## Overview

I am an applied mathematician, with research focused primarily on modeling phenomena in the biological and medical sciences. I am interested in a diverse number of biological disciplines, including the emergence of drug resistance in cancer chemotherapy, epidemiological modeling and control strategies, and collective behavior; other research programs include statistical modeling of cancer epidemiology and COVID-19, as I have recently begun utilizing statistical and machine learning tools to analyze population-level (as opposed to cellular-level) data sets. I am also working on understanding the dynamics of the glucose-insulin regulatory system, as well as its interaction with the immune system, with respect to the onset and progression of type 1 diabetes. In all of my projects, I aim to develop and utilize mathematical techniques to discover scientific principles and/or design novel experimental and clinical applications. My work is extremely interdisciplinary, and often includes biological/experimental collaborators, as well as both undergraduate and graduate students.

Since the Fall of 2019, I have been an Assistant Professor in the Mathematics Department at Clarkson University. From the Fall of 2015 to the Spring of 2019, I was a Hill Assistant Professor (Harold H. Martin Postdoctoral Fellow of Mathematics) in the Mathematics Department at Rutgers University in New Brunswick, NJ. While at Rutgers, my mentor was Professor Eduardo Sontag, and I was also a member of the Center for Quantitative Biology (CQB). In the Spring of 2015, I received my Ph.D. in Mathematics from the University of Maryland, College Park. My work as a graduate student was under the direction of Professor Doron Levy, also of the Department of Mathematics and the Center for Scientific Computation and Mathematical Modeling (CSCAMM), of which I was also a member.

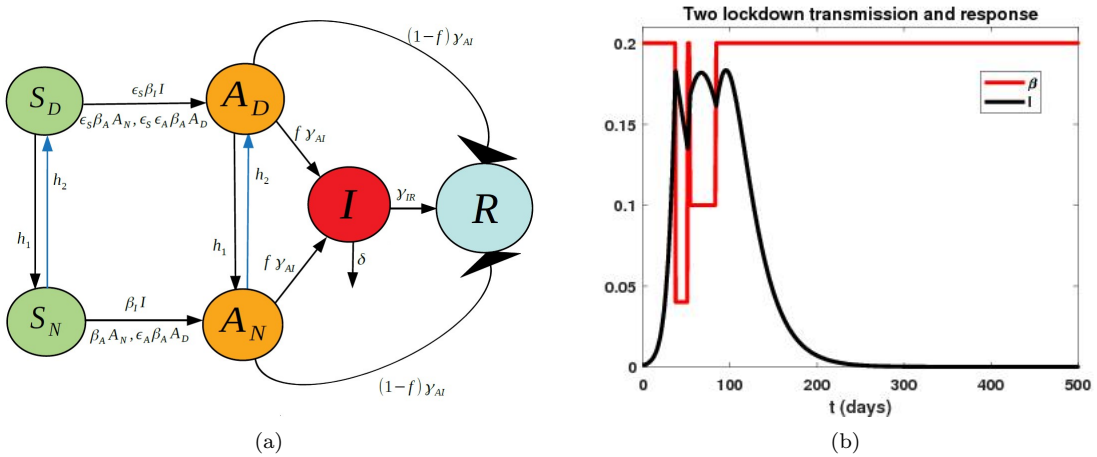
As a mathematical biologist, I integrate a number of branches of mathematics and computer science into my work, including Differential Equations and Dynamics Systems, Control Theory, Differential Geometry, Numerical Analysis, Agent-Based Modeling, and Probability Theory. Using these techniques while working closely with experimentalists, my work (current and future) can be classified as follows:

1. Modeling and control strategies in epidemiology
2. The role of chemotherapeutic induction in drug resistance
3. Combining modeling and machine learning to understand collective motion
4. Analyzing clinical data sets using statistical tools
5. Dynamics of chemical reactions arising from biological circuits
6. Mechanistic modeling as a way to understand the progression of type 1 diabetes
7. The selection/mutation framework of drug resistance
8. Intrinsic heterogeneity and its effect on cancer cell growth
9. The effect of space in drug resistance

In the remainder of this document, I will describe both the biological questions as well as the modeling techniques used in each of the above areas.

## 1 Modeling and control strategies in epidemiology

Early 2020 saw the start of the COVID-19 pandemic, which necessitated both new epidemiological models and analytic tools to answer questions of interest to public health. For example, there was significant interest in formulating models which could incorporate both asymptomatic transmission as well as social distancing as a form of a non-pharmaceutical intervention (NPI), the latter of which at the time was the only available method to inhibit the spread of the disease. Utilizing these models to quantify the effect of the uncertainty of variable transmission rates, as well as the effect of NPIs on disease dynamics, was and remains to be an important aspect of policy design, which seeks to minimize infection rates without complete economic stagnation. Indeed, most of us are now familiar with the phrase "flatten the curve," a central tenet in early pandemic policy, where distancing is implemented to minimize the peak of the infected population, which is generally understood as a measure of strain on the healthcare system. Our initial interest in COVID-19 modeling was to construct a mathematical model capable of incorporating asymptomatic transmission as well as explicit compartments for socially distanced populations [1]. We formulated a




(a) (b)

Figure 1: COVID-19 compartment models and dynamics of social distancing. (a) Model incorporating asymptomatic individuals and social distancing. (b) Optimal timing of social distancing.

compartmental model, represented schematically in Figure 1, which included specific flow rates of individuals from socially distanced to non-distanced compartments; we believe that this was unique at the time, as most models incorporated social distancing via a simple reduction in transmission rates. Using our model, we demonstrated that delays in start times of distancing mandates may have a significant impact on peak infection, and that there exists a "critical implementation delay" after which little can be done to reduce peak infections. We also studied social distancing relaxation strategies, and showed that there exists a critical relaxation rate which prevents a second peak in infections. At the time [1] was first published on medRxiv, very little was known about what would occur once distancing mandates were relaxed; hence we were interested in understanding possible scenarios based on different social distancing protocols.

Motivated by the above work [1], we were then interested in the timing of NPIs (termed *lockdowns*) as to minimize the infected peak. Specifically, we were interested in if there were any commonalities shared by a large class of epidemiological models with respect to optimal timing of social distancing mandates, and how such optimal timings could be characterized. To our surprise, we discovered that many modeling frameworks shared a common trade-off between start time and duration of distancing, which we characterized as a "universal" feature of epidemic models [2]. Furthermore, this trade-off could be explained theoretically by analyzing the classical susceptible-infected-recovered (SIR) model. Fixing a lockdown duration, we were then interested in determining an optimal lockdown schedule based on infected estimates; essentially we wanted a feedback law which would determine when a lockdown should be initiated. Analysis of the SIR model led to us characterize the optimal initiation time of a single non-perfect lockdown [3]. Reopening strategies (i.e. relaxation strategies, as discussed above) were also analyzed in [4], with particular emphasis on university/college campuses. In this work, we showed that by utilizing surveillance testing, a robust *transmission rate independent* lockdown strategy could be implemented which maximized days open for campus communities.

Epidemiological modeling continues to be a topic of my research. Although no longer a policy focus, open questions remain with respect to the timing of social distancing mandates, which may be especially important as new, possibly more dangerous, variants of COVID-19 arise. Similarly, a new direction of my research concerns the competition between disease variants, such as the Omicron and Delta strains of COVID-19. Specifically, I am interested in understanding how epidemiological parameters must be altered to generate novel, *fitter* strains of a disease, and how this can be used to determine the dynamics of the competition between strains, as well as mutational frequencies.

## 2 The role of chemotherapeutic induction in drug resistance

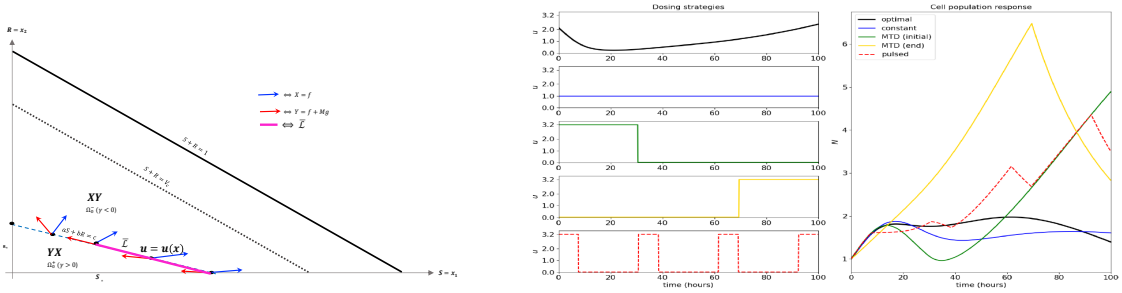
The primary factor limiting the success of chemotherapy in cancer treatment is the phenomenon of drug resistance. Both molecular and microenvironmental factors have been implicated in the development of drug resistance. Molecular mechanisms include the upregulation of drug efflux transporters on the cell membrane, modification of drug targets, enhanced DNA damage repair mechanisms, dysregulation of apoptotic pathways, and the presence of cancer

stem cells. Irregular tumor vasculature, increased regions of acidity, immune cell infiltration and activation, and the tumor stroma are all examples of microenvironmental factors that may inhibit response to chemotherapy. Experimental, clinical, and mathematical research continues to shed light on the multitude of factors that contribute to cancer drug resistance.

As discussed, the mechanisms by which drug resistance presents are extraordinarily diverse. However, a more fundamental question relates to the method by which these resistance-inducing traits arise in a cellular population. Classically, these mechanisms are understood as conferred to the cell by random genetic mutations, from which clonal expansion occurs via Darwian evolution. However, the recent experimental discovery of epigenetics and phenotype plasticity complicates this hypothesis. It is now believed that chemotherapy can produce drug-resistant clones in a dose-dependent manner. That is, the application of chemotherapy has contradictory effects: the agent eliminates a certain subpopulation while simultaneously promoting transitions to a less-sensitive phenotype. It is in this phenomenon, and the resulting control structures, that we are interested in understanding.

In a series of works [5, 6, 7], we have established a mathematical framework to distinguish between drug-selected and drug-induced resistance. In [5], we introduced the mathematical model and studied basic dynamical properties related to treatment outcome and control. Specifically, we demonstrated that the induction rate of a theoretical cancer drug could have a nontrivial impact on the qualitative response to a given treatment strategy, implying this rate must be considered when designing clinical protocols. After establishing the importance of induced resistance, we demonstrated that all parameters in our mathematical model are identifiable utilizing Lie algebraic techniques, meaning that it is theoretically possible to determine the rate at which drug resistance is induced for a given treatment protocol. We also posited novel *in vitro* methods that could be utilized to measure a treatment's induction rate without *a priori* knowledge of the resistance mechanisms.

Having established identifiability, we then investigated how the drug-induced rate of resistance impacts the overall control structure from both mathematical and clinical perspectives [6]. Utilizing a clinically-inspired objective of time to treatment failure (TTF), we applied both the Pontryagin Maximum Principle and differential-geometric techniques to characterize the optimal control and corresponding trajectory (see Figure 2). Numerical solutions are also computed which exhibit close agreement with our analytical results. Mathematical details, including existence results, can be found in [7].




(a) (b)

Figure 2: Characterization of optimal control of drug-induced model of resistance. (a) Switching structure in sensitive-resistant plane. (b) Numerical optimal control (black) and response, together with the response to other standard therapies.

To understand the tradeoff between tumor sensitivity and chemotherapeutic induction, we investigated the success of optimal therapy as a function of these two parameters. Numerical results suggest that cytotoxicity alone does not determine the success of a given therapy, but instead that induction must also be considered when selecting treatments. Our mathematical framework allows us to quantify how effective a potential therapy will be, and can be utilized to personalize therapy based on both patient-specific and pharmacokinetic data. Besides purely theoretical investigations, we have also calibrated our mathematical models to experimental data sets. For example, in [8], we demonstrated that our simple model provided excellent fits to data from **??**; we were further able to predict the time scales for re-population experiments with no additional fitting. Furthermore, we also recently collaborated with the authors of [9, 10] to quantify the role of induction in the adaptive resistance of melanoma cells to RAF inhibitors, as well as to understand the dose-dependency in drug-induced resistance. Utilizing a modified model, we fit all kinetic

parameters non-parametrically as a function of dose, demonstrate that parameters are practically identifiable, and validate the model on doses not used in model calibration. Surprisingly, we discover that the induction rate *decreases* as a function of applied dosage. Using the calibrated model, we numerically investigate an optimal control problem to determine the dosing strategy which minimizes the cancer population at the end of the experiment (100 hours). The computed optimal control is comparable to an intermediate constant therapy, i.e. an idealized metronomic therapy; it is also worth noting that maximum tolerated dose (MTD) is far from optimal. A comparison of standard therapies is provided in the right panel of Figure 2(b). This work has recently been submitted to *NPJ Systems Biology*, and is currently available as a pre-print [11].

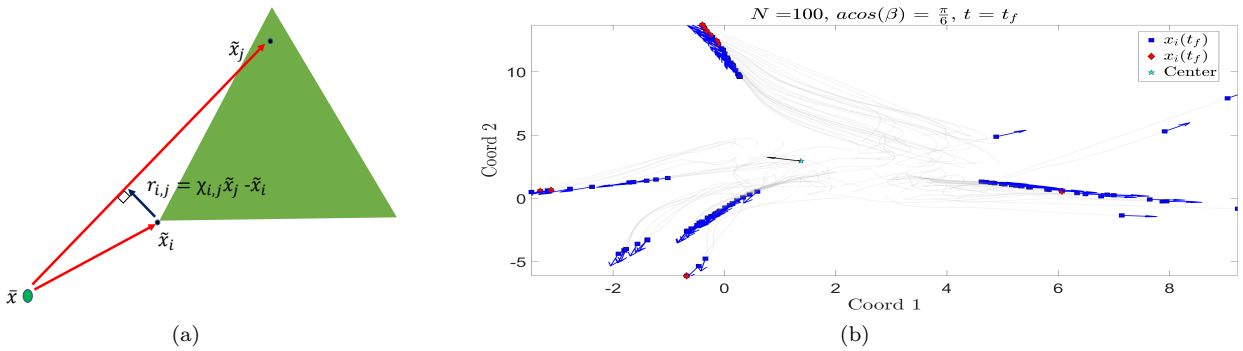
In this framework, we have considered the induction potential of chemotherapy and its effect on control structures from a clinical perspective. Current and future extensions more accurately describing the complex tumor-host microenvironment (see Section 7). Furthermore, we are currently collaborating with the authors of [12] and [13] to precisely characterize the dose-dependence of resistance induction for a class of taxanes, as well as apply our geometric techniques to multi-drug therapies. Indeed, interesting questions arise related to sequential versus combination strategies when chemotherapies with distinct cellular pathways are applied, as is the case in both the experimental setup of [13] and more generally in most clinical scenarios.

## 3 Combining modeling and machine learning to understand collective motion

Collective motion is the phenomenon in which large-scale coherent behavior results from local, generally pairwise interactions among individuals. Collective motion occurs throughout a variety of scales in the biological sciences, including phototaxis/chemotaxis in bacteria, the collective migration of cancer cells, the swarming of locusts and ants, the schooling of fish, the flocking of birds, and lane formation in humans. A fundamental scientific question is understanding what local interaction rules give rise to coordinated movement. Inspired by the emergence of "leader" cells in phototaxis (e.g. the phenomenon of *fingering* [14]), we have recently published [15] a minimal mathematical model that exhibits the emergence of geometric structure in a system of interacting particles. The systems takes the form of a coupled system of nonlinear differential equations, where agents interact not through simple differences of their features, but instead through *projected* differences:

$$\begin{aligned}\dot{\mathbf{x}}_i(t) &= \mathbf{v}_i(t)\\ \dot{\mathbf{v}}_i(t) &= \frac{1}{|\mathcal{N}_i(t)|}\sum_{j\in\mathcal{N}_i(t)}\Big(\phi(|\mathbf{r}_{ij}(t)|)\mathbf{r}_{ij}(t)+\psi(|\mathbf{r}_{ij}(t)|)(\mathbf{v}_j(t)-\mathbf{v}_i(t))\Big)\end{aligned}$$

Here $\mathbf{r}_{ij}$ denotes the orthogonal component of position of the $i^{\text{th}}$ particle projected in the direction of $j^{\text{th}}$ particle, so that agents interact via relative projected directions. Note that the above is a second-order system, but first-order systems are also formulated and analyzed. The resulting dynamics, demonstrated in Figure 3, provide clear local emergence of linear spatial patterns moving coherently, and thus provide a hypothetical mechanism for the emergence of social hierarchy in interacting particle systems. Furthermore, we were able to prove a theorem which quantifies this emergence precisely as a kinetic/potential energy trade-off, and to our knowledge, is the first large-time, large-crowd emergence dynamics based on *local* interactions.




(a) (b)

Figure 3: Emergence of lines in a second-order system inspired by phototaxis in bacterial cells. (a) Projection based interactions, and (b) numerical simulations of the second-order system.

I continue to work in a variety of directions in this field, including collaborative projects with experimentalists to combine modeling and learning methods to *infer* interaction rules based on measured trajectory data. Specifically, in the summer of 2022, a pair of REU students in my Mathematical Biology Team Science (MBioTS) Research Experience for Undergraduates (REU) worked together with myself and a colleague in the Biology department (Professor Shantanu Sur) to study collective motion both experimentally and theoretically. Ongoing and future projects that merge experimental data with collective motion modeling and learning include determining the role calcium plays in cell-cell adhesion, the effect of hypoxia on cell motility, and the therapeutic potential of the toxic peptide Amphiphile in cancer treatment. Outside of modeling (i.e.the "forward" direction) related to the biological questions above, current work with collaborators in machine learning and dynamical systems theory focuses extensively on the "inverse problem," where data-driven techniques can be utilized to learn interaction mechanisms given a set of experimental data; our work aims to both understand this precisely via theoretical analysis of machine learning algorithms and stochastic differential equations, as well as to apply these methods to biological data sets. Specifically, we are working on direct extensions of the techniques introduced in [16, 17, 18, 19] to second-order stochastic differential equations of the form

$$\begin{cases} d\mathbf{x}_i &= \mathbf{v}_i dt + \sigma dW^i, \\ d\mathbf{v}_i &= \frac{1}{N}\sum_{j=1}^N \Big[\phi^E(|\mathbf{x}_j - \mathbf{x}_i|, |\mathbf{v}_j - \mathbf{v}_i|)(\mathbf{x}_j - \mathbf{x}_i) \\ &\qquad +\phi^A(|\mathbf{x}_j - \mathbf{x}_i|, |\mathbf{v}_j - \mathbf{v}_i|)(\mathbf{v}_j - \mathbf{v}_i)\Big] dt + \sigma d\widetilde{W}^i \end{cases}$$

with the goal of inferring the interaction kernels $\phi^E$ and $\phi^A$ from observed (discrete) trajectory data of the form $\{\mathbf{x}_i(t_l), \mathbf{v}_i(t_l)\}_{i,l=1}^{N,L}$ via minimization of the least-squares inspired functional

$$\mathcal{E}(\varphi^E, \varphi^A) = \frac{1}{2\sigma^2 LMN} \sum_{m,l,i=1}^{M,L-1,N} \langle \mathbf{v}_i(t_{l+1}) - \mathbf{v}_i(t_l), \frac{1}{N}\sum_{j=1}^N \Big[\varphi^E(|\mathbf{x}_j(t_l) - \mathbf{x}_i(t_l)|)(\mathbf{x}_j(t_l) - \mathbf{x}_i(t_l)) + \varphi^A(|\mathbf{x}_j(t_l) - \mathbf{x}_i(t_l)|)(\mathbf{v}_j(t_l) - \mathbf{v}_i(t_l))\Big]\rangle.$$

We are also working on learning state-dependent noise, i.e. methods to reverse engineer the functional dependence in the case when $\sigma = \sigma(\mathbf{x}_i)$, which of interest in biological contexts. For example, this could be utilized to unravel the role of technical (e.g. measurement) versus intrinsic (e.g. biological) noise in observation data.

## 4 Analyzing clinical data sets using statistical tools

The analysis of large clinical data sets provides an opportunity to address issues that may have an immediate impact on public health. Furthermore, the relatively recent availability of open-source data sets allows researchers to re-analyze previously collected data from new perspectives. I often incorporate data sets into course-based projects, as well as research (and sometimes both, e.g. [20]). Since becoming the primary investigator (PI) on an NSF-funded Research Experience for Undergraduates (REU) program, I have designed a number of projects related to the statistical/data scientific analysis of publicly available data sets related to both cancer and COVID-19.

During the summer of 2022, and utilizing data from Project Data Sphere (`https://data.projectdatasphere.org`), I began working (together with faculty in both Mathematics and Biology) on a project related to prediction in machine learning algorithms using imbalanced data sets. Specifically, we wanted to understand the role of different explanatory variables in cancer incidence; for example, in data taken from the VITAL trial [21], we were interested in understanding the *relative* importance of predictors (e.g. vitamin D intake, age, sex, race, smoking history, BMI) on the probability of developing cancer by a certain age. To quantify this importance, we applied binary logistic regression in combination with partial-dependence plots (PDPS) [22] to "rank" predictors. However, due to the data imbalance (a small percentage of the participants developed cancer during the study), results of the regression analysis were extremely biased towards the majority class. We then developed a method which combines a novel undersampling technique with binary logistic regression to determine relative importance of biometric and socioeconomic variables in disease incidence. A key aspect of our approach is that it involves a minimal amount of data manipulation, and relies entirely on actual, as opposed to synthetic, data. We applied to our "pipeline" to the VITAL data set to analyze the role of vitamin D across ethnic groups. One finding of our analysis showed that the effect of vitamin D on reducing cancer incidence in the non-Hispanic Black population was much less significant in comparison with other predictors, due mainly to the manner in which the population was sampled. Work related

to this project was published in [23], and was also presented by students from the REU at numerous conferences, including the 2023 Joint Mathematics Meetings.

During the summer of 2023, again utilizing publicly available data from Project Data Sphere and in conjunction with faculty from Biology and REU undergraduate participants, we began re-analyzing data related to the efficacy of combining pemetrexed and bevacizumab as maintenance therapies for advanced nonsquamous non-small-cell lung cancer (NSCLC). Maintenance therapy with either pemetrexed or bevacizumab alone had been shown to increase overall survival in NSCLC, one of the most common subtypes of lung cancer. Other preliminary studies found that the combination of the two in maintenance therapy might have benefits for progression-free survival, and the combination was adopted for routine therapy. However, according to the ECOG-ACRIN 5508 clinical trial, high toxicity induced by the combination therapy offsets its advantages for progression-free survival. We were interested in the relation between the degree of toxicity and response. Utilizing logistic regression, it appears that toxicity is a significant predictor of progression-free survival in the combination therapy, which is not observed in individual therapies alone. Further analysis needs to be performed to determine the exact relationship between treatment efficacy and toxicity, and specifically if we can understand which side effects correlate with improved response. We expect this work to be submitted for publication in late 2024.

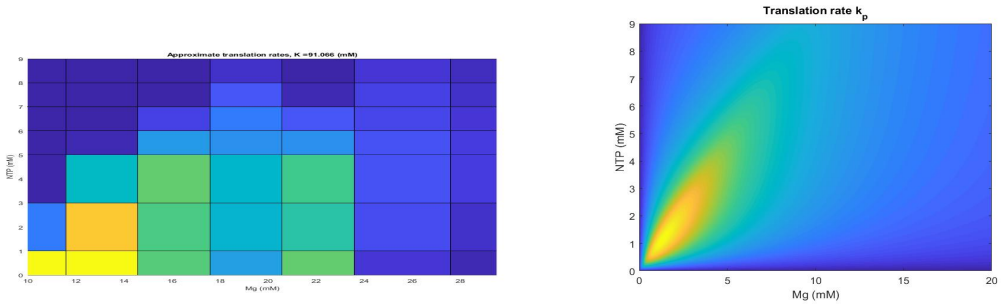
With another group of undergraduates (including a high school student), and related to the study of risk factors and disease as in [23], we have recently showed how multiple risk factors could affect vulnerability to COVID-19. We analyzed a total of 3040 US counties, and showed the potential of utilizing quartile clustering in conjunction with negative binomial regression to better understand the interaction effects on the infection and fatality rates with respect to selected comorbidities, as well as demographic and socioeconomic risk factors. The data was separated into two equal phases that correspond approximately to the first two pandemic waves; each phase was further subdivided into three groups (low, medium, and high) based on infection and fatality rates. Negative binomial regression models without interaction effects were first constructed to identify significant risk factors, followed by the construction of models considering only the significant risk factors and their interaction effects. Our analysis showed that many bivariate interactions, including population density and the age of the population, the male and non-Hispanic Black population, and poverty rate and age of the population are jointly associated with increased infection, while the interactions of non-Hispanic Black ethnicity and less education with the percentage of the population above 65 years were associated with increased fatality. Our findings suggest that selected sub-populations defined by the intersection of specific risk factors could be more vulnerable to COVID-19, and a manuscript related to this work is being prepared for submission.

## 5 Dynamics of chemical reactions arising from biological circuits

In the last twenty years, a major goal of synthetic biology is the design and implementation of biological circuits. Analogous to electronic "breadboarding", researchers attempt to generate tunable modules which can be easily deployed in living systems with the purpose of modifying cellular function, influence development, creating environmental response pathways, etc. While I have not explicitly designed such circuits in the laboratory, I am involved in several projects related to this subfield of systems biology.

Through correspondence with the Biophysics department at Rutgers University, I became interested in resource competition in the processes of transcription and translation [24]. More specifically, there were experiments being performed in a cell-free transcription-translation (TX-TL) system, which provides a minimally complex environment to characterize genetic networks. Resource requirements must be precisely calibrated for optimal expression, and an unexpected phenomenon of decreased expression with increasing nucleotide triphosphate (NTP) concentration past a certain threshold complicated system design and throughput (Figure 4). It then became clear that, even in idealized experimental configuration, the energy landscape for fundamental biological processes was more complicated than initially supposed. Through a combination of further experimental and mathematical modeling, we explained this via a direct suppression of gene expression, specifically that of the translation rate as a function of the overabundance of certain reagents. Using novel numerical techniques to extract time-dependent translation rates from data, we observed a complex interdependence between $Mg^{2+}$ and NTP concentrations on the translation rate dynamics. A mechanistic mathematical model was derived and analyzed which explained this phenomena via a "poisoning" by both unbound NTP molecules and unbound $Mg^{2+}$ ions. We note that the latter was well understood in the literature, but the former was unknown at the time.

A serious weakness of genetic circuits is the slow timescales present in transcription-based circuits. Partnering with member of the Chemistry department at Rutgers, I have been involved in a project related to the design and analysis of enzymatic circuits. From the experimental perspective, self-modifying and evolving enzymatic circuits




(a) (b)

Figure 4: (a) Experimental data and (b) mathematical model of translation rate dynamics as a function of resources.

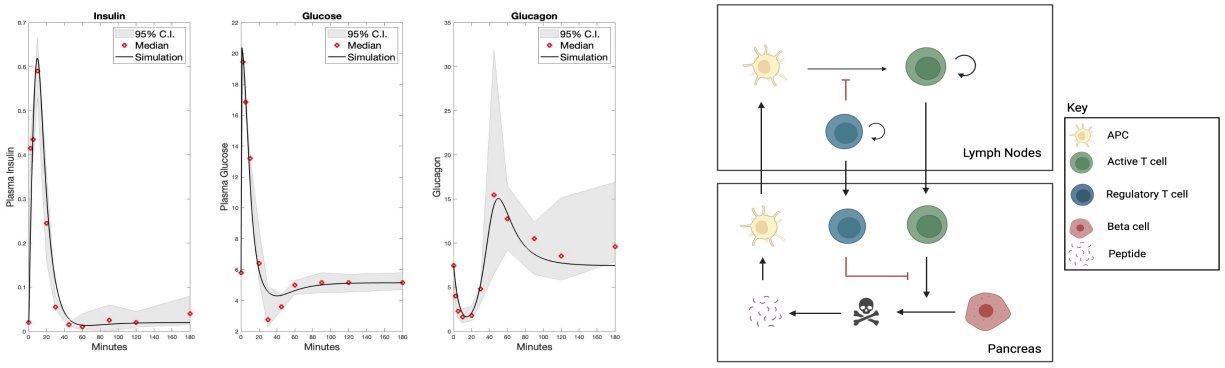
are being implemented using designed/engineered enzymes that can be made responsive to chosen input stimuli such as small molecule chemicals and/or light. One of the many theoretical aspects related to this project includes understanding the dynamical properties of the three-body binding problem, which experimentally is related to the activation of an enzyme via a small molecule. To obtain both maximal output and sensitivity of the designed circuit, a detailed knowledge of the properties of the underlying chemical reaction network must be understood. To our surprise, the system possessed a generally non-monotonic response pattern. Mathematically, using tools from both Horn's and Feinberg's deficiency theory as well as novel analytic/numeric techniques, we are able to precisely quantify this behavior as a function of the kinetic constants of the network. I am currently also studying parameter identifiability and controllability, in the context of input/output systems, of the proposed circuits.

## 6 Mechanistic modeling as a way to understand the progression of type 1 diabetes

Type 1 diabetes is a rare autoimmune disorder with potentially life-threatening consequences if left untreated. The condition arises due to the immune system's targeted destruction of insulin-producing beta cells within the pancreas and constitutes approximately 5% to 10% of all diabetes cases. Presently, there is no known cure for this disease, necessitating regular monitoring of blood glucose levels and the administration of daily subcutaneous insulin injections to uphold healthy glucose levels in affected individuals. However, despite ongoing advancements in diabetes care, those with type 1 diabetes experience a diminished life expectancy, facing a reduction of more than 12 years compared to their non-diabetic counterparts.

The dynamics of the disease are comprised of two inter-connected components: the insulin-glucose regulatory system, and the immune system. Together with a student, I have begun to analyzes both facets of this complex system [25]. We regards to the insulin-glucose regulatory system, we have proposed a novel model which emphasizes the role of glucagon, a hormone secreted by alpha cells to increase glucose production, in regulation; this model is compared to the standard "minimal model," [26] which is widely used in clinical applications, but includes non-observable insulin compartments and non-mechanistic dynamics. In contrast, our proposed model is entirely mechanistic, provides improved fits to longitudinal data (see Figure 5(a)), and yields an identifiable measure of insulin sensitivity, an important patient-specific parameter used to measure insulin resistance. The work from this project is currently under revision at *Mathematical Biosciences*.

We are also in the process of incorporating the above glucose-insulin-glucagon system into a model of autoimmune disease, which includes cytotoxic (activated) T cells (the primary antagonist of beta cells in the pancreas), regulatory T cells, and memory T cells (see Figure 5(b) for a portion of the immune system network). A primary goal here is to quantify the interaction between regulatory and activated T cells, a mechanism that is not well understood biologically, and to see if qualitative features of dynamical response/adaptation can be associated with biological observations and disease progression; the goal is thus to infer cellular mechanisms from our dynamical model. We eventually plan to study the impact of the administration of immunosuppressants on delaying type 1 diabetes progression, as well as apply tools from optimal control theory on designing dosing regimens.




(a) (b)

Figure 5: Type 1 diabetes modeling. (a) Glucose-insulin-glucagon dynamics, including both the experimental data (red) and theoretical model (black). (b) Schematic for T-cell activation and regulation in the pancreas and lymph nodes.

## 7 The selection/mutation framework of drug resistance

Due to high rates of cellular mutation and genetic instability, in comparison to healthy cells, populations of cancerous cells constitute an exceptionally heterogeneous environment. This heterogeneity implies that there are many different cellular genotypes and phenotypes present in the same disease. These different cellular "types" exhibit varying responses to the same chemotherapeutic treatment, and thus give rise to drug resistance. For instance, cells may exhibit genetic diversity in varying levels of drug efflux or uptake, different cell-cycle lengths or apoptotic pathways, drug sequestration, or drug metabolism. Furthermore, non-genetic deviations among cells such as their location relative to the blood supply or local density can also impact the efficacy of cytotoxic agents. As cells can exhibit one or a number of these factors in different degrees, drug resistance presents as a complex problem in heterogeneous population dynamics.

To model this phenomena, we constructed a system of integro-differential equations (IDEs), structured by a parameter representing the resistance level. For example, these equations take the general form

$$\begin{aligned}\frac{\partial n(x,t)}{\partial t} &= \Big(f(\rho(t))[r(x)(1-\theta(x)) - h(D(t),x)] - g(\rho(t))d(x)\Big)n(x,t) \\ &\quad + f(\rho(t))\int_0^1 \theta(y)r(y)M(y,x)n(y,t)\,\mathrm{d}y.\end{aligned} \tag{1}$$

Many classical mathematical works on drug resistance consider the resistance level as discrete (e.g. two population only, sensitive and completely resistant); however the multitude sources of heterogeneity led us to consider resistance as a continuous phenomena (variable $x$ in (1)), which affects all aspects of cellular rate processes. That is, the rates of division ($r(x)$), natural death ($d(x)$), and drug-induced death ($h(D(t),x)$) all depend on the resistance level that the cell exhibits. Furthermore, the model also includes density dependencies on the aforementioned rates ($f(\rho)$ and $g(\rho)$), as well as various types of mutations between resistance levels ($M(y,x)$). Indeed, the novelty of the model arises from the combination of a continuous trait, non-exponential cell growth, and epigenetic as well as genetic mutations.

In [27], we perform a mathematical and numerical analysis of the dynamics of the model. In particular, we focus on the asymptotic heterogeneity profiles of the cellular population in large time. As the characterizing parameter is continuous, the regime of convergence is distributions. Comparing the limiting distribution to that of other classical formulations (which can be derived as particular cases of the same model), we discover that only the full model which incorporates density-limited growth/death and non-zero mutations produces a truly heterogenous population that is biologically reasonable. For example, a model without mutations can be shown to converge to a distribution that is the sum of finitely many $\delta$-functions, i.e. point masses over the trait space, and hence express extremely limited heterogeneity. On the other hand, without density limitations, the population will either grow unbounded or become extinct, both of which are unreasonable for the majority of tumors. Thus, this analysis validates all included

aspects of the model, in terms of biological significance, and provides a framework that can be extended to model more realistic combinations of drug resistance mechanisms.

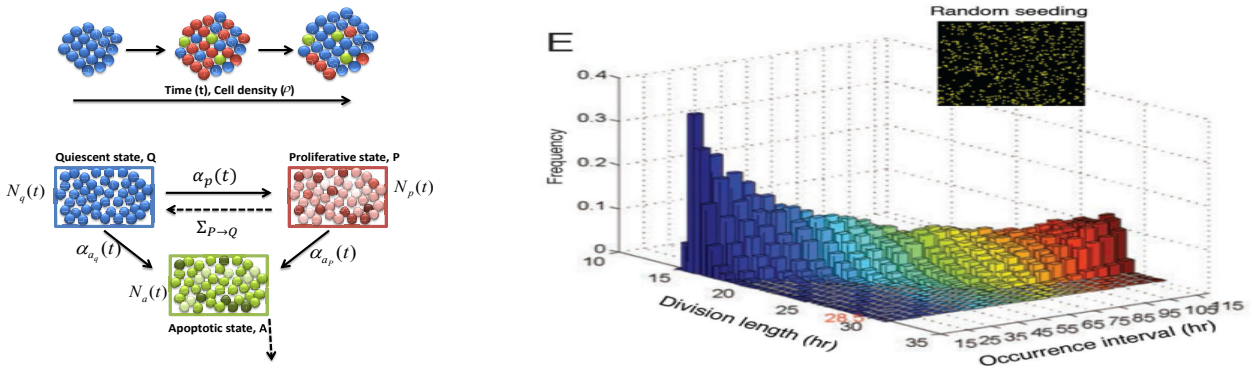
Applications of the above framework to the clinical setting appear in [28] and [29]. In [28], we study the effects of treatment on the size and heterogeneity of tumor cell populations. First, we show qualitative agreement of our model and experimental evidence that suggests that dividing cells are most likely to be eradicated first, whereas the relatively slower growing (or dormant) cells are the major contributors during tumor re-initiation after chemotherapy. Different drug treatment strategies are also investigated, and optimal strategies are proposed. For instance, a strategy of combining multiple drugs, in which the first drug causes an increase in the death rate and the second drug causes a decrease in the cell division rate, is shown to result in a smaller tumor load in our simulations. An extension of the model is also investigated, in which the effects of treatment by mutagenic drugs, i.e. cytotoxic agents that increase the mutational rate of cells during treatment, were incorporated. By investigating numerical simulations of these models, novel treatment strategies were also discovered. We recommend bioengineering treatments that target the mutational rate of cells, so as to yield a relatively homogeneous population, as such a population is more sensitive to traditional chemotherapeutics. However, if the mutation rate for a specific disease cannot be decreased, we suggest the opposite: apply a treatment which temporarily increases the rate of mutations. Our model suggests that the increase in the mutation rate will spread the disease to traits that are ill-favored in their specific environment, and will thus decreasing the overall tumor size.

In [29], we study the application of our model to actual clinical data, with particular emphasis on metastatic cancer. Here we show that a continuous resistance variable is necessary to describe patient data, as well as describe a way of viewing an exponentially growing metastatic disease as a sum of density-limited tumors distributed throughout the body. Furthermore, variations in the rate distributions are analyzed, and are shown to have a large effect on the final tumor size, as well as on the overall heterogeneity.

In this work, we view cancer cells as a structured population. Using a modeling framework which incorporates density dependencies on rate processes, as well as mutations via division errors, we are able to successfully capture real clinical data and observed phenomena, as well as predict treatment strategies for minimizing both overall drug resistance and tumor size. In the future, I plan to continue this work by adapting the models to specific resistance mechanisms, as well as incorporating probabilistic elements to better model the metastatic spread of the disease.

## 8 Intrinsic heterogeneity and its effect on cancer cell growth

The previous work focuses on modeling a population of cancer cells as an aggregate, with heterogeneity arising from distributions specified on the rate parameters. In this work, found in [30], heterogeneity manifests from an explicit mechanism, and we study the dynamics of the growth and death of cancer cells from an individual perspective.




(a) (b)

Figure 6: Cell-cycle dynamics in a cancer population. (a) Cellular transitions in a cancer cell growth model. (b) Distribution of cell-cycle lengths due to the effects of space.

It is well known that the cell-cycle has no fixed duration, but rather appears as a probabilistic event with lengths determined by distributions. Furthermore, as most chemotherapies affect cells primarily in the cell-cycle, understanding the heterogeneity arising from non-constant cell-cycle lengths and how it impacts growth is an important step

towards the successful treatment of cancer. In this work, we describe the dynamics at the cellular level, with cells residing in one of three compartments: quiescent (Q), proliferative (P), or apoptotic (A). The system is described as a continuous time Markov chain, and as such, transitions are governed by probabilistic rates. Furthermore, cells undergoing division remain in compartment P a random amount of time $L$, where $L$ is governed by an intrinsic probability distribution. An analogous assumption holds for cells in the apoptotic compartment. See Figure 6(a) for an outline of all cellular transitions in the model.

This project demonstrates a clear synergy of mathematics and experimental research. As mathematicians, we assisted in the design of experiments with our collaborators at the National Institutes of Health (NIH) for data that was directly implementable into our Markov chain models. For example, measurements of equilibrium distributions of division and apoptotic population fractions were obtained, and provided input into the probabilistic rate terms that we believe are novel and valuable modeling techniques. Cell growth data was also used to determine model parameters that best matched the experimental data. We use analysis and probability theory to derive a corresponding deterministic IDE model for expected values that well-approximates the ABM (both in average value and individual realizations), which we verified through various numerical simulations for a wide range of parameters. For example, the equation approximating the expected value of the proliferative compartment $N_p$ is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}N_p(t) &= \alpha_p(t)N_q(t) - \alpha_{a_p}(t)N_p(t) \\ &\quad - \int_0^t f_p(t-t_*;\mu,\sigma)\alpha_p(t_*)N_q(t_*)\left(1-\int_{t_*}^t \alpha_{a_p}(s)\,\mathrm{d}s\right)\,\mathrm{d}t_*.\end{aligned} \tag{2}$$

Here $N_q$ is the number of cells in compartment Q, $\alpha_p$ and $\alpha_{a_p}$ denote transition rates, and $f_p$ governs the distribution of times spent in the cell-cycle. We note that this IDE system is qualitatively different from that appear in Section 7, as in this work the integration captures a series of distributed and continuous time delays, whereas in (1) it describes the cumulative effect of the structured population on the individual compartments. Using optimization theory on the derived IDE, we also show that a distribution of cell-cycle lengths is indeed necessary to describe the experimental findings.

## 9 The role of space in drug resistance

The work in Section 8 assumes an intrinsic distribution on cellular process lengths. While it is important to understand how the effects of distributions on cell growth, it is also important to understand the mechanism that produces this heterogeneity. Indeed, once understood, it can possibly be controlled and harnessed in treatment protocols. One possible mechanism that explains this observed heterogeneity is the spatial configuration of cells.

In [31], we assumed a similar framework as in [30], with cells undergoing transitions between the three compartments via a continuous Markov chain model. However, the model is updated to include physical space as well, with cells exhibiting aggregation phenomena via stochastic differential equations (SDEs):

$$dX_k(t) = F_a[X(t)](X_k(t))dt + F_r[X(t)](X_k(t))dt + \sigma dW_k(t) \tag{3}$$

Here $F_a$ and $F_r$ denote the attractive and repulsive forces of locomotion, respectively, on a cell due to interactions with other cells, $\sigma W_k$ accounts for the observed random properties of motion via a Wiener process, and $X(t)$ denotes the empirical measure on the physical space due to the distribution of cells. Furthermore, the transitions between compartments (as in Figure 6(a)) are governed by local measurements, such as a cell's local calculation of density. Variation in cell-cycle is then assumed to result from these measurements, with local density determining the speed at which a cell progresses through the cell-cycle, and hence actual cell-cycle length. Preliminary experimental results suggest local density does impact cellular transitions, as well as the efficacy of drug treatment. From a modeling perspective, we show how the emerging spatiotemporal heterogeneity of one cell population can be attributed to differences in local cell density and cell cycle (see Figure 6(b)). Manipulation of the geometric arrangement and spatial density of cancer cells revealed that given a fixed global cell density, significant differences in growth, proliferation, and paclitaxel-induced apoptosis rates were observed based solely on cell movement and local conditions.

In the future, I hope to continue in my study of interesting and salient topics in mathematical biology. Many of the projects discussed above are ongoing, and a continued goal of my research is to further develop collaborations with other mathematicians and scientists, and to use my mathematical and biological knowledge to understand and address problems in science.

# References


[1] Jana L Gevertz, James M Greene, Cynthia H Sanchez-Tapia, and Eduardo D Sontag. A novel covid-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. *Journal of Theoretical Biology*, 510:110539, 2021.

[2] Mahdiar Sadeghi, James M Greene, and Eduardo D Sontag. Universal features of epidemic models under social distancing guidelines. *Annual reviews in control*, 51:426–440, 2021.

[3] James M Greene and Eduardo D Sontag. Minimizing the infected peak utilizing a single lockdown: a technical result regarding equal peaks. In *2022 American Control Conference (ACC)*, pages 3640–3647. IEEE, 2022.

[4] Mackenzie Dalton, Paul Dougall, Frederick Laud Amoah Darko, William Annan, Emmanuel Asante-Asamani, Susan Bailey, James Greene, and Diana White. Modeling optimal reopening strategies for covid-19 and its variants by keeping infections low and fixing testing capacity. *PLoS One*, 17(11):e0274407, 2022.

[5] James M Greene, Jana L Gevertz, and Eduardo D Sontag. Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. *JCO clinical cancer informatics*, 3:1–20, 2019.

[6] James M Greene, Cynthia Sanchez-Tapia, and Eduardo D Sontag. Control structures of drug resistance in cancer chemotherapy. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 5195–5200. IEEE, 2018.

[7] James M Greene, Cynthia Sanchez-Tapia, and Eduardo D Sontag. Mathematical details on a cancer resistance model. *Frontiers in Bioengineering and Biotechnology*, 8:501, 2020.

[8] Jana L Gevertz, James M Greene, and Eduardo D Sontag. Validation of a mathematical model of cancer incorporating spontaneous and induced evolution to drug resistance. *bioRxiv*, pages 2019–12, 2019.

[9] Mohammad Fallahi-Sichani, Verena Becker, Benjamin Izar, Gregory J Baker, Jia-Ren Lin, Sarah A Boswell, Parin Shah, Asaf Rotem, Levi A Garraway, and Peter K Sorger. Adaptive resistance of melanoma cells to raf inhibition via reversible induction of a slowly dividing de-differentiated state. *Molecular systems biology*, 13(1):905, 2017.

[10] Natacha Comandante-Lou, Mehwish Khaliq, Divya Venkat, Mohan Manikkam, and Mohammad Fallahi-Sichani. Phenotype-based probabilistic analysis of heterogeneous responses to cancer drugs and their combination efficacy. *PLOS Computational Biology*, 16(2):e1007688, 2020.

[11] Eduardo D Sontag, Jana L Gevertz, James Greene, Natacha Comandante-Lou, and Samantha Prosperi. Understanding therapeutic tolerance through a mathematical model of drug-induced resistance. *bioRxiv*, pages 2024–09, 2024.

[12] Angela Oliveira Pisco, Amy Brock, Joseph Zhou, Andreas Moor, Mitra Mojtahedi, Dean Jackson, and Sui Huang. Non-darwinian dynamics in therapy-induced cancer drug resistance. *Nature communications*, 4(1):2467, 2013.

[13] Aaron Goldman, Biswanath Majumder, Andrew Dhawan, Sudharshan Ravi, David Goldman, Mohammad Kohandel, Pradip K Majumder, and Shiladitya Sengupta. Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. *Nature communications*, 6(1):6139, 2015.

[14] Rosanna Man Wah Chau, Devaki Bhaya, and Kerwyn Casey Huang. Emergent phototactic responses of cyanobacteria under complex light regimes. *MBio*, 8(2):e02330–16, 2017.

[15] James M Greene, Eitan Tadmor, and Ming Zhong. The emergence of lines of hierarchy in collective motion of biological systems. *Physical Biology*, 2023.

[16] Fei Lu, Ming Zhong, Sui Tang, and Mauro Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. *Proceedings of the National Academy of Sciences*, 116(29):14424–14433, 2019.

[17] Jinchao Feng, Mauro Maggioni, Patrick Martin, and Ming Zhong. Learning interaction variables and kernels from observations of agent-based systems. *IFAC-PapersOnLine*, 55(30):162–167, 2022.

[18] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. *The Journal of Machine Learning Research*, 22(1):1518–1584, 2021.

[19] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories. *arXiv preprint arXiv:2007.15174*, 2020.

[20] Shreejit Poudyal, Alex Lindquist, Nate Smullen, Victoria York, Ali Lotfi, James Greene, and Mohammad Meysami. Unveiling wildfire dynamics: A bayesian county-specific analysis in california. *J*, 7(3):319–333, 2024.

[21] JoAnn E Manson, Shari S Bassuk, Julie E Buring, VITAL Research Group, et al. Principal results of the vitamin d and omega-3 trial (vital) and updated meta-analyses of relevant vitamin d trials. *The Journal of steroid biochemistry and molecular biology*, 198:105522, 2020.

[22] Brandon M Greenwell, Bradley C Boehmke, and B Gray. Variable importance plots-an introduction to the vip package. *R J.*, 12(1):343, 2020.

[23] Mohammad Meysami, Vijay Kumar, McKayah Pugh, Samuel Thomas Lowery, Shantanu Sur, Sumona Mondal, and James M Greene. Utilizing logistic regression to compare risk factors in disease modeling with imbalanced data: a case study in vitamin d and cancer incidence. *Frontiers in Oncology*, 13:1227842, 2023.

[24] Vijayalakshmi H Nagaraj, James M Greene, Anirvan M Sengupta, and Eduardo D Sontag. Translation inhibition and resource balance in the tx-tl cell-free gene expression system. *Synthetic Biology*, 2(1):ysx005, 2017.

[25] Mackenzie Dalton, Emmanuel Asante-Asamani, and James Greene. Evaluating the importance of glucagon in the insulin-glucose regulatory system: A mechanistic modeling approach. *bioRxiv*, pages 2024–07, 2024.

[26] Richard N Bergman. Toward physiological understanding of glucose tolerance: minimal-model approach. *Diabetes*, 38(12):1512–1527, 1989.

[27] James Greene, Orit Lavi, Michael M Gottesman, and Doron Levy. The impact of cell density and mutations in a model of multidrug resistance in solid tumors. *Bulletin of mathematical biology*, 76:627–653, 2014.

[28] Orit Lavi, James M Greene, Doron Levy, and Michael M Gottesman. The role of cell density and intratumoral heterogeneity in multidrug resistancemodeling the intratumoral heterogeneity in multidrug resistance. *Cancer research*, 73(24):7168–7175, 2013.

[29] Orit Lavi, James M Greene, Doron Levy, and Michael M Gottesman. Simplifying the complexity of resistance heterogeneity in metastasis. *Trends in molecular medicine*, 20(3):129–136, 2014.

[30] James M Greene, Doron Levy, King Leung Fung, Paloma S Souza, Michael M Gottesman, and Orit Lavi. Modeling intrinsic heterogeneity and growth of cancer cells. *Journal of theoretical biology*, 367:262–277, 2015.

[31] James M Greene, Doron Levy, Sylvia P Herrada, Michael M Gottesman, and Orit Lavi. Mathematical modeling reveals that changes to local cell density dynamically modulate baseline variations in cell growth and drug responsedynamic baseline variations and drug sensitivity. *Cancer Research*, 76(10):2882–2890, 2016.