Mechanistic modeling of cell viability assays with *in silico* lineage tracing

Arnab Mutsuddy[1], Jonah R. Huggins[1,*], Aurore Amrit[1,2,*], Cemal Erdem[1,3], Jon C. Calhoun[4], and Marc R. Birtwistle[1,5,#]

[1]Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, USA

[2]Faculté de Pharmacie, Université Paris Cité, Paris, France

[3]Department of Medical Biosciences, Umeå University, Umeå, Sweden

[4]Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, USA

[5]Department of Bioengineering, Clemson University, Clemson, SC, USA

[#]Correspondence to: mbirtwi@clemson.edu

*Equal contribution

Abstract

Data from cell viability assays, which measure cumulative division and death events in a population and reflect substantial cellular heterogeneity, are widely available. However, interpreting such data with mechanistic computational models is hindered because direct model/data comparison is often muddled. We developed an algorithm that tracks simulated division and death events in mechanistically detailed single-cell lineages to enable such a model/data comparison and suggest causes of cell-cell drug response variability. Using our previously developed model of mammalian single-cell proliferation and death signaling, we simulated drug dose response experiments for four targeted anti-cancer drugs (alpelisib, neratinib, trametinib and palbociclib) and compared them to experimental data. Simulations are consistent with data for strong growth inhibition by trametinib (MEK inhibitor) and overall lack of efficacy for alpelisib (PI-3K inhibitor), but are inconsistent with data for palbociclib (CDK4/6 inhibitor) and neratinib (EGFR inhibitor). Model/data inconsistencies suggest (i) the importance of CDK4/6 for driving the cell cycle may be overestimated, and (ii) that the cellular balance between basal (tonic) and ligand-induced signaling is a critical determinant of receptor inhibitor response. Simulations show subpopulations of rapidly and slowly dividing cells in both control and drug-treated conditions. Variations in mother cells prior to drug treatment all impinging on ERK pathway activity are associated with the rapidly dividing phenotype and trametinib resistance. This work lays a foundation for the application of mechanistic modeling to large-scale cell viability assay datasets and better understanding determinants of cellular heterogeneity in drug response.

Introduction

One of the grand challenges of systems biology is to build a comprehensive and quantitative understanding of the structure and functionality of living cells. Whole cell models that describe the function of every gene and its products is an attractive manifestation of such a goal. Whole-cell models for some microorganisms exist[1–4], but have not been reported for human cells. Nevertheless, a wide range of individual pathway[5–20] and integrative multiple pathway[21–28] human cell models have been published, which are a stepping stone[29,30]. These models can contribute to better understanding multiscale phenotypes[31–35], diagnosis of disease states and their progression[36,37], and development of efficacious therapeutic procedures[38–40].

Data availability is a primary challenge for human whole cell modeling[29,41,42]. Large-scale databases[43–45] containing cell viability assay data exploring sensitivity of multiple cell lines to multiple drugs are an attractive resource in this regard. Assessing how well computational models based on current knowledge explain this data can reveal existing knowledge gaps and inform next stages of research.

An obvious pre-requisite to leveraging such data sets for modeling purposes is the existence of robust methods for comparing simulations to experimental readouts appropriately. Most viability assays measure cell population size or a proxy. Typical mechanistic models do not explicitly describe cell division and death events, whose balance dictates cell population size, but usually prescribe empirical relations[22,46]. Course-grained agent based modeling[47–49] can account for division and death events and were described in, for example, process control and optimization of production of therapeutic proteins using mammalian cell culture[50], analyzing the impact of cell

population heterogeneity in colony and tissue context[51], and elucidating the role of heterogeneity in IFNβ signaling[52]. However, there remains a need for algorithms that can simultaneously track large-scale mechanistic detail of drug response with cell division/death events. Such algorithms would help interpret the large body of cell viability response data for building better human whole-cell models.

In this work, we present an algorithm that combines detailed mechanistic descriptions of drug action with single cell lineage-resolved division and death events to construct simulation outputs that are directly comparable to cell viability assay data. As a test case, we use a previously developed large-scale model of single mammalian breast epithelial cell (MCF10A) proliferation and death, SPARCED[53], and add to it mechanistic pharmacodynamic models based on known binding interactions between drugs and modeled targets. First, we describe the algorithm and the types of novel analytics that can be derived, such as cell population size dynamics, cross-generational biomarker tracking for cell lineages, and cell population dendrograms. Then, we simulate dose responses to multiple drugs for which experimental data exist, namely, alpelisib (PI-3K inhibitor), trametinib (MEK inhibitor), palbociclib (CDK4/6 inhibitor) and neratinib (EGFR inhibitor). Simulations agree with experiments for strong growth inhibition by trametinib and overall lack of efficacy of alpelisib; however, there is substantial model-experiment discrepancy for palbociclib and neratinib. Analysis of model discrepancies suggests that (i) the importance of CDK4/6 for driving the cell cycle is likely overestimated, and (ii) the cellular balance between basal (tonic) and ligand-induced ERK signaling is a critical determinant of EGFR inhibitor response. Simulations show subpopulations of rapidly and slowly dividing cells in both control and drug-treated conditions. We find that variations in

mother cells prior to drug treatment all impinging on ERK pathway activity are associated with the rapidly dividing phenotype and trametinib responsiveness. This work lays a foundation for the application of mechanistic modeling to large-scale cell viability datasets, which are critical to the development of whole human cell models, and also provide a unique analytical framework to generate hypotheses on molecular drivers of cellular variability across generations.

Results

Lineage-resolved single-cell simulation framework

We first set out to construct a simulation algorithm that mirrors drug dose response viability assays (**Fig. 1**). These assays typically start with a population of asynchronously cycling cells that are treated with drug for ~3 days and then assayed for final cell number (or a metric proportional to it). The final cell number is related to the number of division and death events each initial single cell ultimately experienced. Thus, the simulation algorithm starts with a population of asynchronously cycling cells and counts the individual division and death events from each initial cell. Throughout this manuscript, we use our previously published mechanistic model of single cell proliferation and death signaling[53] representing MCF10A breast epithelial cells (SPARCED). In principle, however, any model with single-cell resolution and division/death readouts is potentially compatible with the approach.

The algorithm begins with the creation of a simulated population of asynchronously-cycling, single cells (**Fig. 1A**). The initial model state is an average, serum-starved cell (non-cycling). The first step is to generate a population of cells with heterogeneous gene expression profiles, enabled by descriptions of intrinsic noise in gene expression as previously described[21]. We refer to this process as “heterogenization”. After 48 simulated hours, when the distribution of most protein levels across the cell population stabilizes, the addition of growth media is simulated (in the case of MCF10A cells and this model—EGF and Insulin). Subsequently, synchronized cell cycle progression is observed in simulations for an additional 48 hours, creating so-called “Generation 0” (Gen 0). To convert these Gen 0 synchronized cells into Gen 1

asynchronously cycling cells, we sample random times during the 48 hour growth media treatment window for each single cell. These selections become initial conditions for Gen 1, which is then subjected to simulated drug treatment for 72 hours (**Fig. 1B**).

Once these simulations are completed, the outputs are analyzed to determine cell division events (based on CyclinB-CDK1 peaks for this model) and their time points (**Fig. 1B**). Based on the cell division time points, the remaining simulation time (difference between division time and 72 hours), and initial conditions for each daughter cell are determined for the next generation, and lineage information is recorded. Subsequently, simulations for the next generation are run and this cycle continues until no division events occur in a given generation. Detected cell death events (based on cleaved PARP dynamics for this model), halt a lineage.

These simulations not only mirror typical drug dose response experiments but also enable lineage-resolved analyses (**Fig. 1C**). Since individual division and death events from a parental cell are tracked, it also allows dynamic tracking of observables (such as ERK or Akt activity) across multiple generations of any single cell lineage. Lineage dendrograms can be also constructed, as is typical in such analyses. Such capability may generate hypotheses linking drug sensitivity or resistance with cell fates and lineage, or variations in biochemistry that predispose cells to response or resistance.

Comparing Simulated Drug Dose Responses to Experimental Measurements

With the above algorithm, we could now compare model predictions to experimental data for cell viability drug responses. Specifically, we focused on four previously-studied, targeted anti-cancer drugs for which our model includes primary targets and significant off-targets: trametinib (MEK inhibitor), alpelisib (PI-3K inhibitor),

neratinib (EGFR inhibitor), and palbociclib (CDK4/6 inhibitor)[54]. We extended SPARCED by including known drug interactions with target species[55–58] and leveraging previously described capabilities to robustly and easily increase the model scope[53] (see Methods).

We then performed lineage-resolved simulations for various doses of the modeled drugs for which experimental data are available[54], with no adjustment to the SPARCED model (besides addition of the drug pharmacodynamic modules—see Methods), using a starting population of 100 cells. This framework allows direct simulation of the dynamic cell population size in response to drug doses (**Fig. 2A**). The effect of drug action can be visualized by cell lineage dendrograms, showing in this example the clear effect of moderate trametinib doses to reduce the number of cell division events (**Fig. 2B-C**). The simulation outputs were used to calculate growth-rate inhibition[59] metrics which is the same method applied to the experimental dataset, allowing direct comparison of experimental and simulation results (**Fig. 2D-G**). The simulation results for trametinib (**Fig. 2F**) demonstrate surprising agreement with experimental data considered no parameter fitting was done, and also expectedly indicate substantial cell-to-cell variability at low doses. The simulations also captured the overall lack of efficacy for alpelisib, albeit with some slight differences in dose response slope (**Fig. 2D**). On the other hand, predicted palbociclib (**Fig. 2G**) and neratinib (**Fig. 2E**) responses were substantially different from experiments. Subsequent analyses explore the nature of these differences, as well as potential reasons for the cell-to-cell variability in the trametinib response.

Palbociclib Dose Response Discrepancies Suggests CDK4/6 is Partially Redundant for Cell Cycle Progression

What could explain the experiment/simulation discrepancy for palbociclib, a potent inhibitor of CDK4/6, canonically understood to be a central mediator of cell cycle progression from G_0 and G_1 to S-phase[60–62] (**Fig. 2G**)? Simulated palbociclib dose response starts to deviate from the experimental results at doses as low as 0.01 μM. Above 0.1 μM, the simulated dose response shows complete cytostasis. On the other hand, experimental results show minimal growth inhibition at 0.1 μM, and even high doses indicate only partial growth inhibition.

One consideration was doubling time. We reasoned that if the experimental doubling time was slower (greater) than the simulated doubling time, then in simulations many more cell divisions would be inhibited by the drug. That may explain why the simulated effects of palbociclib were much greater than that observed in experiments. The used GR metric in principle should help to account for such doubling time-related phenomena[59], but it also relies on assumptions such as exponential growth and constant drug effect on growth, which may not be satisfied. The model predicted a slower doubling time (~48 hours) than was reported in experiments for these MCF10A cells (~18-25 hours) (**Fig. 3A** and[54]), although a wide range is reported for this cell line (~48[63] or even ~96 hrs[64]). This is opposite of the difference we expected may explain the dose response curve discrepancy. Therefore, we conclude doubling time differences are unlikely to explain the observed differences.

Another consideration was restriction point behavior with respect to CDK4/6 activity, where a cell continues to divide even after complete inhibition of CDK4/6 at some

point in the cell cycle[65–67]. We reasoned that if the model does not capture such behavior, simulated palbociclib treatment could immediately stop cell divisions instead of letting already committed cells continue to divide once, leading to more predicted potency than observed. To explore this in simulations, we analyzed the probability of cell division in Gen 1 versus Gen 2 cells, as a function of dynamic progress in the cell cycle at the time of simulated drug treatment (**Fig. 3B**). Prior studies place the restriction point early in the cell cycle[65]. Simulations with saturating palbociclib dose (0.1 μM) reflect such behavior, where most cells divide once if the cell cycle is at least ~10% completed, but subsequent cell divisions are nearly non-existent. This effect is also clear from the simulated lineage dendrogram which shows most cells divide once but not subsequently with this dose of palbociclib (**Fig. 3C**). Thus, we conclude that modeled restriction point behavior is also unlikely to explain discrepancies.

Finally, we considered that the canonically understood role of CDK4/6 as modeled in SPARCED is simply inadequate. That is, the assertion that CDK4/6 activity is a necessary and sufficient step to drive the early cell cycle[65,68,69]. A clinical line of evidence is the fact that CDK4/6 inhibitors have limited efficacy outside of hormone-positive breast cancers[60]. It has also been reported that proliferation can occur in CDK4/6 knockout cells[70]. More recent data have suggested that CDK4/6 activity has more of a probabilistic effect on cell cycle progression[71], and the restriction point may be more reversible than previously thought in response to CDK4/6 inhibition[72]. CDK activities may also be overlapping; for example CDK2 and CDK4/6 may be compensatory[73], and a sensor integrating multiple CDK activities[74] was shown to be highly predictive of restriction point behavior[75]. Therefore, we conclude that most likely, fundamental model reformulation is

needed to capture the effects of palbociclib, and that the canonical view of CDK4/6 as necessary and sufficient for cell cycle progression is inadequate.

The Balance of Tonic Versus Ligand-Induced Growth Factor Signaling is Critical for Capturing Drug Effects

Neratinib is an irreversible inhibitor of the EGFR (with some off-target activity for the closely related ErbB2/HER2), a receptor tyrosine kinase that, upon ligand binding, activates the pro-proliferative and survival ERK and AKT pathways[76–78]. Hence, drug action is expected to block ERK and AKT signaling when a ligand, such as EGF, binds to EGFR. The experimental dose response (**Fig. 2E**) shows strong growth inhibition at doses above 0.1 μM and complete cytostasis at ~3 μM. However, simulation-predicted growth inhibition within this range is significantly weaker.

To explain this discrepancy, we considered that the current modeled balance of ligand-induced versus basal (also called tonic) ERK signaling could be incorrect. Specifically, that basal ERK signaling was too strong and causes non-negligible proliferation in the absence of EGF. If cell cycling is initiated by basal signaling too strongly, coupled with the fact that neratinib cannot inhibit basal signaling, this could explain some of the model-experiment discrepancy.

MCF10A cells are dependent upon EGF for cell cycle progression[79,80]. Thus, in simulations, cells dividing without EGF would support the above explanation. In simulations where the growth media contained only insulin, some cell division events were observed (**Fig. 4A**). Since the proliferative signaling activity that caused these divisions did not originate as a result of simulated EGF-EGFR activity, simulated neratinib treatment cannot inhibit these. This is inconsistent with the experimentally observed cell

behavior and hence may be a major cause of mismatch between simulation and experiment.

How could the model be changed to account for these mismatches? First, we ensured that basal ERK signaling in the presence of insulin minimally induces cell cycle progression. Basal Ras-GDP to Ras-GTP exchange is the main reaction controlling basal ERK activity in the model. We reduced the value of the associated rate constant until the probability of cell division in the absence of EGF and presence of insulin was near zero (**Fig. 4B**—last point on left, $2x10^{-4}$ s^{-1}), and then simulated the dose responses again (**Fig. 4C-F**). The new simulated neratinib dose responses show closer alignment with experiments. However, for all other drugs, experiment-model agreement became significantly worse, most likely now because the absolute levels of EGF-induced ERK signaling are altered. This result reinforces the close interacting nature of signaling mechanisms in the model for influencing broad features of drug response, and cautions against developing models without considering comparison to a compendium of data. Further model refinement in this regard, therefore, will be the scope of future work.

Explaining Single-Cell Heterogeneity in Division Rate and Trametinib Response

A commonly observed phenotypic variation among cells within a population is the division rate[81,82]. Under both control (no drug) and trametinib (~half-maximal response dose = 0.03 nM) treatment conditions, simulations show large variability in the number of divisions arising from a particular Gen 1 mother cell (**Figs. 5A-B**). Since rapidly dividing simulated cells (indicated by red) are present prior to drug treatment and persist after drug treatment, in simulations they are largely responsible for the partial response to trametinib. Could properties in the initial state of Gen 1 mother cells at the time of

trametinib treatment be a predictor of resistance, in this case marked by persistent rapid division in the presence of drug? Do the same mechanisms that drive rapid division in control conditions apply to this drug resistance?

To answer these questions, we first focused on simulated control cells. The initial conditions of Gen 1 cells under control conditions were extracted into a cell-by-species matrix (5,000 simulated cells by 934 initial conditions) (**Fig. 5C**). To identify species that may be associated with rapid division, we performed principal components analysis of this matrix, and colored cells by their division phenotype (**Fig. 5D**). The second principal component (PC2) stratified Gen 1 mother cells based on the number of divisions. To identify the most important model species contributing to PC2, we analyzed the loadings (**Fig. 5E**). Species with large loadings were associated with ERK signaling pathway components or downstream early cell cycle components. This suggests a simple hypothesis—that any fluctuation giving rise to higher ERK signaling capacity is associated with the rapid division phenotype under control conditions.

Does this finding hold true under trametinib treatment conditions? That is, are cells with higher ERK signaling capacity more likely to retain rapid division phenotypes in the presence of sub-saturating doses of trametinib? Trametinib is a MEK inhibitor, and since MEK is a key component of the ERK signaling pathway, there is a clear connection. Trametinib treatment shifts the number of divisions distribution to the left, reducing the number of rapidly dividing cells (**Fig. 5B**). To answer the question, we generated a new initial Gen 1 mother cell state matrix from trametinib-treated simulated cells, and then applied the projection learned from PCA of the control data onto this matrix. The results of this projection indicate again that rapidly-dividing cells cluster towards higher values

on PC2 (**Fig. 5F**). We conclude that in simulations, the rapidly dividing phenotype is driven by a multitude of factors impinging on higher ERK signaling pathway capacity, and these cells are likely to remain rapidly dividing in the presence of trametinib, contributing to acute resistance.

Discussion

Data availability is a major bottleneck for systems biology model development. While there is a wide range of drug dose response viability assay data available, it is difficult to use for large-scale model development because simulation outputs often do not recapitulate the experiment outputs—cell number from cumulative division and death events in single cells. To address this gap, we developed a lineage-resolved simulation framework that tracks individual cell division and death events along with mechanistic detail that enables inference for why single cells have different outcomes. We demonstrate application of this framework using our previously developed model of proliferation and death signaling in single mammalian cells[53], but in principle any model that simulates division and death events should be compatible. We compare model simulations to experimental data for viability response to four different targeted anti-cancer drugs. Discrepancies between model and data for palbociclib and neratinib, elaborated on further below, lead to new biological understanding. Deeper analysis of trametinib cases suggest mechanisms of resistance and what drives rapidly cycling cells in general.

For palbociclib, the simulations overpredicted its efficacy, showing very high growth inhibition at moderate doses to complete cytostasis at high doses. This reflects the indispensability of the drug target, CDK4/6, as per the model of the cell cycle pathway. However, in experiments, even the highest doses resulted in only partial growth inhibition. CDK4/6 is associated with traversing the cell cycle restriction point[83]. In pre-S-phase cells, one of the regulators of restriction point, Rb, is bound to a key transcription factor of the cell cycle process, E2F. In presence of a growth stimulus, CDK4/6 is activated when

bound to cyclin D. This activated cyclin D-CDK4/6 complex can phosphorylate and inactivate Rb, which then releases E2F. Subsequently, there is an upregulation of E2F which then mediates S-phase entry and progression by activating cyclin E and cyclin A. The mechanism of CDK4/6 inhibitors such as palbociclib attempt to induce cytostasis by preventing the inactivation of Rb by CDK4/6[84]. This canonical understanding places CDK4/6 as indispensable, similar to how it is modeled, but experiments did not agree with this. One of the known resistance mechanisms of CDK4/6 inhibition is the loss of Rb function[85,86]. However, since MCF10A cells do not harbor such mutations, it is an unlikely explanation in this case. Another reported resistance mechanism in cancer cells is the overexpression of Cyclin E[87,88], which is a regulator of the later stages of cell cycle, but is also not the case in MCF10A cells. In the results section, we investigated mismatch between model and experimental doubling time and restriction point behavior, finding neither likely to explain the discrepancies. Therefore, we think the most likely explanation is that CDK4/6 is simply not as indispensable for the cell cycle as contemporary views may portray. It has been increasingly reported that CDKs can compensate for one another[73], so the activity of other CDKs could compensate for CDK4/6 activity in actively cycling cells. Such mechanisms were not included in the original cell cycle submodel[89], so these additions are likely important for capturing effects of cell cycle-targeted therapies.

For neratinib, the simulations underpredicted its efficacy, showing weak inhibition for moderate to high doses whereas the experiments showed significant growth inhibition to complete cytostasis within this range. To investigate this discrepancy, we considered the progression of ERK proliferative signaling within the single cell model and how the

neratinib drug action might affect it. Neratinib is an irreversible inhibitor of the EGFR, which attempts to block ERK and Akt proliferative signaling by inhibiting ligand-receptor interactions. The model incorporates both ligand-induced and basal signaling along the ERK pathway. In simulations, if cells enter the cell cycle in the absence of ligand, it would result in proliferation that the drug action would be unable to inhibit. To test this, we performed subsequent simulations where EGF was absent from the growth media, but several simulated cells were still cycling. This is contrary to the experimental observations that MCF10A cells do not proliferate without EGF[79,80], and explains the discrepancy observed between simulation and experimental results of neratinib dose response. Furthermore, we sought to account for this mismatch by altering a key model reaction modulating basal ERK activity, basal Ras-GDP to Ras-GTP exchange rate. We reduced this rate constant to minimize the probability of cell division in absence of EGF and ran all dose response simulations again. This time, the neratinib dose response showed closer alignment with the experimental result, but we observed overprediction of growth inhibition for all other drugs, presumably due to the altered balance between basal and ligand induced ERK signaling. Hence, ideally, the model should incorporate a more improved balance between basal and ligand induced signaling for describing cell proliferation events. In our previous work, stochastic single cell simulations initiated from a representation of a serum-starved MCF10A cell minimally entered S-phase without EGF[21]. However, for cell population simulations done here, single cells are subject to randomized sampling for induction of an asynchronously cycling population which more closely resembles the experimental conditions whereby drug treatment is applied after growth media is introduced to the cells. Also, cells were followed for much longer, which

amplifies small percentages of cells still cycling with Insulin treatment alone. Thus, the model's limitations become more apparent here at the population level.

The neratinib case study highlighted an important future direction focused on parameter estimation for such models with stochastic components. This is a challenging area due to the computational cost of model evaluation, and the wide range of datasets that are needed to constrain large stochastic models. One part of our previous work was to do this for a subset of rate constants by "initialization"[21,90]. In initialization, certain model parameters and initial conditions are determined for a specific cell-line context using a set of focused parameter estimation operations which aim to tune parameters based on constraints placed on model observables. It is a computationally intensive process whereby each parameter estimation step performs iterative execution of deterministic model simulations. The SPARCED model is composed of a stochastic gene expression and a protein biochemistry module which are executed simultaneously. However, communication bottlenecks between the modules caused the computation time to be impractical for the purpose of initialization[90]. Recently, we solved the communication bottleneck problem which speeded up the deterministic execution by over 200-fold[90]. Fast deterministic parameter estimation solvers have been reported for large-scale models as well[22]. This drastic increase in computation speed for deterministic simulations will allow a more exhaustive exploration of the model parameters essential for defining a more robust initialization protocol, but extending this to stochastic evaluation remains an important unsolved problem.

For trametinib, model predictions closely resemble experimental observations. Trametinib has high specificity for MEK1/2; once MEK1/2 is inhibited, it is no longer able

to phosphorylate ERK1/2[91]. ERK signaling controls the G1/S-phase transition of the cell cycle[91] through activation of RSK, which in turn upregulates the production of Cyclin D and CDK4/6. Cyclin D expression through the G1 phase can drive the cell through the G1/S-phase checkpoint, as described above. When these events are inhibited by trametinib, the cell is unable to progress through the G1/S-phase checkpoint. However, results indicate that low-medium doses of trametinib are unable to reduce cycling in every cell in the population. Spencer et al. hypothesize that the signaling response patterns of a mother cell pass on to daughter cells, enabling an expression pattern to continue through multiple generations[92]. Therefore, we hypothesized that the initial values of the Generation 1 mother cells could predict the number of divisions that occur from said mother cell. The number of division events was able to be explained by principal component analysis, with higher ERK signaling capacity in multiple regards was associated with an increased number of division events. Thus, in this case acute resistance to trametinib is simply related to a multitude of biochemical factors all impinging on increasing the activity of the target ERK pathway.

In conclusion, we have developed an algorithm that takes a mechanistically-detailed model of stochastic proliferation and death, and generates lineage-resolved simulations that can be used to interpret dose response viability data and better understand cellular response heterogeneity. Specific demonstrations revealed new insights into drug response, cell cycle biology, rapidly dividing phenotypes, and acute drug resistance. Given the abundance of such drug dose response viability data that is available, we expect this work to help address the modeling bottleneck in data availability, for building mechanistic models of single cell behavior.

Methods

Code Availability

The final model scripts, files, and information are available on the SPARCED GitHub page at github.com/SPARCED/LinResSims.

SPARCED Pharmacodynamic Models

Reactions representing drugs binding to their reported targets with mass action rate laws were added to the SPARCED model (see model input text files). The assumptions and mechanism of action for each drug are described below. We tested each drug action model by observing simulated deterministic response of an average serum-starved cell to EGF and Insulin (growth media doses) with and without drug (10 μM). We required (i) intracellular and extracellular free drug concentration equilibrated rapidly (within a few minutes); (ii) drug-target engagement (complex formation) was observed similarly rapidly; (iii) that the drug had a substantial effect on a downstream biomarker (ppAKT, ppERK, pEGFR, or Cyclin D).

Alpelisib. Alpelisib enters and leaves the cell with first-order kinetics and the same rate constant (0.01 $nM^{-1}s^{-1}$). Cytoplasmic alpelisib binds reversibly to its intracellular targets, all p110 isoforms and free PI3K (p85/p110 heterodimers), with a dissociation constant (K_d) of 2.4 nM[55] and mass action kinetics (k_{on} = 0.001 $nM^{-1}s^{-1}$; k_{off} = 0.0024 s^{-1}). The binding of alpelisib to p110 prevents its dimerization with the regulatory subunit (p85). Any drug-bound species loses its kinase activity. Any drug-bound species undergoes first-order decay with a rate constant equal to that of the non-drug-bound species.

Palbociclib. Palbociclib enters and leaves the cell and the nucleus with first-order kinetics and the same rate constant (0.01 $nM^{-1}s^{-1}$). Nuclear palbociclib reversibly binds to its target, nuclear CDK4/6, with a dissociation constant (K_d) of 1.9 nM and mass action kinetics (k_{on} = 0.001 $nM^{-1}s^{-1}$; k_{off} = 0.0019 s^{-1}). Any drug-bound species loses its kinase activity. Any drug-bound species undergoes first-order decay with a rate constant equal to that of the non-drug-bound species.

Trametinib. Trametinib enters and leaves the cell with first-order kinetics and the same rate constant (0.01 $nM^{-1}s^{-1}$). Cytoplasmic trametinib reversibly binds to its target, unphosphorylated free MEK, with a dissociation constant (K_d) of 0.35 nM and mass action kinetics (k_{on} = 0.001 $nM^{-1}s^{-1}$; k_{off} = 0.00035 s^{-1}). Any drug-bound species loses its kinase activity and ability to bind substrates. This is a simplification of trametinib action but is effective for capturing the broad effects of downregulating the ERK pathway. Any drug-bound species undergoes first-order decay with a rate constant equal to that of the non-drug-bound species.

Neratinib. Neratinib enters and leaves the cell with first-order kinetics and the same rate constant (0.01 $nM^{-1}s^{-1}$). Cytoplasmic neratinib binds irreversibly to free EGFR, ErbB2, and ErbB4 with first-order kinetics (k_{on} = 10^{-4} $nM^{-1}s^{-1}$). While this is a kinase inhibitor, and drug bound complex loses kinase activity, for simplicity we disallow subsequent interaction with other receptors and ligands. Any drug-bound species undergoes first-order decay with a rate constant equal to that of the non-drug-bound species.

Lineage-Resolved Simulations

Asynchronous population. Cell population simulations are initiated by creating a representation of an asynchronously cycling cell population. The starting size of the cell

population is specified by the user. For each starting cell, initial conditions representing an average serum-starved MCF10A cell are used to create a heterogenized cell population[21]. For heterogenization, we run stochastic single cell simulations for 48 simulated hours under serum-starved conditions, using the initial conditions of the average serum-starved MCF10A cell. Thus, the intrinsic gene expression noise incorporated within the single cell model leads to heterogeneity in protein levels across the starting cell population over the duration of simulation time. Then, simulated growth media with EGF (3.3 nM) and insulin (1721 nM) is introduced and another series of stochastic simulations are run for each individual cell for 48 hours. From the generated trajectories, for each cell a timepoint is randomly selected from a uniform distribution using the NumPy `randint` function. The conditions at this time point for each cell are used as the initial conditions for the first generation. Single-cell simulations are executed for all first generation cells for the user-specified duration (typically 72 hours).

Identifying cell division events. Once the single-cell simulations are completed, the generated outputs are analyzed to determine cell division events. The cell division events are detected by identifying troughs in Cyclin B-CDK1 trajectories. For this, we defined a python function combining the `find_peaks` methods in the SciPy signal processing library and the n-th discrete difference calculation method (along any given axis) in the NumPy library. For any individual cell, if a division event is detected, timepoints after the occurrence of cell division events are discarded and the state vector at the time of cell division is selected as the initial condition for two new second generation cells.

Identifying cell death events. Cleaved PARP is the readout for cell death[21]. For any single cell, if more than half of PARP has been cleaved at any time point, the cell is labeled dead at that time point.

Subsequent generations. For each generation, state matrices for individual cells are obtained and saved as part of output dataset. In the event of a cell division, we retain the state matrix of the mother cell until the time point of division and the remaining portion is truncated and discarded. For every cell, we scan the output for the duration of its lifetime to find division events. To determine the required simulation time for next generation of daughter cells, the division time point is subtracted from the total simulation time. The single cell outputs at the time point of division of each mother cell is recorded as initial conditions for the next generation of daughter cells. Thus, we define the required simulation time, population, and initial conditions for the next generation of cells. This process is repeated for the subsequent generations of cell populations. In a given generation, if there is no cell division event observed within the simulation time, the population simulation is terminated.

Implementation. Computation is performed using HPC-compatible parallel processing in Python whereby single cell simulations are run in individual CPU threads. To run the cell population simulation, a computational environment with an implementation of MPI (Message Passing Interface), such as OpenMPI[93] on Linux and MSMPI on Windows systems needs to be set up in addition to the dependencies of the SPARCED model pipeline. Before the simulation can be performed, the SPARCED model is built using the python script under scripts/createModel.py, which creates an executable single cell model based on the specifications in the input files. Once the model build

process is complete, MPI can be used to run cell population simulations using the below command. Upon completion of simulations, the results are saved to disk as Python pickle objects for analysis and visualization.

```
mpirun -np [n_cpu] python cellpop_drs.py --arguments [argument_value]
```

Here, `n_cpu` is the number of MPI processes that the user decides to use for parallelization. The arguments are:

`sim_name`: User-defined string to create a directory under sparced/output where simulation outputs will be saved.

`cellpop`: An integer specifying the number of starting cells for simulation

`exp_time`: Duration of experiment in hours

`drug`: String specifying species name for the drug of interest, available options for our work include: `alpel_EC, nerat_EC, trame_EC` and `palbo_EC`

`dose`: Applied concentration of the drug in μM

`egf`: EGF concentration in nM

`ins`: insulin concentration in nM

`hgf`: HGF concentration in nM

`nrg`: heregulin concentration in nM

`pdgf`: PDGF concentration in nM

`igf`: IGF concentration in nM

`fgf`: FGF concentration in nM

GR Score Calculation

Dose responses from cell population simulations were calculated using the growth rate inhibition metric (GR)[59]. Dose response simulations were run for 10 dose-levels matching experimental data for each drug and 10 replicates of each dose. Outputs from the cell population simulations were read and analyzed to determine the total number of living cells over time for the duration of experiment time. Using the number of living cells at 72 hours, the GR scores were computed for each replicate was computed with the Python script provided as part of GR-metrics git repository.

Calculation of Fractional Cell Cycle Progression

Cell cycle progress estimation. For the palbociclib dose response, the extent of cell cycle progress at the time of drug addition was estimated using a function of average cyclin concentration levels. CyclinE-CDK2, CyclinA-CDK2 and CyclinB-CDK1 species concentrations were converted to a relative measure based on their observed peaks. An average of these three variables over time generated an oscillating function, with trough-to-trough distance representing total cell cycle time. We determined an average trajectory of this function using a deterministic simulation. Then we calculated the function trajectory from individual Gen 1 cells We calculated the relative cell cycle progression as the ratio of temporal distance between timepoint of alignment and its previous trough to distance between two neighboring troughs.

Estimation of cell division probability given cell cycle progression. For any given drug dose, the cell cycle progression of all cells at the time of dose administration was calculated. Then all living cells were grouped into those that divided and did not divide in both Gen 1 and Gen 2. Gaussian kernel density estimation was used to estimate the probability density function for each group. Using the probability density function, the number of cells for dividing and non-dividing groups within cell cycle progress time intervals with increments of 0.01 were estimated. For each interval, the probability of division at Gen 1 and 2 were calculated using the ratio of number of dividing cells and total number of cells.

Principal Components Analysis

The number of progeny arising from each Gen 1 cell ('mother cell') was determined from control condition simulations as above. Z-score normalization was applied to the initial condition matrix (cells-by-species) using `standardScaler.transform` in scikit-learn. Principal component analysis was completed using `decompostion.PCA` in scikit-learn. To apply this projection to the drug-treated simulated cells, we generated a new initial state matrix from simulated mother cells treated with 0.032 nM trametinib and normalized as above.

References

1. Karr JR, Sanghvi JC, Macklin DN, et al. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*. 2012;150(2):389-401. doi:10.1016/j.cell.2012.05.044

2. Ahn-Horst TA, Mille LS, Sun G, Morrison JH, Covert MW. An expanded whole-cell model of E. coli links cellular physiology with mechanisms of growth rate control. *npj Syst Biol Appl*. 2022;8(1):1-21. doi:10.1038/s41540-022-00242-9

3. Macklin DN, Ahn-Horst TA, Choi H, et al. Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. *Science*. 2020;369(6502):eaav3751. doi:10.1126/science.aav3751

4. Thornburg ZR, Bianchi DM, Brier TA, et al. Fundamental behaviors emerge from simulations of a living minimal cell. *Cell*. 2022;185(2):345-360.e28. doi:10.1016/j.cell.2021.12.025

5. Batchelor E, Loewer A, Mock C, Lahav G. Stimulus-dependent dynamics of p53 in single cells. *Mol Syst Biol*. 2011;7:488. doi:10.1038/msb.2011.20

6. Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK. Modeling a Snap-Action, Variable-Delay Switch Controlling Extrinsic Cell Death. *PLOS Biology*. 2008;6(12):e299. doi:10.1371/journal.pbio.0060299

7. Nakakuki T, Birtwistle MR, Saeki Y, et al. Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics. *Cell*. 2010;141(5):884-896. doi:10.1016/j.cell.2010.03.054

8. Hughey JJ, Lee TK, Covert MW. Computational modeling of mammalian signaling networks. *WIREs Systems Biology and Medicine*. 2010;2(2):194-209. doi:10.1002/wsbm.52

9. Singhania R, Sramkoski RM, Jacobberger JW, Tyson JJ. A Hybrid Model of Mammalian Cell Cycle Regulation. *PLOS Computational Biology*. 2011;7(2):e1001077. doi:10.1371/journal.pcbi.1001077

10. Herz AVM, Gollisch T, Machens CK, Jaeger D. Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction. *Science*. 2006;314(5796):80-85. doi:10.1126/science.1127240

11. Bachmann J, Raue A, Schilling M, et al. Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Mol Syst Biol*. 2011;7:516. doi:10.1038/msb.2011.50

12. Schoeberl B, Pace EA, Fitzgerald JB, et al. Therapeutically Targeting ErbB3: A Key Node in Ligand-Induced Activation of the ErbB Receptor–PI3K Axis. *Science Signaling*. 2009;2(77):ra31-ra31. doi:10.1126/scisignal.2000352

13. Eduati F, Doldàn-Martelli V, Klinger B, et al. Drug Resistance Mechanisms in Colorectal Cancer Dissected with Cell Type-Specific Dynamic Logic Models. *Cancer Res*. 2017;77(12):3364-3375. doi:10.1158/0008-5472.CAN-17-0078

14. Fey D, Halasz M, Dreidax D, et al. Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Science Signaling*. 2015;8(408):ra130-ra130. doi:10.1126/scisignal.aab0990

15. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res*. 2002;12(1):9-18. doi:10.1038/sj.cr.7290105

16. Chen WW, Schoeberl B, Jasper PJ, et al. Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular Systems Biology*. 2009;5(1):239. doi:10.1038/msb.2008.74

17. Altan-Bonnet G, Germain RN. Modeling T Cell Antigen Discrimination Based on Feedback Control of Digital ERK Responses. *PLoS Biol*. 2005;3(11):e356. doi:10.1371/journal.pbio.0030356

18. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science*. 2002;298(5596):1241-1245. doi:10.1126/science.1071914

19. Lee E, Salic A, Krüger R, Heinrich R, Kirschner MW. The Roles of APC and Axin Derived from Experimental and Theoretical Analysis of the Wnt Pathway. *PLoS Biol*. 2003;1(1):e10. doi:10.1371/journal.pbio.0000010

20. Park CS, Schneider IC, Haugh JM. Kinetic analysis of platelet-derived growth factor receptor/phosphoinositide 3-kinase/Akt signaling in fibroblasts. *J Biol Chem*. 2003;278(39):37064-37072. doi:10.1074/jbc.M304968200

21. Bouhaddou M, Barrette AM, Stern AD, et al. A mechanistic pan-cancer pathway model informed by multi-omics data interprets stochastic cell fate responses to drugs and mitogens. *PLOS Computational Biology*. 2018;14(3):e1005985. doi:10.1371/journal.pcbi.1005985

22. Fröhlich F, Kessler T, Weindl D, et al. Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model. *Cell Systems*. 2018;7(6):567-579.e6. doi:10.1016/j.cels.2018.10.013

23. Swainston N, Smallbone K, Hefzi H, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*. 2016;12:109. doi:10.1007/s11306-016-1051-4

24. Hass H, Masson K, Wohlgemuth S, et al. Predicting ligand-dependent tumors from multi-dimensional signaling features. *NPJ Syst Biol Appl*. 2017;3:27. doi:10.1038/s41540-017-0030-3

25. Nilsson A, Nielsen J. Genome scale metabolic modeling of cancer. *Metabolic Engineering*. 2017;43:103-112. doi:10.1016/j.ymben.2016.10.022

26. Mardinoglu A, Agren R, Kampf C, et al. Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Molecular Systems Biology*. 2013;9(1):649. doi:10.1038/msb.2013.5

27. Duarte NC, Becker SA, Jamshidi N, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*. 2007;104(6):1777-1782. doi:10.1073/pnas.0610772104

28. Robinson JL, Kocabaş P, Wang H, et al. An atlas of human metabolism. *Sci Signal*. 2020;13(624):eaaz1482. doi:10.1126/scisignal.aaz1482

29. Szigeti B, Roth YD, Sekar JAP, Goldberg AP, Pochiraju SC, Karr JR. A blueprint for human whole-cell modeling. *Curr Opin Syst Biol*. 2018;7:8-15. doi:10.1016/j.coisb.2017.10.005

30. Bajikar SS, Janes KA. Multiscale Models of Cell Signaling. *Ann Biomed Eng*. 2012;40(11):2319-2327. doi:10.1007/s10439-012-0560-1

31. Qutub AA, Popel AS. Elongation, proliferation & migration differentiate endothelial cell phenotypes and determine capillary sprouting. *BMC Syst Biol*. 2009;3:13. doi:10.1186/1752-0509-3-13

32. Ruscone M, Montagud A, Chavrier P, et al. Multiscale model of the different modes of cancer cell invasion. *Bioinformatics*. 2023;39(6):btad374. doi:10.1093/bioinformatics/btad374

33. Groß A, Kracher B, Kraus JM, et al. Representing dynamic biological networks with multi-scale probabilistic models. *Commun Biol*. 2019;2(1):1-12. doi:10.1038/s42003-018-0268-3

34. Wertheim KY, Puniya BL, Fleur AL, Shah AR, Barberis M, Helikar T. A multi-approach and multi-scale platform to model CD4+ T cells responding to infections. *PLOS Computational Biology*. 2021;17(8):e1009209. doi:10.1371/journal.pcbi.1009209

35. Aghamiri SS, Puniya BL, Amin R, Helikar T. A multiscale mechanistic model of human dendritic cells for in-silico investigation of immune responses and novel therapeutics discovery. *Frontiers in Immunology*. 2023;14. Accessed November 14, 2023. https://www.frontiersin.org/articles/10.3389/fimmu.2023.1112985

36. Tripathi S, Park JH, Pudakalakatti S, Bhattacharya PK, Kaipparettu BA, Levine H. A mechanistic modeling framework reveals the key principles underlying tumor metabolism. *PLOS Computational Biology*. 2022;18(2):e1009841. doi:10.1371/journal.pcbi.1009841

37. Hormuth DA, Jarrett AM, Lima EABF, McKenna MT, Fuentes DT, Yankeelov TE. Mechanism-Based Modeling of Tumor Growth and Treatment Response Constrained by Multiparametric Imaging Data. *JCO Clinical Cancer Informatics*. 2019;(3):1-10. doi:10.1200/CCI.18.00055

38. Azer K, Kaddi CD, Barrett JS, et al. History and Future Perspectives on the Discipline of Quantitative Systems Pharmacology Modeling and Its Applications. *Frontiers in Physiology*. 2021;12. Accessed November 14, 2023. https://www.frontiersin.org/articles/10.3389/fphys.2021.637999

39. Aghamiri SS, Amin R, Helikar T. Recent applications of quantitative systems pharmacology and machine learning models across diseases. *J Pharmacokinet Pharmacodyn*. 2022;49(1):19-37. doi:10.1007/s10928-021-09790-9

40. Chung D, Bakshi S, van der Graaf PH. A Review of Quantitative Systems Pharmacology Models of the Coagulation Cascade: Opportunities for Improved Usability. *Pharmaceutics*. 2023;15(3):918. doi:10.3390/pharmaceutics15030918

41. Gonçalves E, Bucher J, Ryll A, et al. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Molecular BioSystems*. 2013;9(7):1576-1583. doi:10.1039/C3MB25489E

42. Malik-Sheriff RS, Glont M, Nguyen TVN, et al. BioModels—15 years of sharing computational models in life science. *Nucleic Acids Research*. 2020;48(D1):D407-D415. doi:10.1093/nar/gkz1055

43. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603-607. doi:10.1038/nature11003

44. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*. 2013;41(Database issue):D955-D961. doi:10.1093/nar/gks1111

45. Rees MG, Seashore-Ludlow B, Cheah JH, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*. 2016;12(2):109-116. doi:10.1038/nchembio.1986

46. He W, Demas DM, Shajahan-Haq AN, Baumann WT. Modeling breast cancer proliferation, drug synergies, and alternating therapies. *iScience*. 2023;26(5):106714. doi:10.1016/j.isci.2023.106714

47. Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. Poisot T, ed. *PLoS Comput Biol*. 2018;14(2):e1005991. doi:10.1371/journal.pcbi.1005991

48. Islam MA, Versypt ANF. *Mathematical Modeling of Impacts of Patient Differences on COVID-19 Lung Fibrosis Outcomes*. Systems Biology; 2022. doi:10.1101/2022.11.06.515367

49. West J, Robertson-Tessi M, Anderson ARA. Agent-based methods facilitate integrative science in cancer. *Trends Cell Biol*. 2023;33(4):300-311. doi:10.1016/j.tcb.2022.10.006

50. Jackson R, Bayrak ES, Wang T, Coufal M, Undey C, Cinar A. High Performance Agent-Based Modeling to Simulate Mammalian Cell Culture Bioreactor. In: Eden MR, Ierapetritou MG, Towler GP, eds. *Computer Aided Chemical Engineering*. Vol 44. 13 International Symposium on Process Systems Engineering (PSE 2018). Elsevier; 2018:1453-1458. doi:10.1016/B978-0-444-64241-7.50237-8

51. Yu JS, Bagheri N. Agent-Based Models Predict Emergent Behavior of Heterogeneous Cell Populations in Dynamic Microenvironments. *Frontiers in Bioengineering and Biotechnology*. 2020;8. Accessed October 3, 2023. https://www.frontiersin.org/articles/10.3389/fbioe.2020.00249

52. Gregg RW, Shabnam F, Shoemaker JE. Agent-based modeling reveals benefits of heterogeneous and stochastic cell populations during cGAS-mediated IFNβ production. *Bioinformatics*. 2021;37(10):1428-1434. doi:10.1093/bioinformatics/btaa969

53. Erdem C, Mutsuddy A, Bensman EM, et al. A scalable, open-source implementation of a large-scale mechanistic model for single cell proliferation and death signaling. *Nat Commun*. 2022;13(1):3555. doi:10.1038/s41467-022-31138-1

54. Niepel M, Hafner M, Mills CE, et al. A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines. *Cell Systems*. 2019;9(1):35-48.e5. doi:10.1016/j.cels.2019.06.005

55. Fritsch C, Huang A, Chatenay-Rivauday C, et al. Characterization of the Novel and Specific PI3Kα Inhibitor NVP-BYL719 and Development of the Patient Stratification Strategy for Clinical Trials. *Molecular Cancer Therapeutics*. 2014;13(5):1117-1129. doi:10.1158/1535-7163.MCT-13-0865

56. Kim S, Tiedt R, Loo A, et al. The potent and selective cyclin-dependent kinases 4 and 6 inhibitor ribociclib (LEE011) is a versatile combination partner in preclinical cancer models. *Oncotarget*. 2018;9(81):35226-35240. doi:10.18632/oncotarget.26215

57. Yoshida T, Kakegawa J, Yamaguchi T, et al. Identification and Characterization of a Novel Chemotype MEK Inhibitor Able to Alter the Phosphorylation State of MEK1/2. *Oncotarget*. 2012;3(12):1533-1545.

58. PubChem. Neratinib. Accessed October 27, 2023. https://pubchem.ncbi.nlm.nih.gov/compound/9915743

59. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat Methods*. 2016;13(6):521-527. doi:10.1038/nmeth.3853

60. O'Leary B, Finn RS, Turner NC. Treating cancer with selective CDK4/6 inhibitors. *Nat Rev Clin Oncol*. 2016;13(7):417-430. doi:10.1038/nrclinonc.2016.26

61. Suski JM, Braun M, Strmiska V, Sicinski P. Targeting cell-cycle machinery in cancer. *Cancer Cell*. 2021;39(6):759-778. doi:10.1016/j.ccell.2021.03.010

62. Tchakarska G, Sola B. The double dealing of cyclin D1. *Cell Cycle*. 2020;19(2):163-178. doi:10.1080/15384101.2019.1706903

63. Gajewski E, Gaur S, Akman SA, Matsumoto L, van Balgooy JNA, Doroshow JH. Oxidative DNA Base Damage in MCF-10A Breast Epithelial Cells at Clinically Achievable Concentrations of Doxorubicin. *Biochem Pharmacol*. 2007;73(12):1947-1956. doi:10.1016/j.bcp.2007.03.022

64. Strauch M, Lüdke A, Münch D, et al. More than apples and oranges - Detecting cancer with a fruit fly's antenna. *Sci Rep*. 2014;4:3576. doi:10.1038/srep03576

65. Blagosklonny MV, Pardee AB. The Restriction Point of the Cell Cycle. In: *Madame Curie Bioscience Database [Internet]*. Landes Bioscience; 2013. Accessed June 6, 2024. https://www.ncbi.nlm.nih.gov/books/NBK6318/

66. Zetterberg A, Larsson O. Kinetic analysis of regulatory events in G1 leading to proliferation or quiescence of Swiss 3T3 cells. *Proc Natl Acad Sci U S A*. 1985;82(16):5365-5369. doi:10.1073/pnas.82.16.5365

67. Pardee AB. A restriction point for control of normal animal cell proliferation. *Proc Natl Acad Sci U S A*. 1974;71(4):1286-1290. doi:10.1073/pnas.71.4.1286

68. Alberts, B. Johnson, A. Lewis, J. Raff, M. Roberts, K. Walter P. *Molecular Biology of the Cell, 5th Edition*.; 2008.

69. Weinberg RA. *The Biology of Cancer*. GS Garland Science; 2007.

70. Malumbres M, Sotillo R, Santamaría D, et al. Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell*. 2004;118(4):493-504. doi:10.1016/j.cell.2004.08.002

71. Chung M, Liu C, Yang HW, Köberlin MS, Cappell SD, Meyer T. Transient Hysteresis in CDK4/6 Activity Underlies Passage of the Restriction Point in G1. *Mol Cell*. 2019;76(4):562-573.e4. doi:10.1016/j.molcel.2019.08.020

72. Cornwell JA, Crncec A, Afifi MM, Tang K, Amin R, Cappell SD. Loss of CDK4/6 activity in S/G2 phase leads to cell cycle reversal. *Nature*. 2023;619(7969):363-370. doi:10.1038/s41586-023-06274-3

73. Arora M, Moser J, Hoffman TE, et al. Rapid adaptation to CDK2 inhibition exposes intrinsic cell-cycle plasticity. *Cell*. 2023;186(12):2628-2643.e21. doi:10.1016/j.cell.2023.05.013

74. Hahn AT, Jones JT, Meyer T. Quantitative analysis of cell cycle phase durations and PC12 differentiation using fluorescent biosensors. *Cell Cycle*. 2009;8(7):1044-1052. doi:10.4161/cc.8.7.8042

75. Schwarz C, Johnson A, Kõivomägi M, et al. A Precise Cdk Activity Threshold Determines Passage through the Restriction Point. *Mol Cell*. 2018;69(2):253-264.e5. doi:10.1016/j.molcel.2017.12.017

76. Takeuchi K, Ito F. EGF receptor in relation to tumor development: molecular basis of responsiveness of cancer cells to EGFR-targeting tyrosine kinase inhibitors. *The FEBS Journal*. 2010;277(2):316-326. doi:10.1111/j.1742-4658.2009.07450.x

77. Seshacharyulu P, Ponnusamy MP, Haridas D, Jain M, Ganti AK, Batra SK. Targeting the EGFR signaling pathway in cancer therapy. *Expert Opinion on Therapeutic Targets*. 2012;16(1):15-31. doi:10.1517/14728222.2011.648617

78. Wee P, Wang Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers*. 2017;9(5):52. doi:10.3390/cancers9050052

79. Chou J, Fan Z, DeBlasio T, Koff A, Rosen N, Mendelsohn J. Constitutive overexpression of cyclin D1 in human breast epithelial cells does not prevent G1 arrest induced by deprivation of epidermal growth factor. *Breast Cancer Res Treat*. 1999;55(3):267-283. doi:10.1023/A:1006217413089

80. Scaling AL, Prossnitz ER, Hathaway HJ. GPER Mediates Estrogen-Induced Signaling and Proliferation in Human Breast Epithelial Cells and Normal and Malignant Breast. *HORM CANC*. 2014;5(3):146-160. doi:10.1007/s12672-014-0174-1

81. Min M, Spencer SL. Spontaneously slow-cycling subpopulations of human cells originate from activation of stress-response pathways. *PLoS Biol*. 2019;17(3):e3000178. doi:10.1371/journal.pbio.3000178

82. Nowak CM, Quarton T, Bleris L. Impact of variability in cell cycle periodicity on cell population dynamics. *PLOS Computational Biology*. 2023;19(6):e1011080. doi:10.1371/journal.pcbi.1011080

83. Li Y, Barbash O, Diehl JA. Regulation of the Cell Cycle. In: *The Molecular Basis of Cancer*. Elsevier; 2015:165-178.e2. doi:10.1016/B978-1-4557-4066-6.00011-1

84. Schmidt M, Sebastian M. Palbociclib—The First of a New Class of Cell Cycle Inhibitors. In: Martens UM, ed. *Small Molecules in Oncology*. Recent Results in Cancer Research. Springer International Publishing; 2018:153-175. doi:10.1007/978-3-319-91442-8_11

85. Lukas J, Parry D, Aagaard L, et al. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature*. 1995;375(6531):503-506. doi:10.1038/375503a0

86. Finn RS, Dering J, Conklin D, et al. PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Res*. 2009;11(5):R77. doi:10.1186/bcr2419

87. Knudsen KE, Weber E, Arden KC, Cavenee WK, Feramisco JR, Knudsen ES. The retinoblastoma tumor suppressor inhibits cellular proliferation through two distinct mechanisms: inhibition of cell cycle progression and induction of cell death. *Oncogene*. 1999;18(37):5239-5245. doi:10.1038/sj.onc.1202910

88. Caldon CE, Sergio CM, Kang J, et al. Cyclin E2 overexpression is associated with endocrine resistance but not insensitivity to CDK2 inhibition in human breast cancer cells. *Mol Cancer Ther*. 2012;11(7):1488-1499. doi:10.1158/1535-7163.MCT-11-0963

89. Gerard C, Goldbeter A. Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. *Proceedings of the National Academy of Sciences*. 2009;106(51):21643-21648. doi:10.1073/pnas.0903827106

90. Mutsuddy A, Erdem C, Huggins JR, et al. Computational speed-up of large-scale, single-cell model simulations via a fully integrated SBML-based format. *Bioinformatics Advances*. 2023;3(1):vbad039. doi:10.1093/bioadv/vbad039

91. Lavoie H, Gagnon J, Therrien M. ERK signalling: a master regulator of cell behaviour, life and fate. *Nat Rev Mol Cell Biol*. 2020;21(10):607-632. doi:10.1038/s41580-020-0255-7

92. Min M, Rong Y, Tian C, Spencer SL. Temporal integration of mitogen history in mother cells controls proliferation of daughter cells. *Science*. 2020;368(6496):1261-1265. doi:10.1126/science.aay8241

93. Gabriel E, Fagg GE, Bosilca G, et al. Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In: Kranzlmüller D, Kacsuk P, Dongarra J, eds. *Recent Advances in Parallel Virtual Machine and Message Passing Interface*. Lecture Notes in Computer Science. Springer; 2004:97-104. doi:10.1007/978-3-540-30218-6_19

Acknowledgements

We thank Michael Salim, Nicole Hobbs, and Daniel Cook for helpful discussions. This work was supported by NIH grant R35GM141891 to MRB. We thank the Clemson University Palmetto team for use of high-performance computing resources.

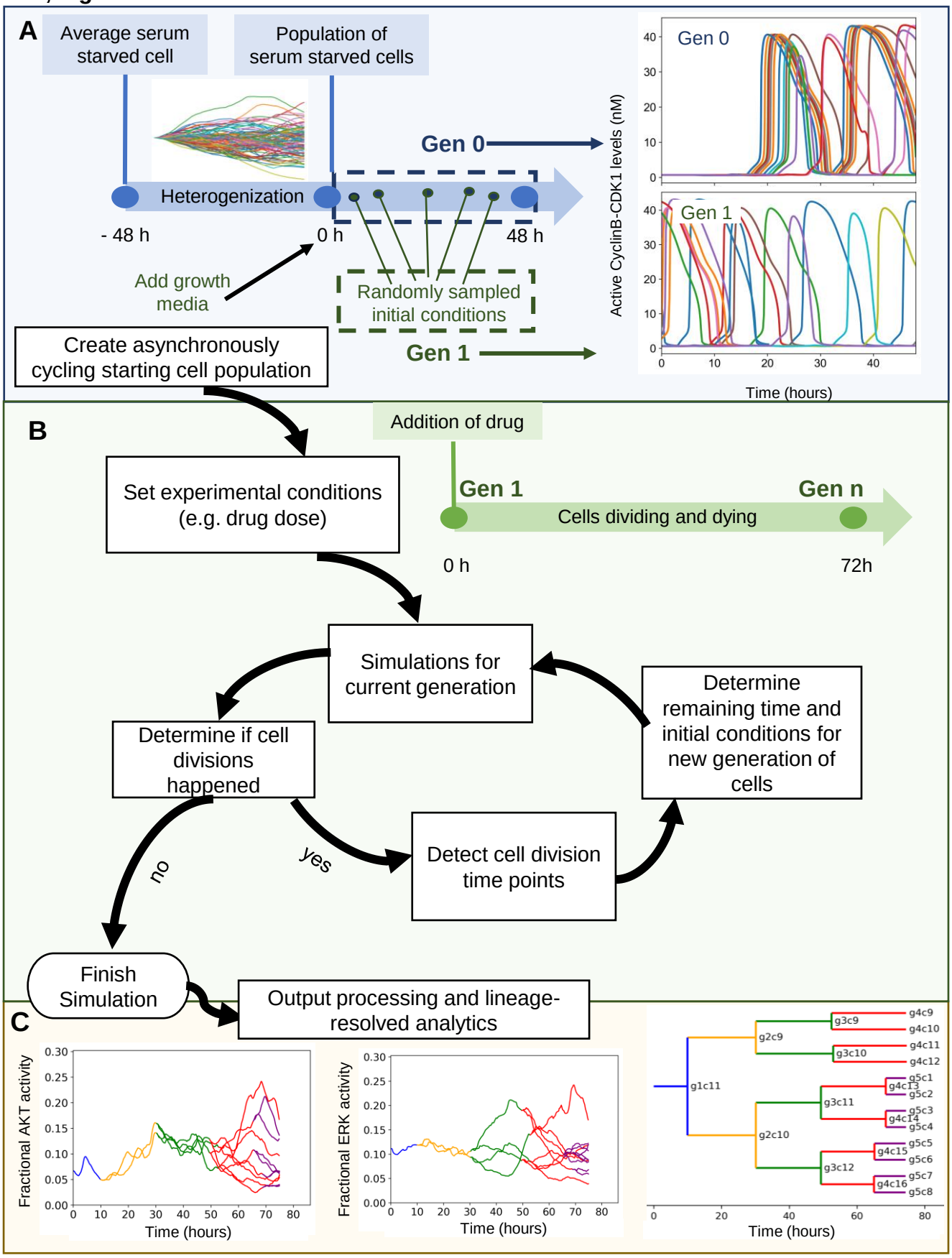

Figure 1. Workflow of the developed simulation algorithm. A. An asynchronously cycling cell population (Gen 1) is initiated by sampling the initial conditions at random time points from a pool of single cell simulations run with growth factor stimulation (Gen 0). **B.** Upon the execution of each generation, detection of new cell division events (or lack thereof) within simulation time determines the creation or not of a next generation. **C.** (left and center) Cross-generational trajectory of observed ERK and AKT activity from a randomly-chosen single-cell lineage. Varying colors represent subsequent generations, starting from Gen 1. (right) *In silico* lineage tracing capability is demonstrated with a lineage dendrogram. Lines representing individual cells are labeled with generation and index.

Mutsuddy et al., Figure 2

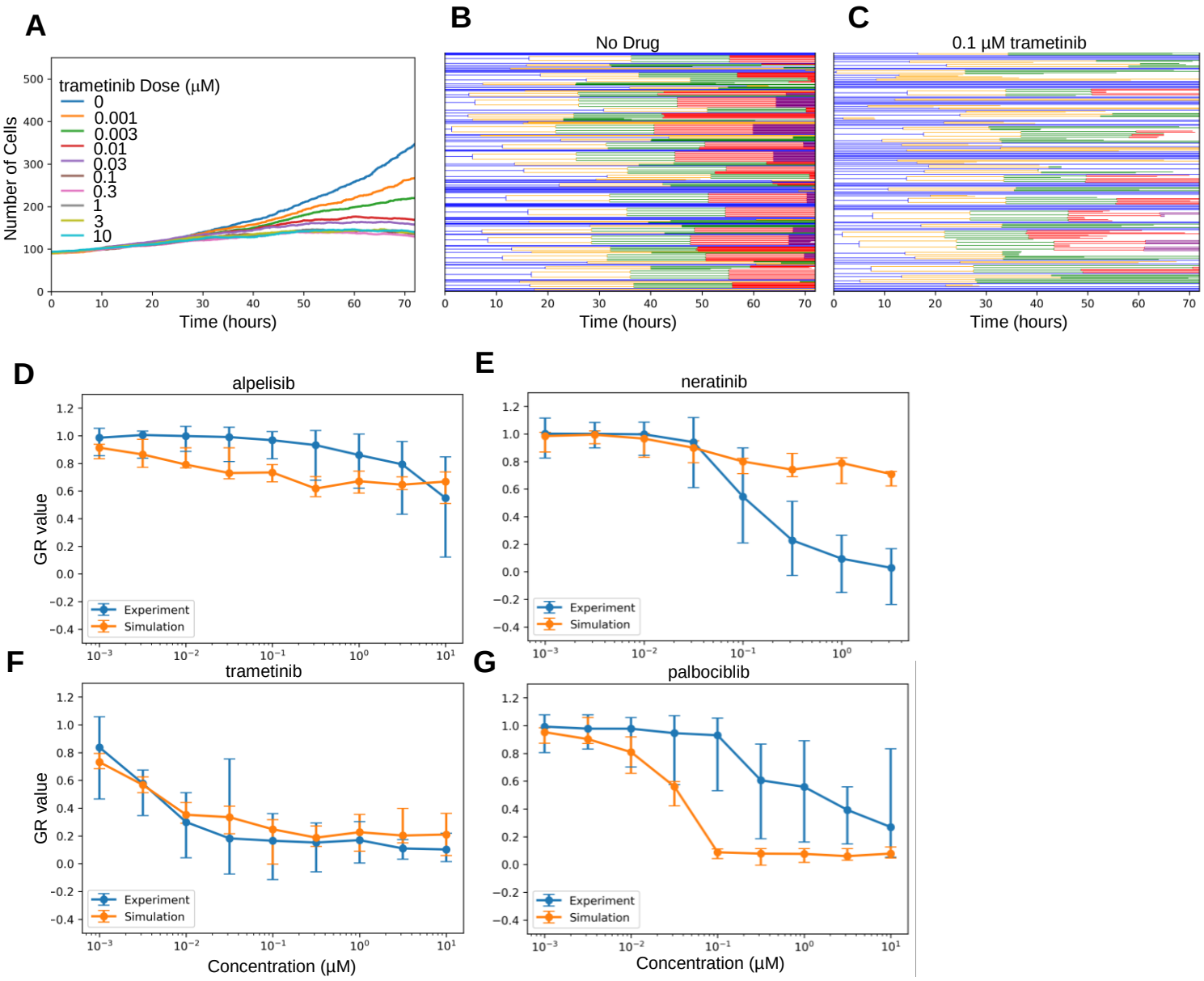

Figure 2. Lineage Resolved Simulations for Comparing Simulated and Experimental Cell Viability Assays. A-C. Dose response simulations for an example drug (trametinib). Median (across simulation replicates) cell population dynamics for several doses (A) and cell population lineage dendrograms for specific doses: 0 μM (B) and 0.1 μM (C) are shown. **D-G.** Simulated dose responses measured in GR-value for four drugs compared experimental data. Error bars are standard error taken from original experimental data, or as calculated across simulation replicates (n=10).

Mutsuddy et al., Figure 3

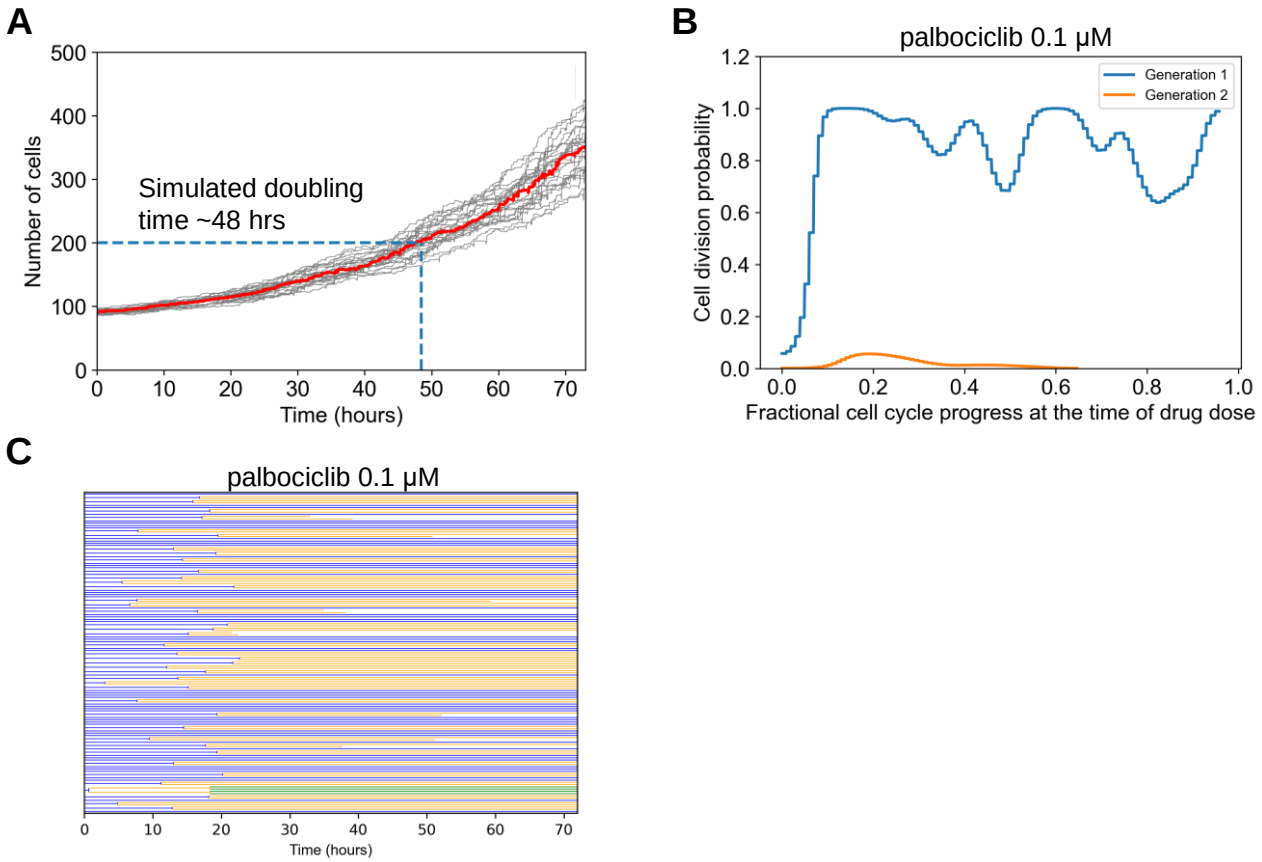

Figure 3. Simulation Analysis to Investigate Palbociclib Dose Response Discrepancy. A. Simulated cell growth curves under control conditions for multiple replicates. The dark black line is the median which was used to estimate doubling time, when the initial cell number (100) doubled (200). **B.** The fractional progression through the cell cycle (see Methods) was estimated for the beginning 100 generation 1 cells. This was the point at which 0.1 μM palbociclib was administered. The division outcome for each cell was then determined for this current generation, and if it exists for the next generation. The probability of division occurring was empirically estimated from this collection of binary outcomes and then plotted. **C.** Cell lineage dendrogram for response to 0.1 μM palbociclib. Most cells divide once early, and then the response is cytostatic.

Mutsuddy et al., Figure 4

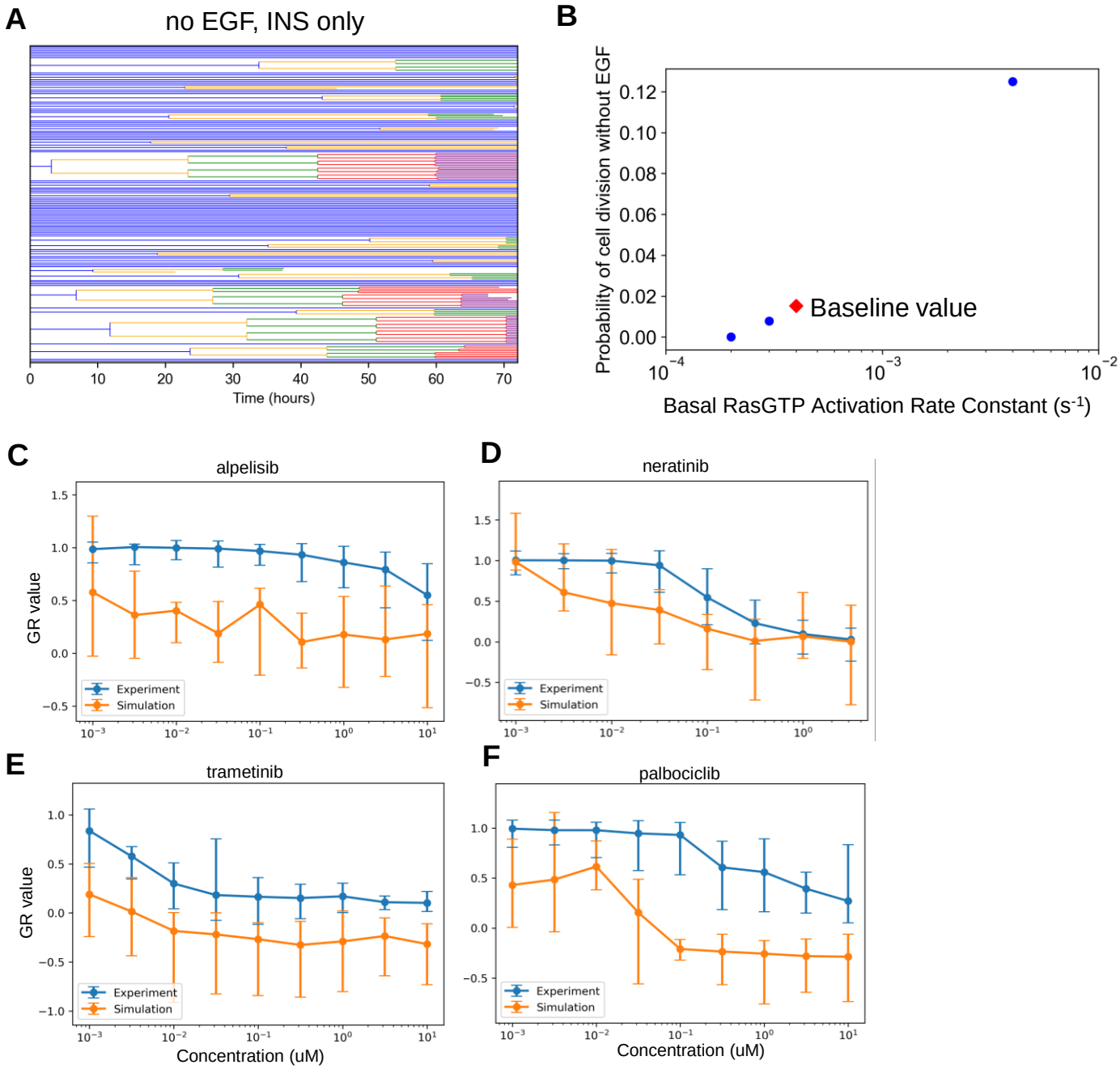

Figure 4. Simulation Analysis to Investigate Neratinib Dose Response Discrepancy. A. Lineage dendrograms under control (no drug) conditions without EGF but with insulin. There are multiple rapidly dividing cells. **B.** Dependence of the probability of cell division as a function of the rate constant controlling basal Ras activation. Simulations were done as in A, without EGF and with insulin. The baseline value for the rate constant in the current published version of the model is designated by the red diamond. **C-F.** Dose response curves as in Figure 2, except with the altered basal RasGTP activation rate constant from panel B ($2x10^{-4}$ s^{-1}).

Mutsuddy et al., Figure 5

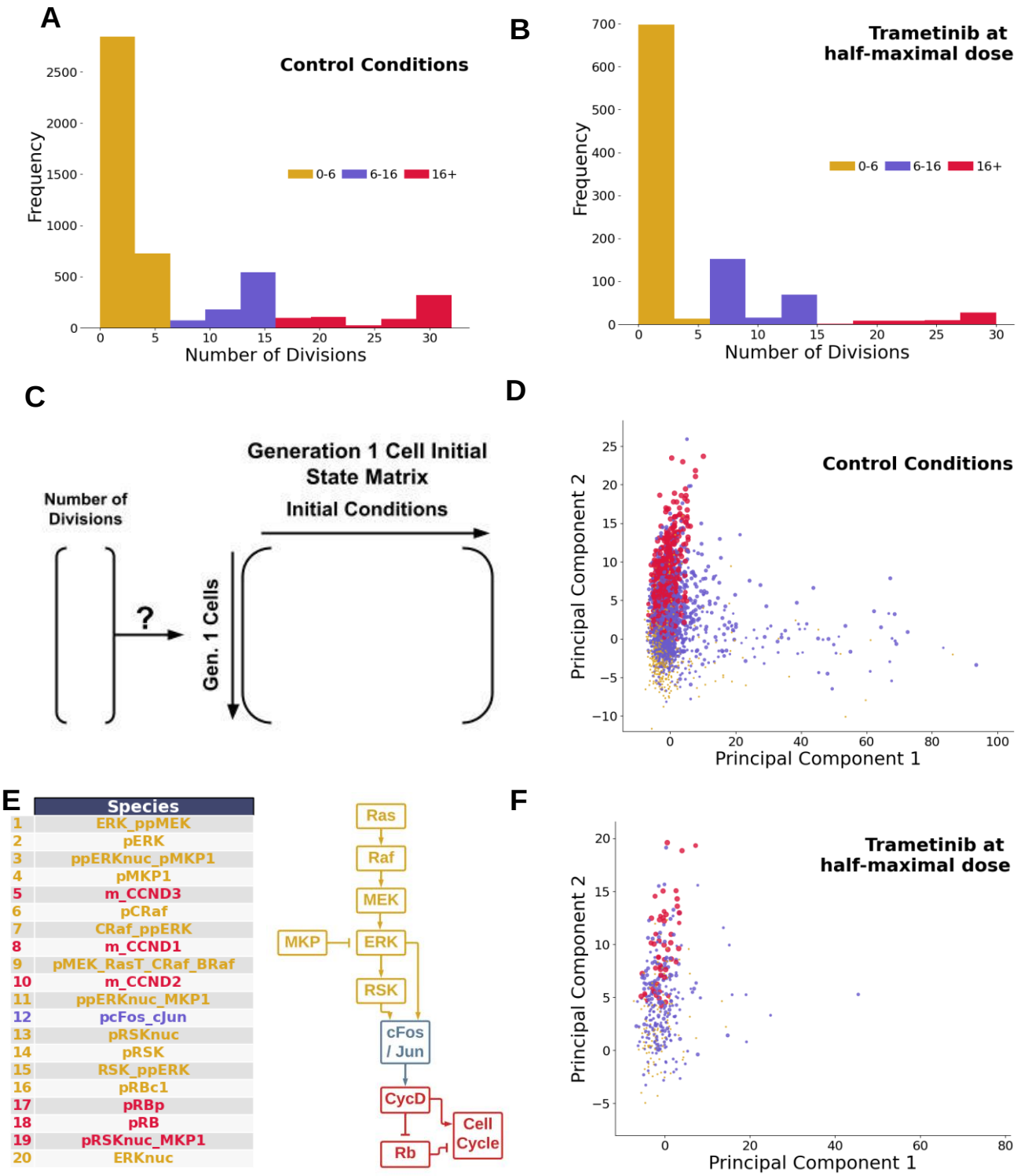

Figure 5. Rapidly Dividing Phenotype in Control and Trametinib-Treated Conditions. **A.** Histogram displaying low (0-6) moderate (6-16) and high (16+) number of division events for simulated cells under control conditions. **B.** Histogram displaying low (0-6) moderate (6-16) and high (16+) number of division events for simulated cells under trametinib treatment conditions (0.03 nM). **C.** Setup of the hypothesis, relating initial conditions of Generation 1 mother cells to the eventual number of divisions arising from them. **D.** Principal component plot of the initial states matrix under control conditions. Points are sized by number of division events, with colors equivalent to panel A. **E.** Top 20 loadings of the second principal component under control conditions, and a cartoon schematic of where they fall along the pathway driving the cell cycle. **F.** Projection of the trametinib data set onto the principal components learned from the control dataset. Points are sized by number of division events, with colors equivalent to panel A.