent proteins with phosphoproteomics data available in Yaffe's dataset. The black and red colored edges represent the activation and inhibition of the signals, respectively. The blue edges indicate the inhibitory drug effect induced by Erlotinib on EGFR signaling pathway, i.e., blockage of signal transduction in BT-20 cells as inferred from the drug-sensitive phosphorylation dataset in [2]. The same model, after calibration with cell line specific drug sensitive data from [2], was used to analyse drug effects on three breast cancer cell lines, i.e., BT-20, MCF7 and MDA-MB-453.

3. RESULTS

3.1 Simulations with Random and Inferred Parameters

To demonstrate the importance of parameter inference from real data, we first simulated signaling and phenotypic responses using the network model with 1000 sets of randomly generated parameters, and compared the predictions with the experimental data in [2] for the BT-20 cell line. The initial phosphorylation level for each protein was fixed to one while the rate constants were varied with uniformly generated random numbers. We computed the sum of Euclidean distances between the simulated and experimental data for each of the 19 proteins. Euclidian distances for signaling proteins were calculated by comparing signaling levels of the 19 proteins in Yaffe's dataset with simulated signaling data at 8 time points (numbered as 0, 1, 2,$\cdots$, 7). Since the experimental data were missing for quite a few time points (*i.e.* 1, 2, 3, 4, 5), we compared the simulated and real time-series datasets only on the available time points. For apoptosis, the Euclidean distance was calculated by comparing the experimental apoptosis data with simulated apoptosis data at 5 time points (*i.e.* 0, 6, 7, 8 and 9). The distribution of the Euclidean distances for the signaling data of the 19 proteins and apoptosis, based on the 1000 sets of random parameters, are shown in Figure 3. Next, we will calibrate the model using the Yaffe's phosphorylation dataset and show that, after model calibration, the goodness of fit between predicted and real apoptosis data can be improved.

To infer parameters from the real data, the input for the ABC-SysBio software was a vector containing phosphorylation levels of 19 proteins from the Yaffe's dataset which was iteratively compared with a vector of simulated signaling data. The output of ABC-SysBio was a vector of estimated model parameters that optimized the goodness of fit between the simulated time-series data with the real data. The initial phosphorylation levels for the 19 proteins in the model were set to the initial levels in the Yaffe's data at time point 0. For other proteins in the network, of which the initial phosphorylation levels were unknown, the initial phosphorylation level was set to the dimensionless numerical value of one [25]. Apoptosis data in [2] was not used for the parameter inference. There are totally 163 parameters in the network model, and all of them were inferred. The nabla symbols (in red) in Figure 3 represent the Euclidean distance between the simulated and experimental data after model calibration, which shows the improvement in goodness of fit for signaling and apoptosis data after model calibration. The plots of the predicted time-course phosphorylation data and apoptosis data (after parameter inference) along with corresponding real data are shown in Figure 4, and 5(a), respectively.

3.2 Prediction of Apoptosis

Apoptosis in a cell depends on various molecular processes including signal transduction, gene regulation and metabolic activities. We aim to test if our model, after being trained with only a limited amount of signaling data from [2], would be able to predict changes in apoptosis accurately when compared with observed *in vitro* measurements in drug sensitive cells. The experimental dataset in [2] contains both signaling and cell fate data. We have used only the signaling data (excluding the cell fate data) to train our model. Initially we predicted apoptosis using the model in Figure 2 with random parameters for 1000 times. The distribution of the Euclidean distances between the simulated data (using random parameters) and real data of apoptosis is shown in the bottom-right panel of Figure 3. We also predicted apoptosis using the calibrated model with parameters inferred from real signaling data (but not apoptosis data), and compared the predicted time-series data with the real data of apoptosis. The comparison result for the BT-20 cell line is shown in Figure 5(a). The Euclidean distance between Yaffe's real data of apoptosis and the data predicted by the trained model was 18.8985. Both the simulated and experimental apoptosis levels were found to be increasing with time. There is a strong correlation between two time series, with Pearson correlation coefficient 0.994. These observations indicate a close association between *in vitro* and *in silico* phenotypic observations of programmed cell death.

We used a similar training procedure to calibrate the computational model in Figure 2 with the time-course drug sensitive data for the hormone-sensitive Michigan Cancer Foundation-7 (MCF7) cell line from [2]. The MCF7 cancer cell line is insensitive to the ERL/DOX combination of drug treatment, which otherwise, caused high apoptosis level in BT-20 cells. Our calibrated model was able to predict the apoptosis with the Pearson correlation coefficient of 0.9569. The predicted apoptosis is shown in Figure 5(b). We further trained the network model in Figure 2 with drug sensitive data for the MDA-MB-453 cell line. The predicted apoptosis for the MDA-MB-453 cell line is shown in Figure 5(c). The Pearson correlation coefficient is 0.90, indicating a very strong linear correlation between the *in silico* and experimental apoptosis data.

3.3 Virtual Knockdown of Caspases

The caspase proteins are key players in regulating apoptosis. Caspases can be categorized into Initiator (*e.g.* CASP8, CASP9) and Executioner caspases (*e.g.* CASP3, CASP6). Initiator caspases are responsible for activating executioner caspases via chain reactions. Once activated, Executioner caspases degrade several cellular components in order to induce the morphological changes for apoptosis [26]. Enhanced cell death in cancer cells is often driven by caspases after EGFR inhibition by anti-cancer drugs like Erlotinib. Pro-apoptotic molecules, such as CASP8 and CASP6, have been found to be important variables for regulating apoptosis in the BT-20 cell line [2]. To further understand the roles of caspases in the regulation of apoptosis, we performed *in silico* knockdown of CASP8 and CASP6. The predicted effects of knockdown in the BT-20 cell line are shown in Figure 6(a). We did three knockdown experiments, namely, CASP8 knockdown, CASP6 knockdown and the simultaneous knockdown of both CASP8 and CASP6. We set the activity level of CASP8 or CASP6 to zero, initially one by one, and then simultaneously both. All other model parameters were kept unchanged during the knockdown experiments. A decrease in apoptosis level was observed for each of the knockdown experiments. The decrease in the apoptosis level was higher for CASP8 knockdown than for CASP6 knockdown for the BT-20 cell line. The effect of combined CASP8

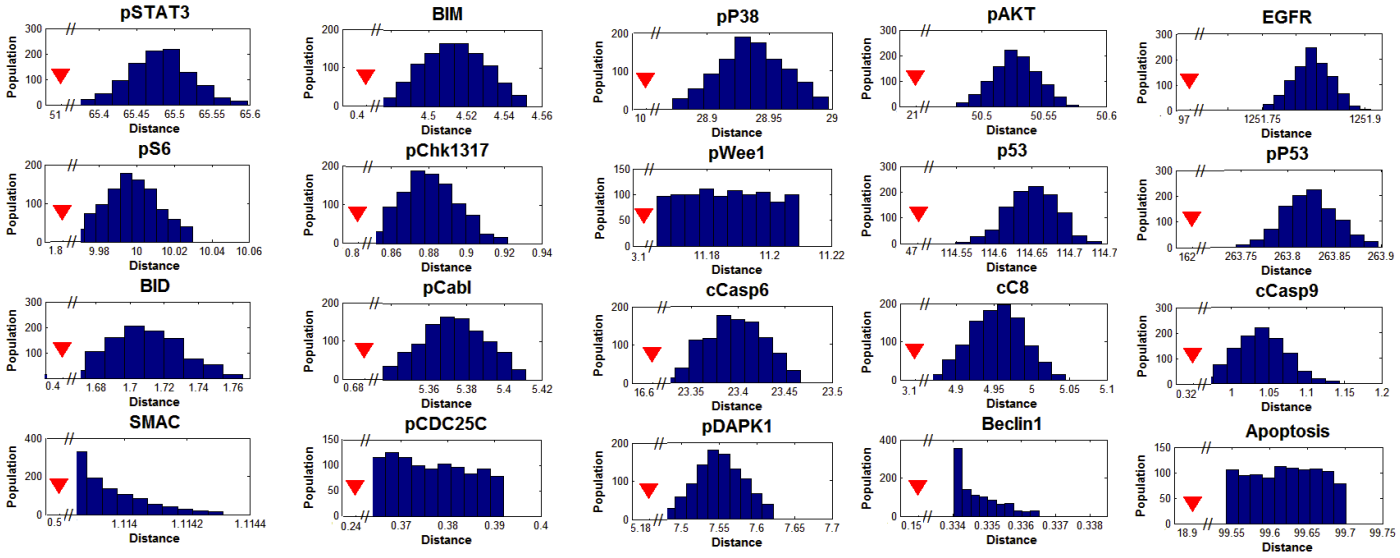


Figure 3: Distributions of Euclidean distances (in blue) between the simulated and the Yaffe's data for each of the 19 signaling proteins and apoptosis (bottom-right panel) for 1000 sets of random parameters for the BT-20 cell line. The nabla symbol (in red) represents the Euclidean distance between the simulated and experimental data after model calibration.

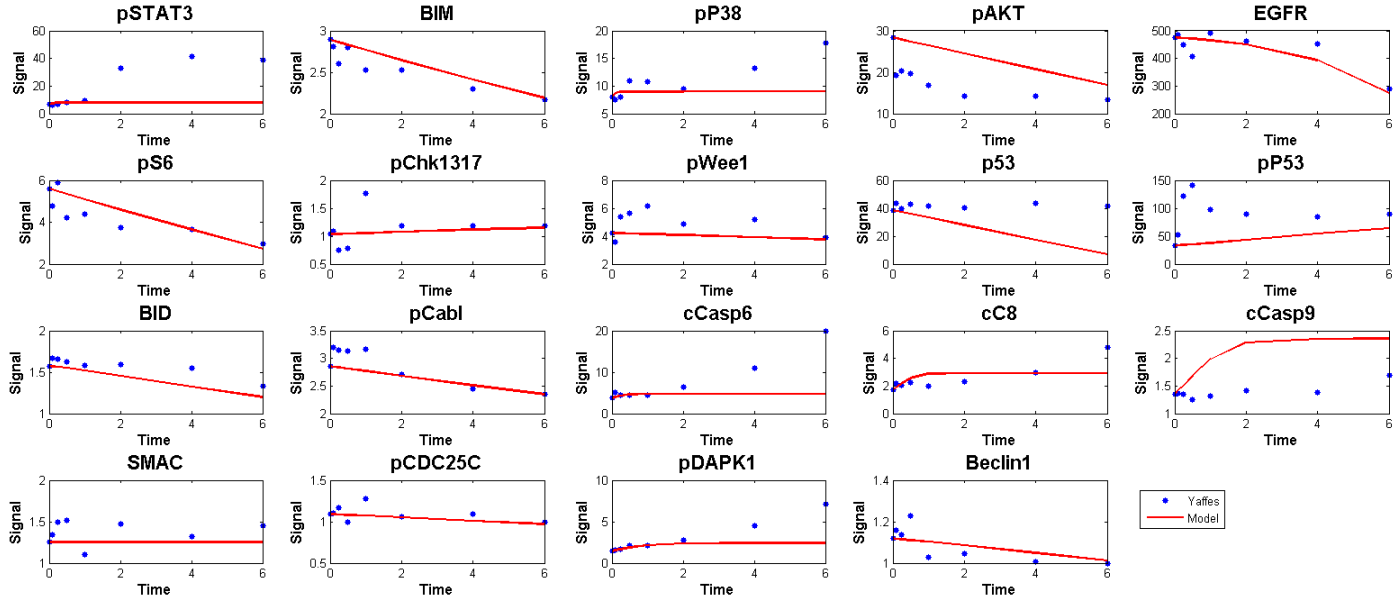
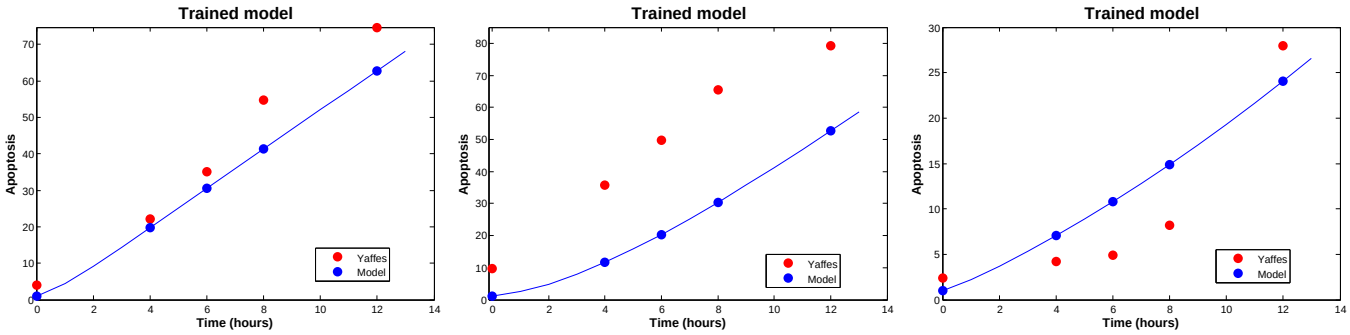


Figure 4: Fitting of simulation data to experimental data for the 19 proteins for the BT-20 cell line from [2] based on inferred parameters for the signaling network.

and CASP6 knockdown was higher than individual CASP8 or CASP6 knockdown. The simulation result suggests that the caspases were important molecules for regulating apoptosis. There is evidence of CASP8 contributing to enhanced cell death when EGFR is suppressed [27, 28].

We also simulated CASP8 and CASP6 knockdown in the MCF7 cells. We found that the CASP8 and CASP6 knockdown did not affect the apoptosis in the MCF7 drug sensitive cells. The apoptotic responses, before and after knockdowns, were similar at different time points indicating that these two knockdowns had almost no impact on the enhanced drug sensitivity. The effects of CASP8 and CASP6 knockdown on apoptosis in trained model for the MCF7 cell line are shown in Figure 6(b). Despite caspases knockdown, the apoptosis remained similar to the apoptosis before caspases knockdownas. The different rate constants for edges downstream of EGFR are shown in Table 1. The rate constants in the trained model for the edges CASP6-CASP8, CASP8-CASP9

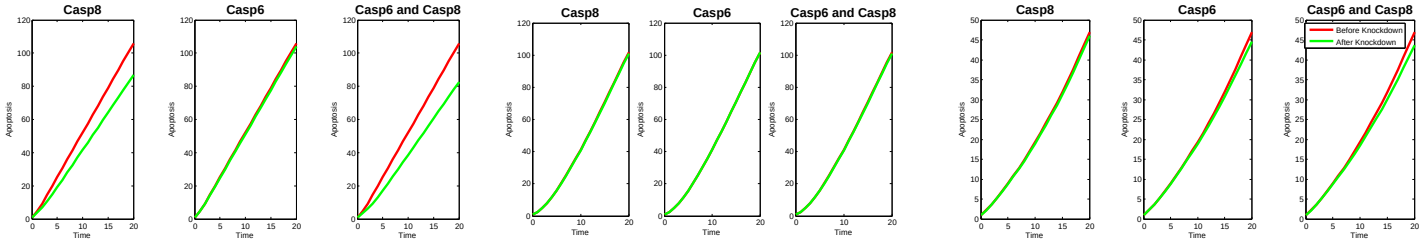


(a) BT-20 **(b) MCF7** **(c) MDA-MB-453**

Figure 5: Prediction of apoptosis and comparison of cell line specific computational apoptosis data with experimental apoptosis data from [2] for three cell lines, i.e., BT-20, MCF7 and MDA-MB-453. The calibrated computational model in Figure 2 was able to predict the cell line specific apoptosis with high Pearson correlation coefficients, such as 0.994 in BT-20, 0.0.9569 in MCF7 and 0.904 in MDA-MB-453 cell lines.

Table 1: Activation rate constants downstream of EGFR among different breast cancer cell lines

Network edge	BT-20 (DMSO)	BT-20 (Drug)	MCF7 (DMSO)	MCF7(Drug)	MDA-MB-453 (DMSO)	MDA-MB-453 (Drug)
EGFR-RAS	0.99994	6.53E-06	0.59745	6.63E-06	0.39985	9.64E-06
RAS-bRaf	0.79519	7.12E-06	0.39931	6.55E-06	0.54668	9.51E-06
EGFR-Stat3	1.1956	0.00089811	0.49992	0.19786	0.69579	9.99E-06
Stat3-PI3K	0.14384	7.73E-06	0.29652	9.37E-06	0.29368	9.96E-06
CASP6-CASP8	0.094878	0.20336	3.31E-06	6.89E-06	0.049838	0.09958
CASP8-CASP3	0.096599	0.50483	0.00097789	0.0072124	0.00073772	0.0099999
CASP8-CASP9	0.0090839	0.60083	3.91E-06	7.88E-06	0.00090093	9.85E-06
CASP9-CASP3	0.89505	1.5073	0.29154	0.69989	0.24888	0.49941



(a) BT-20 **(b) MCF7** **(c) MDA-MB-453**

Figure 6: Effects of CASP8 and CASP6 knockdown on apoptosis in trained models for the BT-20, MCF7, and MDA-MB-453 cell lines. The increase of apoptosis is weakened in response to the knockdown in BT-20 and MDA-MB-453 cells. Moreover, the combined effect of knockdown of CASP6 and CASP8 was higher than the individual knockdown. However, the caspase knockdown has little impact on apoptosis of MCF7 cell line as the apoptotic curves for CASP6 and CASP8 knockdown remained similar to the apoptosis before the knockdown.

and CASP8-CASP3 were 6.889e-06, 7.8777e-06 and 7.2124e-03, respectively. This indicates that, due to the blockage of signals, the CASP6 and CASP8 were not able to send pro-apoptotic signal to induce cell death, and therefore CASP6 and CASP8 were not significant drivers of apoptosis in the drug sensitive MCF7 cells. However, for CASP9-CASP3 the rate constant was 0.69989, indicating that CASP9 may play an active role in stimulating apoptosis in the MCF7 cells. There are several pieces of experimental evidence indicating the down regulation of CASP8 in different breast cancer cells including MCF7, MB231, SKBR3 and HCC1937 [2, 29]. Promoter methylation has been reported to be one important reason for reduced CASP8 signaling profile in MCF7 cell line [29]. The study in [2] found that CASP6 was not a strong driver of apoptosis in MCF7 cell line. The activation of the CASP9 in MCF7 cells following treatment with DNA damaging agents has been reported in [30].

Subsequently, we performed similar virtual knockdown of CASP6 and CASP8 for the MDA-MB-453 cell with results, shown in Figure 6(c). We found that CASP6 knockdown has slightly more impact on apoptosis than CASP8. The rate constants for the caspase edges (Table 1) CASP6-CASP8, CASP8-CASP9 and CASP8-CASP3 were 0.09958, 9.9999e-03 and 9.8506e-06, respectively. Since the rate for CASP6-CASP8 was higher than CASP8-CASP3 and CASP8-CASP9, CASP6 seems able to stimulate apoptosis more strongly than CASP8. The edge CASP9-CASP3 with rate constant of 0.49941 may also contribute to the apoptosis. In summary, our virtual knockdown experiments suggest that CASP6 appeared to be more effective in inducing apoptosis for the BT-20 and MDA-MB-453 cell lines, but not for the MCF7 cell line. By contrast, CASP8 seems not important for regulating apoptosis of the MCF7 and MDA-MB-453 cells. CASP9 is predicted to play an important role in stimulating apoptosis for all the three cell lines.

3.4 Predicting Drug-Induced Pathway Alterations

Although in this paper we have focused on the parameter estimation on a network of fixed topology, the predicted changes of parameters (in Table 1) can be used further to predict network rewiring events. For example, if a rate constant in Table 1 is lower than a threshold of say 0.001, then we consider the corresponding edge deleted from the network. In Figure 2, we plotted several blue arrows to represent such edge deletion events, which represent predicted drug effects where signal transductions were inhibited in the downstream of EGFR, according to the predicted values of rate constants in Table 1. The activation rate constants in the blue edges were significantly low. For example, in BT-20 cells following the drug treatment, the rate constants for the edges EGFR-RAS, RAS-bRAF, EGFR-STAT3, STAT3-PI3K, CASP8-CASP9, and CASP9-CASP3 were 6.5294e-06, 7.1158e-06, 8.9811e-04, 7.7281e-06, 0.60083, and 1.5073, respectively (Table 1).

The inferred parameters show that in the DMSO traetment data (that basically represented cancer cells), several edges (*e.g.* EGFR-RAS) that are known to be causal for breast cancer, were transmitting signals at a higher rate, whereas the signaling from CASP6 and CASP8 were inhibited. By contrast, the signaling from CASP9 remained effective for different treatment. The model calibration predicted higher rate constants for several edges which could be activated by the inhibition of EGFR. Moreover, the inferred rate constants suggest that the signaling from the Initiator (e.g. CASP8) to the Executioner Caspase (e.g. CASP3) could be blocked in the cancer cells. However, in the drug sensitive cells, the signaling between the Initiator and Executioner caspases can be activated. Our simulation-based findings about the blockage and activation of signaling pathways as drug effects may be used to explain some experimental observations reported in the literature [31, 32, 33].

4. CONCLUSION AND DISCUSSION

In this paper, we proposed a hybrid modeling approach which integrates dynamical ODE modeling with data-driven statistical learning to uncover the mechanism of anti-cancer drug effects on cancer cells. The ODE based computational model of signaling pathways was calibrated for three breast cancer cell lines, i.e., BT-20, MCF7 and MDA-MB-453, by iteratively fitting to the cell-line specific phosphoproteomics data. The calibrated model was then used to simulate the cellular responses (e.g. cell death) to different drug-induced perturbations. We also identified the blockage in the pathways downstream of EGFR leading to enhanced drug sensitivity in the BT-20 cells. Our results of comparing model prediction with real data suggested that model calibration by Bayesian inference of parameters from real data can improve the accuracy in predicting time-course signaling and apoptosis data. Our virtual experiments of caspase knockdown have also yielded some insights into the mechanisms of apoptotic regulation. The changes in activation rates inferred from the real data can be used to map network rewiring events induced by the anticancer drug of Erlotinib. In conclusion, our hybrid modeling method can take into account both prior knowledge and real data, to capture the context-specific dynamics of cancer cell fate. The models constructed and calibrated as such can be used to do simulation, prediction, and data analysis. Therefore, they will be instrumental for the precision medicine, as more biomedical knowledge is accumulated and more patient-specific data become available.

Despite the promising performance and prospect of the hybrid modeling in cancer systems biology, a few challenges remain open. Without attempting to be complete, we list some of the challenges here, as a discussion about future work. First, to calibrate a dynamic model against time-series real data, it is crucial to align time points between simulated and real time-course data. This is challenging for multiple reasons, such as, there may be missing time points in the real data; cellular processes tend to be multiscale by nature (e.g. changes in signal transduction occur much faster than gene expression or apoptosis). Secondly, the accurate inference of network models from limited amounts of data could be hindered by issues like overfitting, lack of identifiability, etc. Different models might fit the same dataset equally well, if the information from the real data is insufficient to distinguish the models. Hybrid modeling that takes into account of prior knowledge may help address this issue, but the prior knowledge itself might be incomplete or biased. Thirdly, the relation between network structures and parameters has long been a fundamental question in systems biology. In this paper, we assumed that a fixed network structure with parameters being estimated from real data could capture the context-specific alterations in cancer

cells. Moreover, we have attempted to reconcile the parameter changes with network rewiring by predicting edge deletions from the activation rate constants that were inferred to be low. However, this may not be sufficient to fully address the issue, as some network rewiring events like the addition of a new edge cannot be represented by the change of parameters of the existing edges. Fourthly, different proteins in a network could have very different ranges of signaling activities, thus an appropriate normalization method would be needed to avoid potential bias in the optimization process of parameter inference from real data. This is more crucial when there is noise in the measurement, especially for the single-cell data such as the Yaffe's dataset used in this paper. Another long-standing challenge is how to validate computational predictions. In this paper, we have mainly used qualitative comparisons of real and predicted time-course data and cited evidence from the literature. Such types of validation are indirect and may not be conclusive in some cases. Last but not the least, for the context-specific modeling of cancer cell fate and drug effects, we need to integrate diverse types of data. Here we have mainly used the phosphoproteomics data. Other types of data, such as somatic mutations, copy number variations, epigenetic modifications, gene expression, etc., should also be taken into account. Indeed, the fusion of dynamic network models with genomic information would be particularly important for the applications of computational techniques in precision medicine.

5. ACKNOWLEDGMENTS

This research has been supported by MOE AcRF Tier 2 Grant ARC39/13 (MOE2013-T2-1-079) from Ministry of Education, Singapore.

6. REFERENCES

[1] Susan Elmore. Apoptosis: a review of programmed cell death. *Toxicologic pathology*, 35(4):495–516, 2007.

[2] Michael J Lee, S Ye Albert, Alexandra K Gardino, Anne Margriet Heijink, Peter K Sorger, Gavin MacBeath, and Michael B Yaffe. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, 149(4):780–794, 2012.

[3] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.

[4] Heinrich J Huber, Markus Rehm, Martin Plchut, Heiko Düssmann, and Jochen HM Prehn. Apopto cell:a simulation tool and interactive database for analyzing cellular susceptibility to apoptosis. *Bioinformatics*, 23(5):648–650, 2007.

[5] Chang Gu, Junjie Zhang, Yingyu Chen, and Jinzhi Lei. A trigger model of apoptosis induced by tumor necrosis factor signaling. *BMC systems biology*, 5(1):1, 2011.

[6] Madalena Chaves, Thomas Eissing, and Frank Allgöwer. Regulation of apoptosis via the NFκB pathway: modeling and analysis. In *Dynamics On and Of Complex Networks*, pages 19–33. Springer, 2009.

[7] Luca Grieco, Laurence Calzone, Isabelle Bernard-Pierrot, François Radvanyi, Brigitte Kahn-Perles, and Denis Thieffry. Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput Biol*, 9(10):e1003286, 2013.

[8] Milsee Mol, Milind S Patole, and Shailza Singh. Immune signal transduction in leishmaniasis from natural to artificial systems: role of feedback loop insertion. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1840(1):71–79, 2014.

[9] Mario Niepel, Marc Hafner, Emily A Pace, Mirra Chung, Diana H Chai, Lili Zhou, Birgit Schoeberl, and Peter K Sorger. Profiles of basal and stimulated receptor signaling networks predict drug response in breast cancer lines. *Science signaling*, 6(294), 2013.

[10] Mario Niepel, Marc Hafner, Emily A Pace, Mirra Chung, Diana H Chai, Lili Zhou, Jeremy L Muhlich, Birgit Schoeberl, and Peter K Sorger. Analysis of growth factor signaling in genetically diverse breast cancer lines. *BMC biology*, 12(1):20, 2014.

[11] Chunhui Cai, Lujia Chen, Xia Jiang, and Xinghua Lu. Modeling signal transduction from protein phosphorylation to gene expression. *Cancer Informatics*, 13(Suppl 1):59, 2014.

[12] Shital K Mishra, Sourav S Bhowmick, Huey Eng Chua, Fan Zhang, and Jie Zheng. Computational cell fate modelling for discovery of rewiring in apoptotic network for enhanced cancer drug sensitivity. *BMC systems biology*, 9(1):1, 2015.

[13] David Wan-Cheng Li, Jin-Ping Liu, Ying-Wei Mao, Hua Xiang, Juan Wang, Wei-Ya Ma, Zigang Dong, Helen M Pike, Rhoderick E Brown, and John C Reed. Calcium-activated raf/mek/erk signaling pathway mediates p53-dependent apoptosis and is abrogated by αb-crystallin through inhibition of ras activation. *Molecular biology of the cell*, 16(9):4437–4453, 2005.

[14] Juliane Liepe, Paul Kirk, Sarah Filippi, Tina Toni, Chris P Barnes, and Michael PH Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate bayesian computation. *Nature protocols*, 9(2):439–456, 2014.

[15] Thomas W Kensler, Avrum Spira, Judy E Garber, Eva Szabo, J Jack Lee, Zigang Dong, Andrew J Dannenberg, William N Hait, Elizabeth Blackburn, Nancy E Davidson, et al. Transforming cancer prevention through precision medicine and immune-oncology. *Cancer Prevention Research*, 9(1):2–10, 2016.

[16] Sean Ekins, Yuri Nikolsky, Andrej Bugrim, Eugene Kirillov, and Tatiana Nikolskaya. Pathway mapping tools for analysis of high content data. *High content screening: A powerful approach to systems cell biology and drug discovery*, pages 319–350, 2006.

[17] Akira Funahashi, Yukiko Matsuoka, Akiya Jouraku, Mineo Morohashi, Norihiro Kikuchi, and Hiroaki Kitano. Celldesigner 3.5: a versatile modeling tool for biochemical networks. *Proceedings of the IEEE*, 96(8):1254–1265, 2008.

[18] Jeremy M Berg, John L Tymoczko, and Lubert Stryer. Biochemistry. 5th, 2002.

[19] M Peifer and J Timmer. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting.

Systems Biology, IET, 1(2):78–88, 2007.

[20] Jim O Ramsay, G Hooker, D Campbell, and J Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007.

[21] Hua Liang and Hulin Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 2012.

[22] Vladislav Vyshemirsky and Mark Girolami. Biobayes: a software package for bayesian inference in systems biology. *Bioinformatics*, 24(17):1933–1934, 2008.

[23] Richard M Neve, Koei Chin, Jane Fridlyand, Jennifer Yeh, Frederick L Baehner, Tea Fevr, Laura Clark, Nora Bayani, Jean-Philippe Coppe, Frances Tong, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer cell*, 10(6):515–527, 2006.

[24] Deborah L Holliday and Valerie Speirs. Choosing the right cell line for breast cancer research. *Breast Cancer Res*, 13(4):215, 2011.

[25] Tongli Zhang, Paul Brazhnik, and John J Tyson. Computational analysis of dynamical responses to the intrinsic pathway of programmed cell death. *Biophysical journal*, 97(2):415–434, 2009.

[26] Gabriel Sollberger, Gerhard E Strittmatter, Martha Garstkiewicz, Jennifer Sand, and Hans-Dietmar Beer. Caspase-1: the inflammasome and beyond. *Innate immunity*, 20(2):115–125, 2014.

[27] Ning Kang, Jing-Hai Zhang, Feng Qiu, Shin-ichi Tashiro, Satoshi Onodera, and Takashi Ikejima. Inhibition of egfr signaling augments oridonin-induced apoptosis in human laryngeal cancer cells via enhancing oxidative stress coincident with activation of both the intrinsic and extrinsic apoptotic pathways. *Cancer letters*, 294(2):147–158, 2010.

[28] Floriana Morgillo, Elena D'Aiuto, Teresa Troiani, Erika Martinelli, Tina Cascone, Raffaele De Palma, Michele Orditura, Ferdinando De Vita, and Fortunato Ciardiello. Antitumor activity of bortezomib in human cancer cells with acquired resistance to anti-epidermal growth factor receptor tyrosine kinase inhibitors. *Lung Cancer*, 71(3):283–290, 2011.

[29] Yanyuan Wu, Monica Alvarez, Dennis J Slamon, Phillip Koeffler, and Jaydutt V Vadgama. Caspase 8 and maspin are downregulated in breast cancer cells due to cpg site promoter methylation. *BMC cancer*, 10(1):32, 2010.

[30] Ye Liang, Chaohua Yan, Nina F Schor, et al. Apoptosis in the absence of caspase 3. *Oncogene*, 20(45):6570–6578, 2001.

[31] Alexander Mitsos, Ioannis N Melas, Paraskeuas Siminelakis, Aikaterini D Chairakaki, Julio Saez-Rodriguez, and Leonidas G Alexopoulos. Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data. *PLoS Comput Biol*, 5(12):e1000591, 2009.

[32] Toshiki Iwai, Yoichiro Moriya, Masatoshi Shirane, Kaori Fujimoto-Ouchi, and Kazushige Mori. Continuous inhibition of epidermal growth factor receptor phosphorylation by erlotinib enhances antitumor activity of chemotherapy in erlotinib-resistant tumor xenografts. *Oncology reports*, 27(4):923–928, 2012.

[33] Ryungsa Kim. Recent advances in understanding the cell death pathways activated by anticancer therapy. *Cancer*, 103(8):1551–1560, 2005.