A Stochastic, Molecular Model of Non-small Cell Lung Cancer Fate Decision

By

Austin Nicholes Oleskie

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

December 12, 2020

Nashville, Tennessee

Approved:

Matthew John Lang, Ph.D.

Vito Quaranta, M.D.

Matthew John Tyska, Ph.D.

Michael Shane Hutson, Ph.D.

John Peter Wikswo, Jr., Ph.D.

DEDICATION

To my family – a source of constant support in all I do.

ACKNOWLEDGEMENTS

During my time at Vanderbilt, I have had the privilege of working with many great researchers through the interdisciplinary Chemical and Physical Biology program. The program fosters a highly collaborative environment by creating a space to bring together people from diverse academic backgrounds. They have also allowed me the freedom to explore my research interests across multiple fields, and I finally landed in the laboratory of Dr. Erin Rericha.

I would like to thank my advisor Dr. Erin Rericha for the opportunity to work in her lab. I will be forever indebted to her for sparking my interest in complex systems, permanently changing how I view the world. Her mentorship has pushed me to become an independent researcher, find interesting questions, and communicate my findings to the broader field. Erin has the acute ability to illustrate complicated scientific problems in a clear, easy to understand manner. I strive to reach her aptitude in this skill.

Through my work, I had the pleasure of collaborating with Dr. Vito Quaranta and many members within his lab. Vito has provided endless support and critical feedback during the course of my graduate training. His guidance was instrumental in the completion of my time at Vanderbilt and I will be eternally thankful for his help. The care he shows to his colleagues and interest in all aspects of their lives is an attribute I hope to emulate. I would also like to thank the members of my committee for their guidance and critique of my work. Specifically, thank you to Dr. Matt Lang for serving as my chair and to Dr. Matt Tyska for the use of his laboratory space and equipment.

Thank you to all my friends and colleagues at Vanderbilt, they have made my time here enjoyable and unforgettable. In the Rericha lab, Hala, Jie, Abigail, and Jon have all offered encouragement and I will always appreciate the work environment they have created and the

friendship they have provided. To the members of the Quaranta lab, thank you for allowing me to use your research space and treating me as though I was also a member.

I am grateful to my parents, Nick and Shari, for their constant support throughout this process. Their love and reassuring words were instrumental to the completion of my graduate work. From an early age they have encouraged me to explore my interests, inspired me to always continue learning, and taught me the tenacity to pursue my dreams. Finally, I have immense gratitude to Meghan Bowler, who without her help, this work would not be complete. She has provided continuous encouragement through countless revisions and late nights working on this document. Her wit, intellect, and love are an endless source of respite and I am fortunate to have such a wonderful person in my life.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Description
AKT	Protein kinase B
BAX	BCL2-like protein 4
BCL2	B-cell lymphoma 2
BH3	BCL-like protein 11 (also BIM)
Cdc20	Cell-division cycle protein 20
CDH1	Cadherin-1 (also E-cadherin)
Cdk	Cyclin-dependent kinase
cMet	Hepatocyte growth factor receptor (also HGFR)
CT	Computed tomography
CXCR	CXC chemokine receptor
DIP	Drug-induced proliferation
DM	Direct method
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DS	Discrete subline
EBUS-TBNA	Endobronchial ultrasound-guided transbronchial needle aspiration
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor (also ErbB1 or Her1)
EMT	Epithelial to mesenchymal transition
ERK	Extracellular signal-regulated kinase (also MAPK)
ERL	Erlotinib
FAK	Focal adhesion kinase
FAP	Fibroblast activation protein
FDG	18-fluorodeoxyglucose
G0	Gap 0
G1	Gap 1
G2	Gap 2
GRB2	Growth factor receptor-bound protein 2
GSK3	Glycogen synthase kinase 3
IDO	Indoleamine 2,3-deoxygenase
IL	Interleukin

IMT	Intermitotic time
M	Mitosis
MAPK	Mitogen-activated protein kinase (also ERK)
MEK	Mitogen-activated protein kinase kinase (also MKK)
MMP	Matrix metalloproteinase
MRI	Magnetic resonance imaging
NSCLC	Non-small cell lung cancer
ODE	Ordinary differential equation
PA	Polyacrylamide
PBS	Phosphate buffered saline
PDE	Partial differential equation
PD-L1	Programmed cell death protein-ligand 1
PET	Positron emission tomography
PI3K	Phosphoinositide 3-kinase
PIP2	Phosphatidylinositol (4,5)-bisphosphate
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate
PTK	Protein tyrosine kinase
Raf	Rapidly accelerated fibrosarcoma
Ras	Rat sarcoma
Rb	Retinoblastoma protein
RNA	Ribonucleic acid
RTK	Receptor tyrosine kinase
S	DNA synthesis
Skp2	S-phase kinase-associated protein 2
SOS	Son of Sevenless
SSA	Stochastic simulation algorithm
STAT3	Signal transducer and activator of transcription 3
TKI	Tyrosine kinase inhibitor
VEGF	Vascular endothelial growth factor

CHAPTER I

1. Introduction

Lung cancer is the leading cause of all cancer-associated mortalities within the United States. In 2020, there will be an estimated 135,720 lung cancer related deaths [1]. The mortality rate has been steadily declining over recent years with the advent of preventative measures and new treatments. However, the survivorship for lung cancer remains low, ranging from 92% for patients in stage IA1 to under 10% for patients with stage IV [2]. There are several important factors dictating the onset, treatment, and recurrence of the disease. One especially salient feature, similar across many cancer types, is that cells within the same tumor respond heterogeneously to chemotherapy. Extensive research has focused on revealing the origin of this disparate treatment response; first, identifying the range of cell responses to chemotherapy, and second, uncovering the molecular discrepancies between divergently responding cells. The resulting outcome has been concurrent advances in both identification of molecular drivers of the disease and specific targeting of those drivers as new lines of therapy. Even with these novel tools, lung cancer often progresses with new mutations and changes within the cell signaling network. Discerning how and why these changes occur within the cell, and what role external cell stimuli perform in cell fate decisions, are key details in developing practical, patient-specific treatments.

When inspecting the sheer scale and interconnectivity of the cell signaling network, it quickly becomes apparent that making any type of network response prediction is challenging. Nonetheless, a thorough understanding of how the networks are controlled and how they evolve in response to dynamic stimuli is needed to develop novel treatments. Mathematical models of

biological systems have long been used to inform researchers on their complex behavior. Models provide a means of assessing the current knowledge regarding a system and extending that knowledge through testable predictions. Cancer, broadly speaking, is a disease caused by the misappropriation of cell signaling machinery. The resulting network changes manifest as altered cell behavior defined by seven widely recognized hallmarks of cancer: increased proliferative advantage, improved survival in response to stress, angiogenesis, metastasis, modified metabolic activities, a remodeled microenvironment, and an abated immune response [3]. Models are useful for interpreting how stimuli are integrated by the cellular signaling network into higher order phenomena and, as in the case of cancer, how these systems can go amiss.

This work examines the effect of cancer cell microenvironment on chemotherapeutic response with experimental studies and connects those results with changes in key regulators of the mitotic and apoptotic processes within the cell. The connection is made through the use of a molecular-level model. This project helps us move toward interpreting cell population behavior through the lens of single-cell actions at a molecular-level. In doing so, a bridge is forming between two biological scales, the molecular and cellular. By understanding the mathematical rules underpinning complex cellular responses, novel treatments can be developed more quickly because a more thorough perception of the underlying biology is achieved.

CHAPTER II

2. Background

Lung cancer is an immensely heterogenous disease, making identification and treatment particularly difficult. Due to its varied nature, patients present with diverse symptoms depending on specific disease site. Symptoms also usually appear in advanced stage disease, resulting in approximately 70% of lung cancer diagnoses occurring at regionally advanced or metastatic stages. Additionally, the link between tobacco use and lung cancer was obscured through a major tobacco industry campaign designed to fabricate scientific uncertainty. The high incidence, heterogenous and complex biology, and late diagnosis contribute to the poor prognosis of individuals with lung cancer. A better understanding of this complex biological environment can be achieved through mathematical modeling of the signaling networks and resulting decisions of these cancer cells. We use this chapter to provide necessary background information for the biology and mathematics underlying this project.

2.1. Lung cancer history

Lung cancer incidence over the past century provides a compelling narrative of clinical identification, discernment of the cause of disease, a public health battle, and treatment. At the beginning of the twentieth century, lung cancer incidence was low. There were few diagnosed cases and, surprisingly, even very few autopsy-identified cases. The disease had been established as a unique affliction in 1761 and relevant information was consolidated into the book *Primary Malignant Growths of the Lungs and Bronchi*, published in 1912. However, occurrences were still

rare. Alton Ochsner, a pioneering surgeon of the early twentieth century, describes observing an autopsy with the notion that the disease was so uncommon he may never witness another case. Seventeen years later, he observed his second. Unexpectedly, though, he encountered an additional eight cases within the following six months, marking a sharp rise in frequency. Searching for a causative element, Ochsner landed on increased cigarette consumption. Cigarette smoking had become fashionable through the First World War, and all the patients were heavy smokers that had starting smoking during the war. Even by his own evaluation, his speculation was correlative at best, though he had gathered some experimental evidence by placing tobacco on the surface of animals and observing malignant tumors.

Simultaneously, a German physician named Fritz Lickint was investigating associations between alcohol and tobacco consumption and health problems. In 1929, he published a paper statistically showing that lung cancer patients are likely to also be smokers. Due to his political affiliations as a social democrat, he lost his job at the hospital in 1934 and was later conscripted into military service in 1939. His work conceived terms and ideas still in use today. He coined the phrase “passive smoking” as the effects to those individuals not directly smoking cigarettes and described the physical and psychological dependencies of tobacco as addiction. It would not be until 1950 that the first paper decisively linking lung cancer to the increase of tobacco consumption was published. The authors surveyed lung cancer patients concerning their habits and environment and then compared the results to patients with non-lung cancers and without cancer. They found the highest correlative risk factor was smoking tobacco. Fortuitously, they had included the smoking history question because the final page of the questionnaire had space.

2.2. Lung cancer epidemiology

Lung cancer is the leading cause of cancer-related mortality in the United States, exceeding the next four most common cancers (breast, prostate, colorectal, and brain) combined. An estimated 228,820 lung and bronchus cancer patients will be diagnosed and 135,720 patients will die in 2020. This translates to a lung cancer related death every 3.9 minutes. Though the incidence and mortality rates have been leveling off for women and decreasing for men since 2000 in the United States, it still accounts for nearly a quarter of all cancer-related deaths [1]. Lung cancer is also the most common cancer worldwide with an incidence of 2.1 million and mortality of 1.8 million in 2018 [4]. The World Health Organization estimates that lung cancer mortality rates will continue to increase worldwide from increased tobacco use, particularly in Asia [5].

Lung cancer is more common in men than women across all ethnicities and races with the highest incidence in Black men (0.0712%) and the lowest in Hispanic women (0.0248%). Mortality rates follow similar trends. More than 90% of lung cancer diagnoses and deaths occur in patients aged 55 or over; the median age of diagnosis is 71 and the median age of death is 72. The five-year survival rate is highly dependent on the stage at which the cancer was diagnosed. For lung cancer still localized in the primary site, five-year survival is 59%. Once it has spread to lymph nodes, five-year survival decreases to 31.7%, and after metastasis to secondary sites, it is only 5.8% [6].

On top of it being a major health burden, lung cancer represents a substantial financial burden. In 2018, the National Institutes of Health reported a lung cancer care cost of $14.2 billion. In addition, they estimated the lost productivity cost in 2005 as $36.1 billion, which was higher than any other cancer [7].

2.3. Non-small cell lung cancer pathogenesis

Lung cancer is classified as small cell lung cancer and non-small cell lung cancer (NSCLC), which account for 15% and 85% of cases, respectively. The World Health Organization further divides NSCLCs into three categories: adenocarcinoma, squamous cell carcinoma, and large cell. Adenocarcinomas are the most common, accounting for 40% of all lung cancers. Adenocarcinomas originate in glandular cells, especially in smaller airways near the periphery of the lung [8]. Lung adenocarcinomas are more often found in women. Younger people (20-46 years old) and non-smokers are also more likely to have adenocarcinomas than other lung cancers [9]. The experimental model used for our investigations is a cell line derived from a human lung adenocarcinoma, hence our focus on NSCLCs.

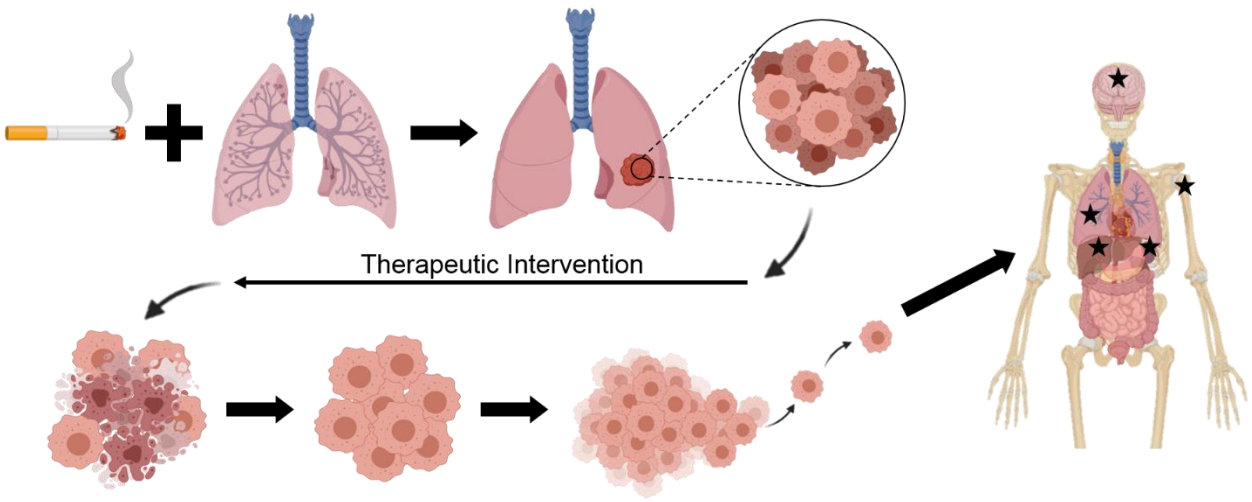


Figure 2.1 Typical progression of non-small cell lung cancer. Damage to the lung accumulates mutations in healthy cells, resulting in unrestricted proliferation into a heterogenous tumor. After therapeutic intervention, many cells undergo apoptosis but often, a resistant subpopulation survives and continues proliferating and metastasizing. For lung adenocarcinoma, the five most common sites of secondary tumors are, in order: bone, brain, lung, liver, and adrenal glands (indicated by black stars). Created with images from BioRender.

NSCLCs, like most cancers, are heterogeneous in their origin and their biological signature. However, the broadly shared paradigm of cancer pathogenesis (Figure 2.1) involves an initial injury, mutation, or perturbation to the genome of cells at the primary site, resulting in a dysregulated cell cycle. This leads to uncontrolled growth at the primary site, or the formation of a tumor. As the tumor grows, angiogenesis is increased to better transport the necessary oxygen and nutrients. However, many tumors eventually exceed a sustainable size and the core becomes necrotic. The malignant tumor cells are able to evade a normal immune response through a variety of mechanisms, including recruitment of immunosuppressive cells, downregulation of tumor antigens, and production of immunosuppressive factors. Individual cells or clusters of cells will migrate from the primary tumor and begin circulating through the bloodstream and lymphatics. Most of these escaped cells die but eventually, a few will extravasate and establish themselves in secondary sites to proliferate, creating secondary tumors. At this point, the cancer is metastatic. While these metastases are founded by a cell from the primary tumor, the primary tumor is heterogenous, so the secondary tumor has a related but unique biological profile. Most cancers spread to the lymph nodes first. Other common sites of metastatic lung cancer are bone, brain, lungs, adrenal glands, and liver. In a 2014 study of 4,316 patients with metastatic lung adenocarcinoma, 39% presented in bone, 38% in the nervous system, 22% in the respiratory system, 17% in the liver, and 8% in the adrenal glands [10], [11].

While each tumor is unique, there are some frequently altered genes in NSCLCs that provide opportunities for identification and therapeutic targeting. Some of these genes and their frequencies in lung adenocarcinoma and lung squamous cell carcinoma have been summarized in Table 2.1. They provide insight into signaling pathways involved in NSCLC pathogenesis and serve as potentially druggable targets.

Table 2.1 Selected genes with altered regulation in lung adenocarcinoma and lung squamous cell carcinoma [12], [13].

Gene	Broad Functions	Activated / Inactivated	Adenocarcinoma Frequency	Squamous Cell Carcinoma Frequency
TP53	Tumor suppressor	Inactivated	46%	90%
CDKN2A	Tumor suppressor; Senescence	Inactivated	43%	70%
KRAS	Proliferation; Differentiation	Activated	32%	3%
EGFR	Proliferation	Activated	27%	<9%
ALK	Proliferation; Differentiation	Activated	<8%	-
MET	Survival; Migration	Activated	7%	-
RB1	Tumor suppressor	Inactivated	7%	7%
AKT1	Proliferation; Differentiation; Survival; Apoptosis	Activated	1%	<1%
AKT2	Proliferation; Survival; Metabolism; Angiogenesis	Activated	-	4%
AKT3	Proliferation; Survival; Protein synthesis	Activated	-	16%

NSCLC can be caused by a range of genetic and environmental factors. The abnormalities to genes for epidermal growth factor receptor and anaplastic lymphoma kinase (*EGFR* and *ALK*, respectively) have identified them as driver oncogenes, primarily present in adenocarcinomas. These abnormalities vary inversely with smoking habit and appear dependent on race and ethnicity; white Europeans have a lower prevalence (~10%) than Asians (~60%). Less common aberrations have highlighted other potential targets, including *MET*, *HER2*, *KRAS*, and mutations in the protein kinase B (AKT) and extracellular signal-regulated kinase (ERK or MAPK) signaling pathways. There are heritable genetic factors as well, especially for lung adenocarcinoma, though they have not been clearly identified. Recent data suggest a germline mutation in *EGFR* in exon 20 ($EGFR^{T790M}$) and mutations in the transmembrane region of *HER2* may be oncogenic, though

rare. The primary environmental risk factor is tobacco smoking. Globally, 80% of men and 50% of women with lung cancer have smoked more than 100 cigarettes in their lifetime [14]. Interestingly, the incidence of lung cancer in non-smoking young women (19%) is rising in the United States [5]. Secondhand or passive smoking is also associated with an increased risk. Smoking marijuana and using electronic cigarettes may also increase risk, but their contribution is not yet quantified. Alcohol consumption has also been identified as a risk factor. Some environmental factors are, unfortunately, less controllable. Air pollution, specifically the presence of particulate matter <10μm in diameter, is associated with an increased risk of lung adenocarcinoma. The decay of uranium (^{238}U) in soil produces radon that, when inhaled, exposes lung tissue to decay products of the radon gas that irradiate the cells. Radon inhalation contributes to 9% of lung cancer mortality [14]. Asbestos is a carcinogen that contributes to lung cancer risk; insulation workers in the United States had a sevenfold higher risk of lung cancer than the general population. Additionally, when exposure to asbestos is combined with exposure to tobacco smoke, the factors act synergistically to dramatically increase risk [15]. Ionizing radiation (e.g. X-rays, alpha particles) and other toxins such as arsenic, chromium, nickel, and polycyclic aromatic hydrocarbons are also risk factors. Patients affected by pulmonary fibrosis or human immunodeficiency virus are at a higher risk for lung cancer [5].

2.4. Tumor microenvironment

As the tumor cells proliferate, they remodel their environment, optimizing it for growth. The extracellular matrix (ECM) is typically composed of collagens I, III, IV, and V (mostly collagen I), fibronectin, laminin, elastin, and osteonectin arranged in specific orientations to provide structure, a network that guides migration, and present ligands that stimulate signaling. In the case

of NSCLC, excess osteonectin can be synthesized to promote invasion and migration. Tumor migration is also supported by increased and organized collagen deposition. Conversely, a matrix composition with higher elastin content prevents cancer cell migration. The excessive growth of fibrotic tissue around the tumor, or desmoplasia, is characterized by thick bundles of collagen fibers and can inhibit tumor migration. This increased ECM density has also been reported to prevent T cells from infiltrating the tumor microenvironment [11], [16].

The tumor is initially surrounded by healthy stroma containing healthy immune cells (e.g. macrophages, neutrophils, lymphocytes, dendritic cells) and tissue maintenance cells (e.g. fibroblasts, myofibroblasts). As the tumor expands, the supply of oxygen and nutrients must be maintained, so the tumor promotes angiogenesis. Angiogenesis is one of the seven hallmarks of cancer. Malignant cells can directly release vascular endothelial growth factor (VEGF) or cooperate with macrophages to release angiogenic growth factors. Macrophages have been classified into two subpopulations: M1/pro-inflammatory macrophages secrete interleukin (IL)-2 and coordinate with natural killer cells and cytotoxic lymphocytes to attack tumor cells; M2/anti-inflammatory macrophages produce IL-10 and promote tumor growth and evasion of the immune system. When angiogenesis cannot provide enough oxygen to a rapidly growing tumor, the tumor becomes hypoxic. Hypoxia has been shown to induce VEGF transcription as well as promote cell migration and epithelial to mesenchymal transition (EMT). EMT is a process by which epithelial cells acquire migratory and invasive properties instrumental to escaping the tumor mass and entering a blood or lymph vessel to metastasize [11], [16].

Macrophages also cooperate with resident fibroblasts and myofibroblasts to remodel the ECM for easier migration. Fibroblasts associated with tumors more closely resemble myofibroblasts than healthy lung fibroblasts. For example, tumor-associated fibroblasts have higher expression of

fibroblast activation protein (FAP) and expression of matrix metalloproteinases (MMPs), proteases that degrade collagen and other ECM proteins. When FAP was reduced in a mouse model of lung cancer, collagen accumulation increased, myofibroblasts decreased, and blood vessel density decreased, all of which resulted in decreased tumor cell proliferation. These cancer-associated fibroblasts also secrete IL-6, which activates signal transducer and activator of transcription 3 (STAT3) to increase metastasis. The fibroblasts increase expression of α11β1 integrin, which leads to increased ECM stiffness. Cancer-associated fibroblasts also contribute to immune modulation of the tumor microenvironment. They can suppress T cell function through the expression of programmed cell death protein-ligand 1 (PD-L1) and PD-L2. They can also present antigens to $CD8^+$ T cells that result in T cell elimination, effectively evading the typical killing process. Tumor cells express indoleamine 2,3-dioxygenase (IDO), which catabolizes tryptophan to decrease cytotoxic T lymphocytes, accumulate T_{reg} cells, and cause proliferation of myeloid-derived suppressor cells that inhibit T cells. IDO can also increase metastasis, increase vascular density, and reprogram immune cells (e.g. inactivating CD8+ T cells). By these and more mechanisms, the cancer cells modify their physical, chemical, and biological environment to optimize it for tumor survival and growth [11], [16].

Interestingly, when lung cancer cells do metastasize, they tend to localize to consistent secondary organs. Investigation of brain metastasis has been shown to involve astrocytes, which secrete inflammatory cytokines and MMPs to create a more hospitable microenvironment for metastases. Lung adenocarcinoma metastasis to bone is mediated by lymphoid enhancer-binding factor 1, homeobox B9, osteoclast-derived CXC chemokine receptor 12 (CXCR12), CXCR4, activated ERK, and MMP9 signaling. Comparison of the metastatic site to primary site immune profiles show a more tumor-promoting microenvironment at the metastatic site (i.e. lower

CD8+:CD4+ T cell ratios and CD8+:CD68+ T cell ratios). Adenocarcinoma cells have also been shown to secrete IL-6 to mobilize inflammatory myeloid cells to basically seed a metastatic niche in the liver that the cancer cells can later colonize [16], [17].

Though some characteristics of the tumor microenvironment are highly similar, one striking difference between the lung and bone, where 39% of metastases occur, is stiffness. A healthy lung has an elastic modulus around 0.44-7.5 kPa and a cancerous lung has a stiffness of about 20-100 kPa but can be as high as 150 kPa [18]–[20]. Cortical bone is by far the stiffest tissue at 10-20 GPa [21], [22]. Other common sites of NSCLC metastasis have stiffnesses more similar to lung: healthy brain is approximately 0.5-7.3 kPa [23], [24] and healthy liver is 0.3-0.6 kPa [25]. The stiffness of a cell's environment is well known to influence its phenotype. For instance, quiescent fibroblasts can become activated myofibroblasts from exposure to increased substrate stiffness, as well as other mechanical forces (e.g. tension in wound healing) [19]. In a pathological positive feedback loop, myofibroblasts deposit more ECM, stiffening their environment and subsequently activating more myofibroblasts [16]. Stiffness can also affect metastasis. In a study of lung adenocarcinoma cells, stiffer substrates induced more collective, slow migration and cells on softer substrates had higher migration velocity and fewer cell contacts [20]. Integrin α11, previously noted to be increased by cancer-associated fibroblasts, is associated with collagen stiffness, which can induce the cancer-associated fibroblast phenotype in another positive feedback loop of myofibroblast activation. It also contributes to systemic metastasis of NSCLC [26]. While the mechanisms by which stroma stiffness modulates cancer progression are complex and not fully understood, the importance of stiffness is clear.

Collectively, these variations constitute intertumor and intratumor heterogeneity characteristic of NSCLC. This diversity between individuals necessitates a large number of specialized or

targeted diagnostic and therapeutic strategies for effective recovery. The diversity within tumors makes it difficult to effectively eradicate every cancer cell and prohibit recurrence. Another barrier to remission is the late stage of diagnosis. Since over 45% of lung cancer patients are diagnosed with distant metastases [6], it is nearly impossible to kill every heterogenous cancer cell. The only potentially curative treatment is surgical resection at an early stage, and a remarkable 30-55% of these cases still recur [27]. Better understanding the basic biology of the cells in the context of varied microenvironments and how they respond to therapies will help garner better lung cancer treatment results.

2.5. Non-small cell lung cancer diagnosis

NCSLC is often not diagnosed until an advanced stage of the disease when symptoms become more apparent. The most common symptom is a cough, which is present in 50-75% of cases. Other symptoms include hemoptysis (bleeding in the airway), chest pain, dyspnea (labored breathing), and weight loss. In fewer cases, venous congestion in the upper chest, shoulder and forearm (ulnar side) pain together, and Horner syndrome (characterized by a small pupil, drooping eyelid, and reduced sweating on the affected side) can be symptoms. Though these symptoms may indicate NSCLC, histology of a tumor biopsy is required to confirm the diagnosis [5], [14].

Various minimally invasive and non-invasive procedures are employed to assess the presence of lung cancer, identify its subtype, and characterize its stage. The first tests often performed are posterolateral and lateral chest radiography (i.e. X-rays), though these can be negative in up to 4% of lung cancer patients. Patients will have a contrast-enhanced computed tomography (CT) scan of the thorax and upper abdomen. CT scans highlight the location and shape of lesions and are important for evaluation of the mediastinum for proper staging. Magnetic resonance imaging

(MRI) is often performed for patients suspected of having brain metastases because of its ability to identify smaller lesions. There is some debate over the cost-effectiveness of brain imaging because CT and MRI scans of NSCLC patients' brains can yield up to 11% false-positive results due to other lesions, gliomas, or abscesses. Positron emission tomography (PET) imaging with radiotracer 18-fluorodeoxyglucose (FDG) can provide metabolic information about lesions. A maximum FDG standardized uptake value of 2.5 has been designated as the threshold between benign and malignant tumors, though some low-grade tumors yield false-negatives and some inflammatory lesions yield false-positives. FDG PET scans cannot define lesion size but they are a relatively low-risk method for gathering information on lesion malignancy [14].

In addition to imaging techniques, clinicians evaluate various tissues for evidence of cancerous cells. Sputum, or a mixture of saliva and mucus from the respiratory tract, can be collected and evaluated for the presence of abnormal cytology. This test has a sensitivity of 66% and a specificity of 99% for lung cancer. In cases where the lung tumor is centrally located, a bronchoscopy can be performed. A flexible tube with a light and camera on the end is inserted down the airway and has the capability to biopsy a nearby tumor. Bronchoscopy can be combined with ultrasonography and needle aspiration biopsy (EBUS-TBNA) to sample peribronchial or mediastinal masses and enlarged mediastinal or hilar lymph nodes. EBUS-TBNA is the preferred method for mediastinal staging as it is less invasive than the traditional mediastinoscopy and has an equivalent sensitivity and specificity. Transthoracic needle aspiration is a diagnostic technique used when a bronchoscopy has negative results from peripheral lesions. This has a 90% sensitivity except in tumors <2cm in diameter, where it is lower. During the procedure, a cytopathologist is on site to analyze the sample. Mediastinoscopy, parasternal mediastinoscopy, and extended cervical spine mediastinoscopy are more invasive, surgical procedures used when less invasive options are

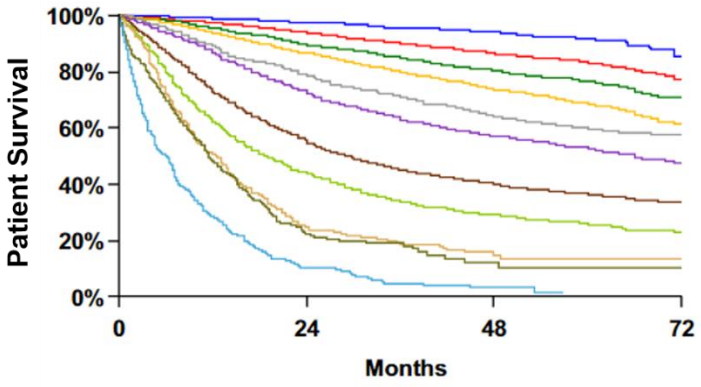
inconclusive. A thoracoscopy can be performed by inserting a thin, flexible scope inside the chest (pleural cavity) to view and potentially biopsy peripheral areas of the lung. In lung cancer patients with pleural effusion, or fluid accumulation around the lungs, the pleural fluid should be sampled via thoracentesis and undergo cytopathological evaluation [14]. Once samples have been obtained, pathologists analyze the histology and cytology of the tissue to determine the presence, type, and grade of cancerous cells.

The diagnostic process culminates in staging of the disease to track the cancer's development and to identify the best treatment options. The most widely used staging system is the TNM staging system, which evaluates the size and extent of the primary tumor (T), identifies nearby lymph nodes (N) with cancer, and describes whether the cancer has metastasized (M). The International Association for the Study of Lung Cancer organized an International Staging Project that utilized a database of 94,708 cases from 16 countries to standardize the T, N, and M classification definitions and how they translate to the Roman numeral stage groupings. Briefly, the tumor classifications range from Tis (in situ) to T4 (major axis >7cm), the lymph node classifications range from N0 (no regional lymph node metastasis) to N3 (metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph nodes), and the distant metastasis classifications range from M0 (no distant metastasis) to M1c (multiple extrathoracic metastases in one or more organs) [2]. The TMN classifications are organized into stage groupings that indicate the progression of the disease. Other factors may also be included in the staging such as tumor grade and biomarkers. Grade is a description of the cancer cytology and ranges from zero to four. If the cancer cells are differentiated, appearing similar to the surrounding healthy tissue, the tumor is low-grade. If the cells are poorly differentiated, the tumor is high-grade and in general, more likely to metastasize. Relevant biomarkers such as cell surface marker

expression or the presence of a genetic mutation will also be recorded with the staging since they can affect treatment plans [28].

Most cancers are staged from I to IV in order of severity. Stage 0 lung cancer is in situ, meaning the primary tumor has not grown into surrounding tissue. Stage I lung cancer describes a primary tumor that is <4cm in its greatest dimension and has not spread to any lymph nodes. Stage I is subdivided based on tumor size: IA1 is ≤1cm, IA2 is >1cm and ≤2cm, IA3 is >2cm and ≤3cm, and IB is >3cm and ≤4cm. Stage IIA describes a primary tumor >4cm and ≤5cm without lymph node involvement, and Stage IIB describes a tumor >5cm with no lymph node involvement or a tumor ≤5cm with metastasis to the ipsilateral peribronchial or ipsilateral hilar lymph nodes and intrapulmonary lymph nodes, ipsilateral mediastinal lymph nodes, or subcranial lymph nodes. Stage III lung cancer is subdivided into A, B, and C based on tumor size and which lymph nodes are affected. Stage IIIA describes a primary tumor >7cm with no lymph node involvement or a tumor >5cm with metastasis to the ipsilateral peribronchial or ipsilateral hilar lymph nodes and intrapulmonary lymph nodes, ipsilateral mediastinal lymph nodes, or subcranial lymph nodes. Stage IIIB describes primary tumors >5cm with metastasis in the ipsilateral mediastinal or subcranial lymph nodes or a tumor ≤5cm with metastasis in the contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph nodes. Stage IIIC describes tumors >5cm with metastasis in the contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph nodes. Stage IV lung cancer is by definition metastatic. Stage IVA is defined by separate tumor lesions in a contralateral lobe, a tumor with pleural or pericardial nodules, a malignant pleural or pericardial effusion, or a single extrathoracic metastasis. Stage IVB is defined by multiple extrathoracic metastases in one or more

organs. The International Association for the Study of Lung Cancer also summarized the survival for almost 95,000 lung cancer cases based on these stage groupings (Figure 2.2) [2].



Proposed	Events / N	MST	24 Month	60 Month
IA1	68 / 781	NR	97%	92%
IA2	505 / 3105	NR	94%	83%
IA3	546 / 2417	NR	90%	77%
IB	560 / 1928	NR	87%	68%
IIA	215 / 585	NR	79%	60%
IIB	605 / 1453	66.0	72%	53%
IIIA	2052 / 3200	29.3	55%	36%
IIIB	1551 / 2140	19.0	44%	26%
IIIC	831 / 986	12.6	24%	13%
IVA	336 / 484	11.5	23%	10%
IVB	328 / 398	6.0	10%	0%

Figure 2.2 Survival curves and median survival time (MST) for non-small cell lung cancer patients by diagnosis stage, compiled from 16 countries and diagnosed between 1999 and 2010. Shared with permission from [2].

The majority of NSCLC cases, approximately 70%, are diagnosed in a locally advanced or metastatic stage, either III or IV [29]. With earlier stage detection, prognosis significantly improves [2]. This highlights the importance of screening high-risk patients. The National Lung Screening Trial evaluated X-ray against CT screening in over 50,000 high-risk patients and found a benefit of 20% for lung cancer mortality. Therefore, the US Preventative Services Task Force recommends annual low-dose CT screening for patients 55 to 80 years old who have smoked within the last 15 years for a total of ≥30 pack-years (the equivalent number of cigarettes to one pack each day for one year, or 7,305 cigarettes). A randomized study implementing low-dose CT screening in the Netherlands and Belgium (NELSON trial) found a 26% reduction in lung cancer mortality for high-risk men at 10-year follow-up [5]. With approximately 8 million high-risk individuals in the United States, this translates to an ability to prevent over 24,000 lung cancer deaths [30]. Despite this, there is some debate over CT scanning because of a high false-positive

rate (24.2%), of a false-negative rate of 6%, 66% of tumors detected are already advanced (stage IIIB-IV), around 18% of cancers will be overdiagnosed, and the radiation used for the scans might increase the risk of lung cancer – it is believed to cause death in 0.024% of screened individuals [5], [14].

2.6. Non-small cell lung cancer treatment

Treatment of NSCLC depends on its stage. Stages 0 – IIA have no metastases or lymph node involvement, so they are typically surgically resected. Lobectomy, or removal of a single lung lobe, is prognostically superior to sublobar resection. A trial from the Lung Cancer Study Group observed a threefold increase in local cancer recurrence in patients given sublobar resections instead of lobectomies. Though this trial was performed before PET, survival still appears to be lower in populations with sublobar resection instead of lobectomy. There has been some evidence for the benefit of neoadjuvant chemotherapy to treat micrometastases and downstage the tumor to allow for complete resection in cases with limited or no lymph node involvement (N0-N1). After surgical resection, adjuvant chemotherapy is indicated for patients with stage II and IIIA disease. It has been shown to yield a 5.4% survival benefit at five years. Induction chemotherapy has shown a significant survival benefit for patients with stage IIB and IIIA lung cancer as well [5]. For stage III lung cancer, chemoradiation has been the standard of care for the past decade, but survival remains poor. More recently, the PACIFIC trial reported a benefit to inhibiting a PD-L1 with durvalumab. Progression-free survival with this targeted immunotherapy was 16.8 months compared to 5.6 months in controls, leading to FDA approval of durvalumab for unresectable stage III NSCLC. Oligometastatic NSCLC is defined by limited metastases to one or two organs that can be locally treated in combination with any systemic therapy. For example, brain metastases

are typically surgically resected or treated with precise radiation via stereotactic radiosurgery. Metastatic NSCLC requires systemic treatment. Historically, this has meant four to six cycles of platinum doublet chemotherapy (i.e. carboplatin or cisplatin with gemcitabine, vinorelbine, paclitaxel, or docetaxel). In nonsquamous NSCLC, treatment with pemetrexed and cisplatin had better tolerance and survival, so it is the current standard of care [5].

Immunotherapy is a relatively new advancement in the treatment of NSCLC and cancer in general. Broadly, immunotherapies strive to support the immune system in the recognition and killing of malignant cells. PD-L1, typically expressed in T cells, is often expressed by malignant cells as a defense mechanism against the immune system. This ligand can be blocked and a subset of treated NSCLC patients were alive five years after diagnosis, which is about five times longer than the typical case. There have been disparate results in the clinical trials currently run for a few PD-L1 inhibitors. Pembrolizumab treatment increased progression-free survival from 6 months to 10.3 months and overall survival from 14.2 months to 30 months. On the other hand, there was no significant difference between nivolumab and platinum doublet chemotherapy. Notably, the combination of platinum doublet chemotherapy with pembrolizumab improved treatment efficacy for patients with metastatic NSCLC, especially those whose tumors highly express PD-L1. Another immune checkpoint inhibitor expressed by T cells is cytotoxic T-lymphocyte-associated protein 4 (CTLA-4). The inhibition of PD-L1 and CTLA-4 has improved melanoma prognoses and is being evaluated as a lung cancer therapy. The combination of nivolumab with anti-CTLA-4 antibody ipilimumab improved progression-free survival over nivolumab alone. However, this improvement was correlated with PD-L1 expression and there were more higher-grade adverse events. Though they remain under evaluation, immunotherapies represent an important tool in the clinician’s armory [5].

Targeted therapy takes advantage of specific mutations often present in NSCLC and modifies the affected proteins' function. *EGFR* mutations are present in 15% of NSCLC patients in the United States and can be targeted with two approaches: monoclonal antibodies and tyrosine kinase inhibitors (TKIs). Erlotinib (ERL), gefitinib, afatinib, dacomitinib, and osimertinib are all *EGFR* TKIs. These have significantly improved progression-free survival in advanced NSCLC compared with platinum-based chemotherapy, but they had little effect on overall survival. Recently, osimertinib and dacomitinib have shown improved progression-free survival over first generation TKIs (i.e. erlotinib and gefitinib), qualifying them for frontline treatment of *EGFR*-mutated advanced NSCLC. Unfortunately, the disease progresses in most NSCLC patients after six to 12 months of *EGFR* targeted therapy. Though the etiology of this resistance is not fully understood, 40 to 60% of patients develop a secondary mutation in exon 20 ($EGFR^{T790M}$). Osimertinib specifically targets this mutation and has a 61% response rate in patients already treated with an *EGFR* TKI. Median progression-free survival on osimertinib is 10 months. Another factor in this resistance appears to be amplification of the *MET* oncogene in 5 to 20% of patients treated with erlotinib or gefitinib. Mutations in *PIK3CA* and *HER2* may also play a role in this resistance. Monoclonal antibodies against *EGFR* cetuximab and necitumumab have demonstrated only modest improvement in overall survival in combination with systemic chemotherapy [5].

Another TKI, crizotinib, was developed as a *c-MET* inhibitor and later discovered to have a beneficial effect in patients with translocations in the *ALK* or *ROS1* genes. Ceritinib and alectinib are *ALK* TKIs that have shown improved efficacy over combination chemotherapy. Alectinib had longer progression-free survival than crizotinib (34.8 months versus 10.9 months) and is better tolerated, so it is now approved for treating *ALK*-positive NSCLC. Crizotinib is, however, the preferred treatment for patients with *ROS1*-positive NSCLC, with a progression-free survival of

19.2 months. Bevacizumab is a monoclonal antibody against VEGF, which is instrumental in angiogenesis. In the ECOG trial, bevacizumab in combination with chemotherapy showed an improved progression-free survival over chemotherapy alone (12.3 months versus 10.3 months). However, due to some bleeding adverse effects, bevacizumab is only recommended for nonsquamous NSCLC. Other therapies against VEGF like aflibercept and ramucirumab demonstrated improved progression-free survival as well. V-raf murine sarcoma viral oncogene homolog B1 (*BRAF)* mutations exist in 2% of NSCLCs and can be targeted with vemurafenib or combination dabrafenib and trametinib, which have a median progression-free survival of 7.3 months and 9.7 months, respectively. Larotrectinib is FDA approved for *NTRK*-positive NSCLC without an acquired resistance mutation and no alternative treatment. *KRAS* mutations are present in nearly a quarter of NSCLCs but molecular agents under review have not yet shown benefits over current therapeutics [5].

There are no curative treatments for metastatic disease, which is when most cases are diagnosed. Even for those in stage I receiving surgical resection, the survival rate is 60-80% [31]. Fortunately, there are currently over 3,000 clinical trials for NSCLC in progress [32]. Better understanding the underlying biology of NSCLC progression and the mechanisms of resistance will lead to more effective treatment strategies.

2.7. Mathematical models of protein signaling networks

We begin with the basic assumptions that a cell is composed of many interacting parts and that, through the interaction of those parts with each other and their surroundings, a cell can perform a particular function including division, movement, apoptosis or many other behaviors. A major question within biology is how do the interactions of cellular components give rise to higher order

events? This has specific application in how cells respond to external stimuli and integrate what they sense into a behavior response. The first step in answering this question involves identifying the cellular components that are necessary to a particular phenomenon. Cells sense their surroundings through proteins associated with the cell surface and transfer the information internally through modification of proteins along a signaling cascade. Modifications to proteins include phosphorylation, dephosphorylation, association, and glycosylation, amongst others. Estimates for the number of protein species in humans ranges from 0.62-6.13 million, including variants and modified proteins [33]. Many of the protein species are capable of interacting with each other, further complicating the system. When observing the network of proteins, we recognize that the collective activity cannot be understood by summing the activity of the individual proteins, defined as a nonlinear system. The system can also typically form self-organized network structures, said to be 'emergent,' that require additional description beyond the component functions. Additionally, the fact that the proteins can modify each other acts as a form of coupling within the system. These particular features together classify the cellular signaling network as a 'complex system'.

Complex systems benefit from an interpretation using a mathematical modeling approach, helping to provide a formalism to describe the systems behavior. Models provide a means to simplify, or coarse-grain, the system to better understand the connection of the system and behaviors that can emerge. They have been used extensively in the past to describe complex systems that often contain behavior nonlinear in nature and, therefore, difficult to predict [34], [35]. Accordingly, models are integral for understanding the complex systems that are presented in cell signaling networks composed of a large collection of tightly coupled signaling molecules. Keeping track of all these components and their interactions and, then, making predictions about

their behavior quickly becomes insurmountable without computational assistance. Cells communicate with and integrate signals from their surrounding environment through their signaling networks. The behavior of a cell is based on the proper response to an external signal [36]. The information is encoded through the network by multiple chronological steps that include protein binding, post-translational modification, and compartmentalization [37]. Many past studies have focused on identifying the immediate upstream and downstream components of signaling cascades, resulting in large, highly connected maps of known interactions. What has emerged from these previous studies is a picture composed of coupled modules, or motifs, that perform a unique signaling behavior. That is, though the specific components may be different, the behavior of the system will be based on how the components are connected, known as the system topology.

Discerning the topology, or architecture of a system, represents an integral first step into understanding how that system operates. Once the connections of the system are defined, the components can be organized into modules based on function. In this section, we illustrate two common modules found in cell signaling networks: a bistable switch and an oscillating circuit. Both will be later used to build a molecular model of NSCLC fate decision. The behaviors of a system that arise from the interaction of its components are called its emergent properties. Often, it is difficult to predict emergent properties from the topology alone. Additionally, because the emergent behavior is the result of the interaction of all components, properties of the system are not the result of a single element. This particular feature makes a reductionist approach of understanding difficult. Instead, it is necessary to view the system as a whole.

Furthermore, a model, by its definition, is used to represent a process. When a model is constructed, the approach and the modeling methods employed make inherent assumptions about that system. Some approaches are applicable in certain instances but not in others. We review two

common methods for representing the protein signaling cascades present in the cell. It is important to remember that the model representation is a proxy for the underlying physical machinations that are occurring. In the descriptions, we will cover the important assumptions made by the model, in what applications the model can be useful, and limitations of the model.

2.7.1. Differential equation implementations

The most common and simplest approach in modeling protein signaling dynamics uses the kinetic equation framework. Here, the kinetics, or change, in a molecule concentration is equal to the rate that the molecule is being created or destroyed:

$$\frac{dC}{dt} = (generation) - (consumption) \tag{2.1}$$

The terms within the equation can represent different cellular processes if a set of particular assumptions are made. Each term, generation and consumption, can be symbolized as a constant (e.g. synthesis), first order reaction (e.g. degradation), or nonlinear component (e.g. Michaelis-Menten kinetics) [38]. We make use of all three options in our eventual model implementation. The expression ordinary differential equation (ODE) refers to the system only evolving in time, not space. The use of differential equations as a basis for the model classifies the approach as deterministic, which is the first of the two formalisms commonly used to model protein signaling networks. One of the major assumptions underpinning this model is that the system is a large, well-stirred reactor, functionally meaning that any reaction can occur readily. There are no limitations to molecule count or location. The use of ODEs as a model system is well characterized and although analytical solutions for a system of coupled equations may not exist, numerical methods are well established to solve the system. Additionally, solving large systems of ODEs is relatively computationally inexpensive in comparison to other approaches, allowing for large ranges of

parameter sets to be explored. As models become more complex, both in size and detail, the computational cost increases, resulting in a greater time to reach a solution. Models generally trade detail to reach a solution more quickly. Properly defining the model involves finding an appropriate balance between detail and speed.

We know that cells do not exist as a single, large compartment as the ODE model assumes. Instead, they contain a multitude of physically separated regions. Taking this fact into consideration when building a model, either a partial differential equation (PDE) or compartmental model is employed. A PDE approach allows for evolution in both time and space, and the dynamics of the system are described by the reaction-diffusion equation:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x} + R \tag{2.2}$$

where C is molecule concentration, t is time, D is the diffusion coefficient, v is the convective velocity, x is the spatial variable, and R is the rate of production and depletion. Equation 2.2 states that the change in concentration of a molecule at a particular point is dictated by diffusion (the first term), active transport (the second term), and the reaction rates (third term). While this approach more closely realizes actual cellular processes, it comes with the disadvantage of a large increase in parameters to a system that may already contain a multitude.

The difference between the two approaches highlights an important question when building a model – what is the degree of simplification that should be used? That is, what is the appropriate amount of abstraction of the system that can occur while still maintaining the phenomena of interest. The details that are included in a model (i.e. what steps can be simplified and what steps cannot) can offer important insight into the dynamics and highlight significant variables of the system.

2.7.2. Michaelis-Menten kinetics

Here, we pay particular attention to Michaelis-Menten kinetics, as a thorough understanding of enzyme behavior is helpful in building more complex protein signaling networks. Historically, the proposed model was developed to describe the enzymatic reaction mechanism of invertase [39]. The process can be represented as:

$$\mathrm{E} + \mathrm{S} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P \tag{2.3}$$

where an enzyme E binds to a substrate S to form complex ES at a forward rate k_f and a reverse rate k_r. The complex then releases the product P and original enzyme E at the catalytic rate constant k_{cat}.

To derive the rate at which the product P is formed, specific assumptions about the system must be applied. The first is the application of the law of mass action, stating that the rate of a chemical reaction is proportional to the product of the reactants' concentrations, resulting in a system of four coupled non-linear ODEs that describe the evolution of all four components of the process. The differential equations are as follows:

$$\frac{d[E]}{dt} = -k_f[E][S] + k_r[ES] + k_{cat}[ES] \tag{2.4}$$

$$\frac{d[S]}{dt} = -k_f[E][S] + k_r[ES] \tag{2.5}$$

$$\frac{d[ES]}{dt} = k_f[E][S] - k_r[ES] - k_{cat}[ES] \tag{2.6}$$

$$\frac{d[P]}{dt} = k_{cat}[ES] \tag{2.7}$$

At the beginning, the substrate binds to enzyme forming a reversible complex. The enzyme then catalyzes a reaction, forming a final product and freeing the enzyme once again. An example time course is provided in Figure 2.3.

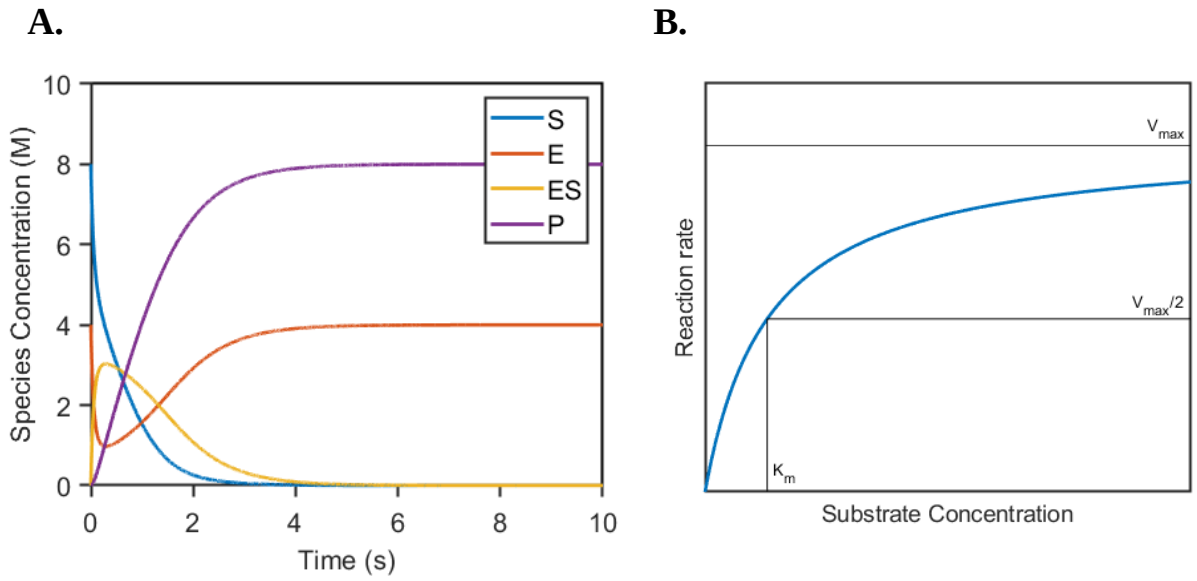


Figure 2.3 Michaelis-Menten system kinetics. **(A)** Species time course for the evolution of the Michaelis-Menten model. [S] represents substrate, [E] represents enzyme, [ES] represents substrate-enzyme complex, and [P] represents product concentrations. **(B)** Michaelis-Menten saturation curve. The rate of reaction is dictated by substrate concentration. At low substrate concentrations, the change in reaction time is nearly linear (first-order reaction), at high substrate concentration the rate reaches the asymptotic value of V_{max} (zero-order reaction). Rate parameters and initial conditions are found in Appendix A.

Using the derived ODEs, we can describe the rate of product formation as:

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{2.8}$$

where [P] represents the concentration of the product and [S], the concentration of the substrate. V_{max} is the maximum velocity of the reaction and is the product of the total enzyme concentration [E] and k_{cat}. K_m is the dissociation constant for the enzyme-substrate complex and is the ratio of the reverse rate k_r and the forward rate k_f, or k_r / k_f.

There are distinct system behaviors depending on the relative size between K_m and [S]. When substrate concentration is low, [S] << K_m, the reaction rate varies approximately linearly with substrate concentration, resulting in first-order kinetics. However, when substrate concentration is much greater than K_m, [S] >> K_m, the reaction rate becomes independent of substrate concentration, giving rise to zero-order kinetics. At this point, the reaction is referred to as saturated since all of the available enzyme is bound to substrate. Overall, the relationship between substrate concentration and reaction rate is hyperbolic.

2.7.3. Zero-order ultrasensitivity

In 1981, Goldbeter and Koshland published seminal work showing that the response curve to a given input within a signaling network can be sigmoidal [40]. The term they used to describe this phenomenon is called zero-order ultrasensitivity. Zero-order refers to the regime where the reaction occurs, at enzyme saturation. Ultrasensitivity is in reference to the response seen; the response can largely vary with small changes in input strength. Ultrasensitive activity has been implicated in signaling cascades, bistable systems, and oscillating systems [41]–[43]. Ultrasensitivity is an important concept in molecular signaling as it allows for filtering of low-level stimulus and a large difference in response for small changes in input. Functionally, within the signaling networks, ultrasensitivity enables the formation of more complex systems, including bistable switches and oscillators.

To examine ultrasensitivity in practice, we can look at a protein that can be activated by phosphorylation and inactivated through dephosphorylation. Assuming mass action kinetics, the rate of phosphorylation for a protein X is:

$$\frac{dpX}{dt} = k_1[kinase](X_{tot} - \text{pX}) - k_{-1}[phosphatase](pX) \tag{2.9}$$

where k_1 is the rate constant for phosphorylation, X_{tot} is total protein, pX is the phosphorylated fraction of X, and k_{-1} is the rate of dephosphorylation. From here, we can solve for the phosphorylated fraction of X as a function of the concentration of kinase:

$$\frac{pX}{X_{tot}} = \frac{[kinase]^n}{{K_m}^n + [kinase]^n} \tag{2.10}$$

We can describe the change by including n, an exponential factor transforming the above relationship into the Hill function. Ultrasensitive responses are generally sigmoidal, with flat responses at low and high inputs and a steep response between the two. The curve is well approximated by the Hill function and provides a means of quantifying ultrasensitivity through the Hill exponent n. In this form, when n = 1, the response is Michaelian and when n > 1, the response is ultrasensitive, with the system becoming more switch-like as n increases (Figure 2.4).

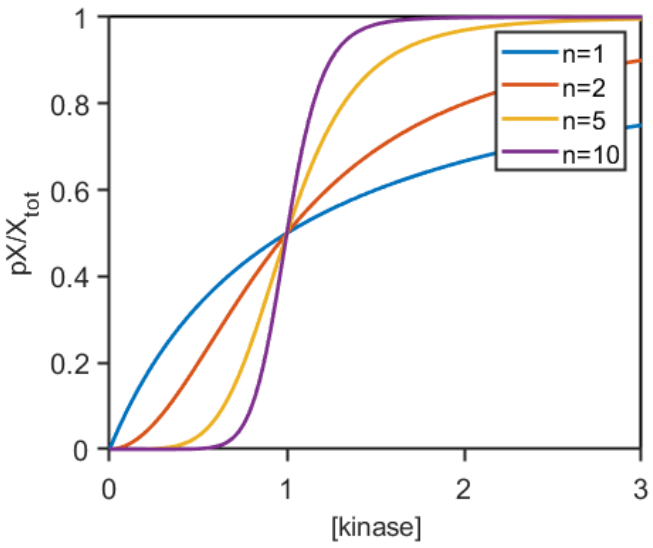


Figure 2.4 Hill function response curves. The input (kinase concentration) and output (fraction of phosphorylated protein X) relationship for Hill functions with exponent 1, 2, 5, and 10. As the exponent increases in value, the response becomes more ultrasensitive and switch-like in nature.

2.7.4. Stochastic implementations

Many biological systems exist at low concentration of species (e.g. proteins, RNA, DNA, signaling chemicals). At this level, small changes in the overall number of signaling molecules can have large and profound effects on the behavior of the system. When this is the case, using a deterministic model is not appropriate to study system dynamics. Instead, a formalism that treats the time evolution of the system as a type of random walk process can be used. This approach uses a chemical master equation that describes a set of phenomenological first-order differential equations.

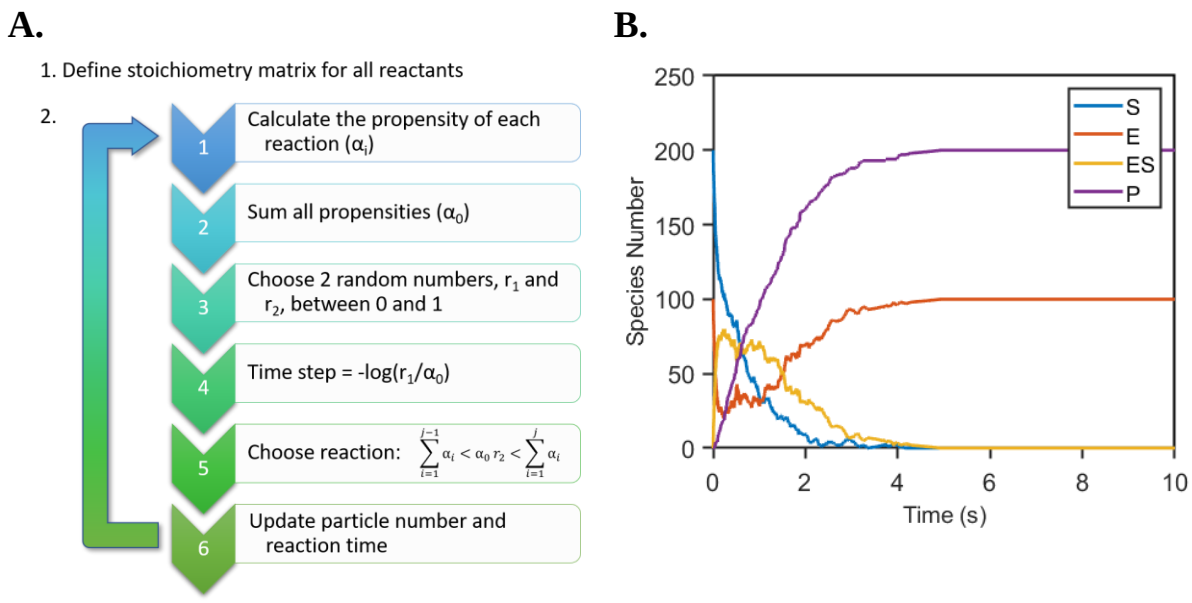


Figure 2.5 Schematic of the Gillespie algorithm and representative stochastic Michaelis-Menten time course. **(A)** The direct method (DM) procedure developed by Gillespie for describing the evolution of a chemical system. **(B)** The stochastic implementation of Gillespie's DM procedure.

For our stochastic approach, we used the Gillespie algorithm, or stochastic simulation algorithm (SSA). This method, while the slowest of the stochastic techniques, offers an exact solution by explicitly simulating every reaction. The approach uses a Monte Carlo procedure to

reproduce evolution of a chemical system. The algorithm, called the direct method (DM), was initially described by Gillespie and is presented as a schematic in Figure 2.5A [44]. Briefly, it consists of the following steps: a stoichiometry matrix for all reactants is constructed, each reaction probability is calculated given the state (number of species at the current time) of the system and the probabilities are summed, the time step and a specific reaction are chosen based on two random numbers, the species number and current time are updated based on the previous step, and the procedure is repeated. When the number of reactions or species becomes large, the DM becomes too slow to be practical. As an example, we present the Michaelis-Menten equation, solved stochastically. We first define the three reactions from the Michaelis-Menten equation as follows:

$$1.\,\mathrm{S} + \mathrm{E} \xrightarrow{k_f} ES \tag{2.11}$$

$$2.\,ES \xrightarrow{k_r} E + S \tag{2.12}$$

$$3.\,ES \xrightarrow{k_{cat}} E + P \tag{2.13}$$

The resulting stoichiometry matrix is:

Table 2.2 Stochastic solution for system of Michaelis-Menten equation.

	E	**S**	**ES**	**P**
1.	-1	-1	1	0
2.	1	1	-1	0
3.	1	0	-1	1

All concentrations in the original ODE implementation must be converted to molecule number. The conversion is achieved by defining a system size for the stochastic implementation. As the system size increases, the time course evolution approaches the ODE solution. Conversely, as the

system size decreases, stochastic effects are more pronounced. An example of one stochastic solution for the Michaelis-Menten equation is shown in Figure 2.5B.

2.7.5. Rb-E2F switch

Turning to our first molecular motif, we examine a bistable switch. The idea and characteristics of a bistable switch are well known within the fields of mathematics and engineering, and beginning to see extensive use in biology as a control mechanism of molecular signaling. The restriction point (R point) marks the time that a cell commits to entering the cell cycle. From an ideal standpoint, once the cell chooses to enter the cell cycle, it should fully commit to completing its own division. At a molecular level, this is achieved through the implementation of a bistable switch (Figure 2.6A). The system exists in three regimes. The first, at low input signal, results in a low output response. As input signal is increased, a threshold is reached, and the output quickly increases, marking the second regime. The third region, a bistable central area, exists between the two extremes and results from hysteresis of the system, meaning the initial state of the entire system dictates the observed response (Figure 2.6B). The behavior of a bistable switch captures two important biological purposes: a high threshold for activation from a system with low input signal and continued signaling once the switch is activated. The first property results in effective signal from noise filtering, a necessary function in noisy biological systems. The second ensures that once the circuit has been activated, the response remains high and the cell can complete mitosis. The bistable mechanism marking the entry into mitosis was suggested theoretically before being confirmed experimentally, highlighting the value in modeling approaches [45]–[47].

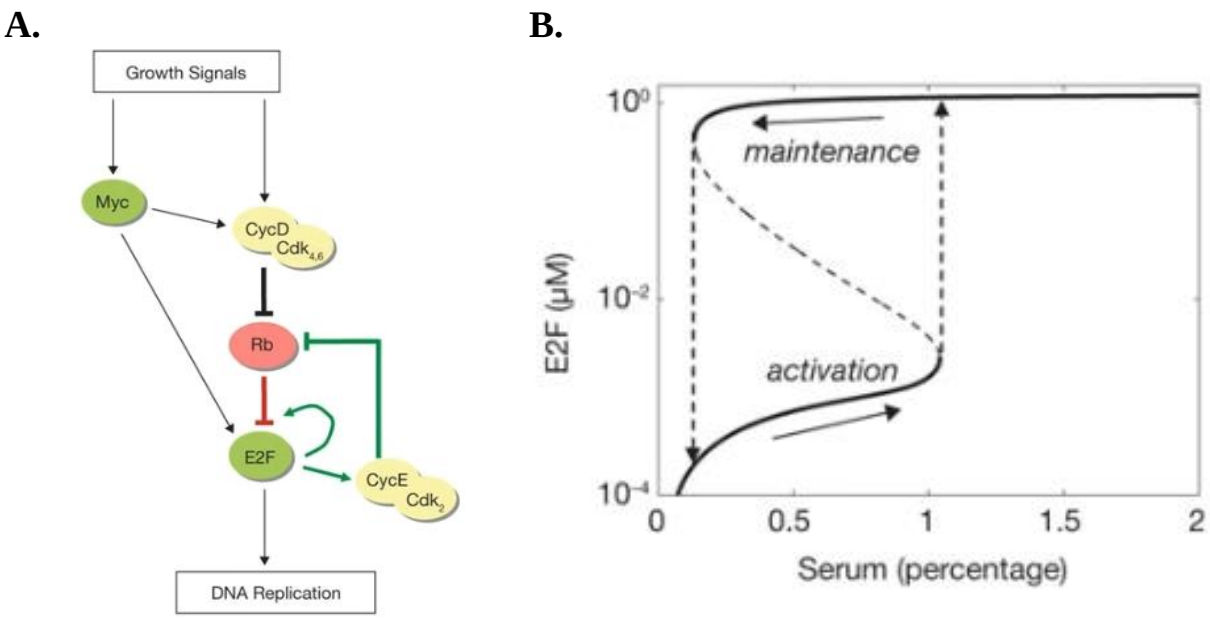


Figure 2.6 Model of the bistable Rb-E2F switch. **(A)** Schematic representation showing the topology of the Rb-E2F switch model. The system behavior is the result of two positive feedback loops, between CycE/Cdk2 and Rb and E2F autoactivation. **(B)** E2F concentration as a function of serum treatment showing hysteresis. Figure is adapted from Yao *et al.* with journal permission [48].

Outlining a molecular example, the retinoblastoma protein (Rb)-E2F mechanism forms the gateway for the cellular decision to enter mitosis. In response to external stimuli (growth signals), the transcriptional factor E2F is activated. E2F is involved in the regulation of proteins controlling DNA replication and cell-cycle progression, and is critical in the control of cell proliferation [49]–[51]. Mechanistically, the growth signals lead to increased activation of Myc and cyclin D/cyclin-dependent kinase (Cdk) 4-6 complex. Myc additionally activates cyclin D/Cdk 4-6 and induces E2F transcription [52]. At low signal, the cellular quiescent state, E2F is bound to Rb, which results in preventing the cell cycle. Cyclin D phosphorylates Rb and the free E2F is able to activate cyclin E/Cdk 2. There exist two positive feedback loops within the network: cyclin E phosphorylates Rb enhancing Rb inactivation from cyclin D, and E2F activates its own transcription (Figure 2.6A).

2.7.6. Oscillating circuit

A defining property of a cell is the ability to self-replicate. Mammalian cell division is carried out through the successive activation of four different cyclins with their dependent kinases [53]. Historically, the cell cycle was divided into four phases: G1 (gap 1 phase) corresponding to the interval between division and DNA replication, S (synthesis phase) when DNA replication occurs, G2 (gap 2 phase) during which the cell continues to grow and the molecular machinery for mitosis is produced, and M (mitosis phase) when the cell finally divides [54]. Each of the different phases is characterized by the distinct dynamics of the separate cyclin/Cdk complexes. Cyclin D/Cdk4-6 and cyclin E/Cdk2 characterize G1 and drive the G1/S transition (the Rb-E2F bistable switch highlighted in the previous section). Cyclin A/Cdk2 facilitates the progression of S phase and the S/G2 transition, and cyclin B/Cdk1 enables the G2/M transition. A full detailed model of the mammalian cell cycle proposed by Goldbeter and colleagues is driven by the Cdk network and is composed of 39 variables [55].

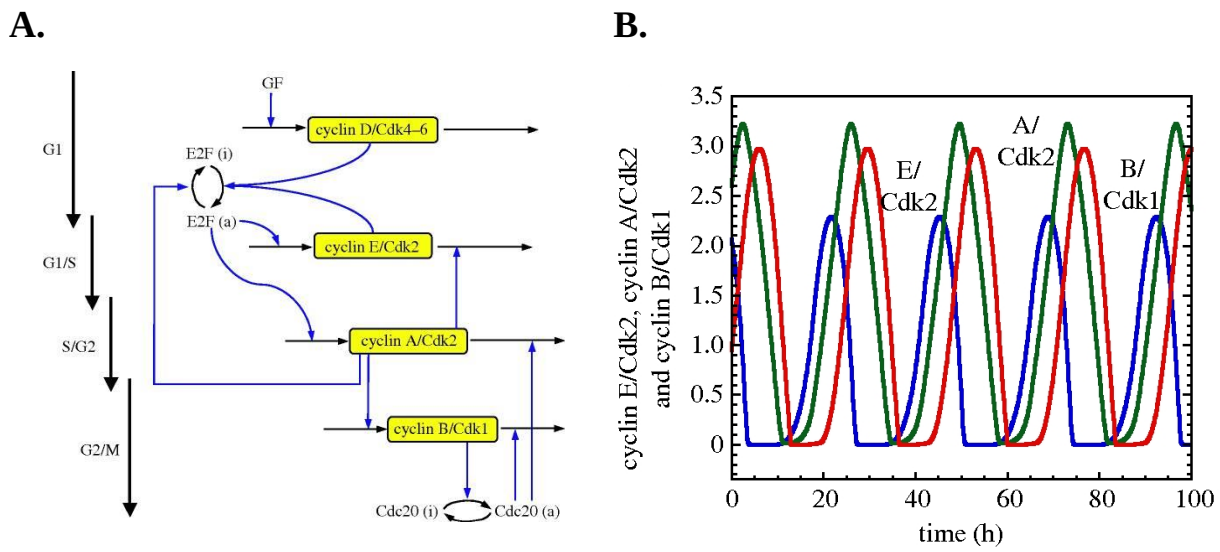


Figure 2.7 Schematic for minimal cell cycle model. **(A)** Network topology for the Goldbeter skeleton model of mitosis is driven by progressive activation of the four cyclin proteins and their respective cyclin dependent kinases. Note the absence of the protein Rb, integral to the bistable switch. **(B)** Evolution of the five coupled ODE system at high growth factor input. Oscillations are phenomenologically the same as the large 39 component model. Figure is adapted from Goldbeter et al. with journal permission [56].

Next, Goldbeter undertook defining the minimum molecular architecture necessary for maintaining sustained oscillations (Figure 2.7) [56]. The reduced form is more tractable for quicker modeling implementations and for a greater search of parameter space. Additionally, it helps define the critical topology necessary for mitotic behavior. The model itself is composed of only the four cyclins in complex with their respective Cdks, the transcription factor E2F, and cell-division cycle protein 20 homolog (Cdc20). The model shows the same phenomenological behavior as its larger relative. A key component missing in this model iteration is the protein Rb. The skeleton model still oscillates at high input signal and reaches a constant steady state at low signal but the sharp, switch-like transition between the two is missing.

The wiring of the cell cycle contains multiple oscillatory circuits driven by negative feedback loops. Even the simple five variable model contains three such circuits [55], [56]. Feedback loops are one of the necessary components for a circuit to oscillate in addition to ultrasensitivity and

enough delay within the feedback loop. The existence of multiple, coupled oscillators can give rise to complex oscillations and chaos [57], [58]. The behavior of the system when the multiple oscillators are tightly coupled results in self-organized, sequential activation. However, as the coupling is decreased, complex periodic oscillations can develop. The existence of decoupled oscillations has been proposed as an explanation for endoreplication in yeast and eukaryotic cells [59], [60].

2.8. Experiment-model iteration process

The understanding of such complex systems as a cell, organ, or a body requires not only the identification of components, but also discernment of how the components interact. The interactions give rise to a host of phenomena difficult to calculate from a list of the parts. Developing a deep comprehension of emergent phenomena requires the concurrent development of a mathematical framework to describe such a system. The study of systems biology, whose definition has itself been somewhat emergent, attempts to define how such phenomena come about from biological organization. The field relies heavily on the quantitative measurement of biological systems, modeling those systems, and then development of theory. Each part of the approach is interrelated, and the theory development occurs from feedback from both ‘wet’ and ‘dry’ experiments.

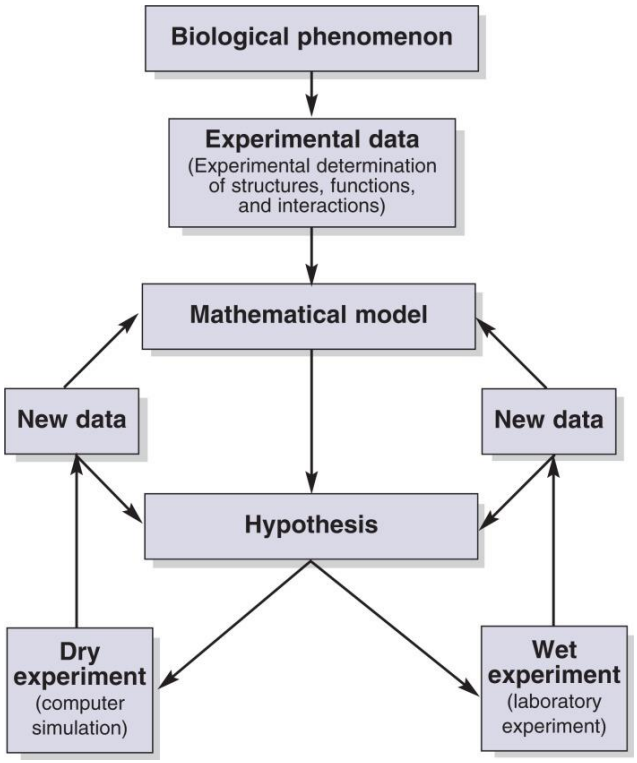


Figure 2.8 Schematic outlining the general procedure for understanding a biological phenomenon using a systems biology approach. Used with permission from [61].

We employ a systems biology approach to examine the heterogenous response seen in NSCLC to targeted drug therapy using the schematic shown in Figure 2.8 as an outline for our approach. The biological phenomenon we are examining is the fate decision of an individual cell. At the single cell level, whether NSCLC continues to develop or not is based on the proportion of cancer cells proliferating, dying, and entering quiescence. The following chapters follow the progression of this outline. In Chapter III, we collect the necessary molecular components from the literature to build a model of cell fate decision. In Chapter IV, we take quantitative single-cell results from defined perturbations to the system. Finally, in Chapter V, using the molecular model generated in Chapter III and the wet experimental data from Chapter IV, we compare experimental and model

results, highlighting discrepancies between the two and extending the model for future experimental testing.

CHAPTER III

3. Building a molecular model of cell fate decision

We constructed a literature-based, minimal protein network model with the goal of capturing the heterogeneous cell response seen in our experiments. Overall, this work examines whether our current understanding of the cellular signaling network can be used to evaluate and predict tumor growth and treatment response. Whether a tumor will grow or shrink ultimately depends on the distribution of decisions by the population of cells to either continue dividing, die, or enter quiescence. To gain a better understanding of how heterogeneity in cell fate decision is linked to the underlying protein signaling network, we propose a simplified model of the process. When designing our model, we highlight several important features. We have chosen to focus on the fate decision process that directs a cell into mitosis or apoptosis. However, the entirety of the process is complex and not fully known. First, it was necessary for our model to be able to integrate external signals through receptor tyrosine kinases (RTKs) and integrin. Both are canonical signaling axes that drive AKT and ERK activation. Second, we needed to include modules to incorporate cell death and division. Third, we needed to return population dynamics and capture stochastic features to link protein network signaling to cell population. The model is purposefully simple to focus on network structure and behavior. As a starting point, we have simplified the signaling upstream of mitosis, and instead implement a mapping of cell surface molecules to AKT and ERK activation levels. In our simplified model, AKT and ERK signaling set the threshold for fate decision.

In this chapter, we present an overview of the signaling pathway as a review for our model, highlighting important caveats of our approach and highlighting the complexity of the cellular signaling network. Building on the design principles laid out by Goldbeter, we use a similar framework to model the cellular decision between division, apoptosis, and quiescence [56]. The model contains three modules: the mitosis module positioned around the cyclin network, a simple apoptosis switch, and a component that integrates the signal transduction network from external cues (Figure 3.1). This system allows us to ask: What does the topology of the signaling network tell us? What types of cellular phenomena can be captured with a simple model? Where do we need to add more complexity? Finally, we can probe the model to gain insight into expected cancer cell behavior and potential mechanisms thereof.

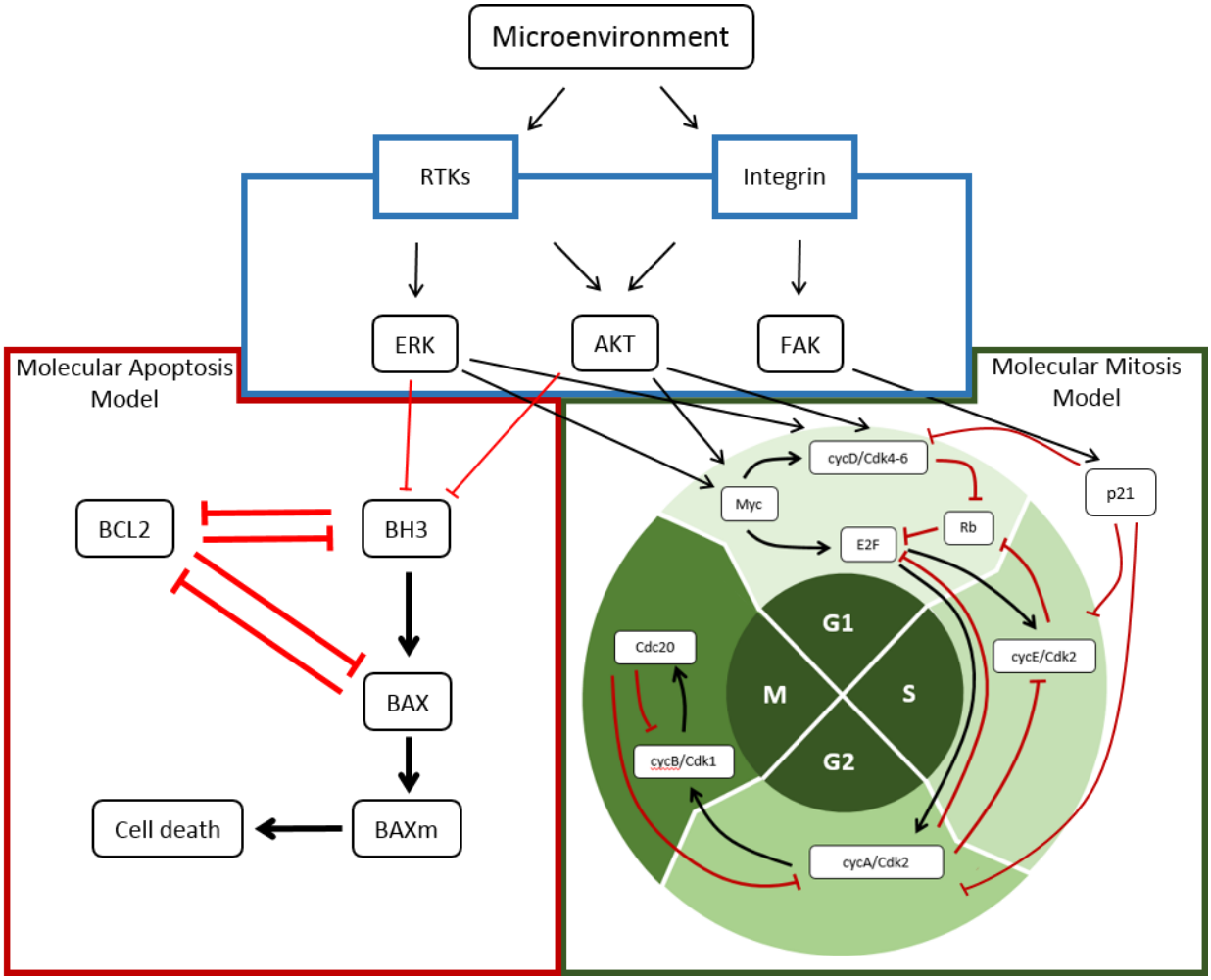


Figure 3.1 Schematic representation of the cell fate model. Microenvironment cues are integrated through RTKs and integrins via molecular signaling cascades. AKT, ERK, and FAK represent major axes within the signaling network and serve as the inputs to our molecular apoptosis model and molecular mitosis model.

3.1. Signal transduction network

Our goal is to connect cell surface signaling to cell fate decision. The cell surface contains a multitude of proteins that allow the cell to sense its surroundings. The microenvironment has been shown to have a significant effect, both in our studies and prior work, on cell behavior and tumor treatment response. However, modeling the full transduction network is outside the scope of this project and the capability of our stochastic technique. Understanding the signal transduction

network and its feedback has its own set of challenges. In the current study, we have chosen to map RTK expression and signaling directly to AKT and ERK activation levels. However, we offer a brief review of the pathway from receptor to AKT and ERK activation. Regardless of mechanism, both integrins and RTKs signal through AKT and ERK. These signaling axes are major signaling nodes within the cell. The idea is part of an emerging theory of so-called "bow-tie" signaling that represents a characteristic feature of robust, evolvable systems [62]. The amount of plasticity afforded to the signaling network from RTKs and integrins is considerable.

3.1.1. Receptor tyrosine kinases (RTKs)

RTKs represent an important class of proteins within oncogene-addicted cancers. The receptor family is often overexpressed or possesses activating mutations, leading to disease progression. This cancer subclass derives its name from the apparent reliance on oncogene signaling for continued survival. In this study, we focus on two of the receptors from the RTK family: hepatocyte growth factor receptor (HGFR or cMet) and EGFR (or ErbB1 or Her1). RTKs have similar molecular architecture with a single transmembrane helix that connects an extracellular ligand binding domain to an intracellular, conserved protein tyrosine kinase (PTK) domain [63]. Upon ligand binding, the receptor can either homo- or heterodimerize with subsequent tyrosine autophosphorylation. Both cMet and EGFR have shown the capability to signal through the same pathways, ultimately activating both AKT and ERK.

The ErbB family consists of four different members that are structurally similar. We focus on EGFR as it is often mutated in NSCLC. The cytoplasmic domain has at minimum 20 tyrosine residues capable of phosphorylation, seven of which are autophosphorylation sites [64], [65]. Our model cell line, PC9, harbors an exon-19 deletion within the PTK domain, making EGFR

constitutively active. This particular mutation occurs in approximately half of all EGFR-mutated lung tumors [66].

Aberrant cMet expression and activation has been correlated with poor prognosis in cancer. Furthermore, we have shown that elevated cMet expression correlates with the erlotinib-resistant phenotype characterized by an increased drug-induced proliferation (DIP) rate. It has also been shown that MET amplification leads to gefitinib resistance by heterodimerizing and activating ErbB3, a member of the ErbB family mentioned previously [67].

Upon activation, cMet recruits and activates many downstream signaling molecules at the plasma membrane. A key adaptor is growth factor receptor-bound protein 2 (GRB2), which facilitates the binding of the nucleotide exchange factor Son of Sevenless (SOS) [68]. SOS functions as a guanine nucleotide exchange factor that acts on the proto-oncogene c-rat sarcoma viral oncogene homolog (Ras) family of small GTPases. Ras operates by binding rapidly accelerated fibrosarcoma (Raf) kinases and translocating Raf to the cell membrane for activation. The activation of Raf represents the first step of the integral MAPK cellular signaling module. The module is composed of MAPK kinase kinase (MAP3K, Raf), MAPK kinase (MEK, MKK), and ERK.

The AKT signaling cascade is also initiated by RTK activation. Upon RTK activation, the phosphoinositide 3-kinase (PI3K) family can bind to the intracellular domain of the RTK, resulting in the conversion of phosphatidylinositol (4,5)-bisphosphate (PIP2) to phosphatidylinositol (3,4,5)-trisphosphate (PIP3) at the plasma membrane. PIP3 then recruits AKT to the membrane, where it is activated.

The signaling cascade between receptor and AKT and ERK contains a multitude of molecules with significant crosstalk. We highlight canonical pathways from our selected RTKs to AKT and

ERK but have chosen to measure AKT and ERK activation directly for our model. We pay particular attention to AKT and ERK signaling, as they represent a point of convergence within the transduction pathway.

3.1.2. Integrins

Integrins are class of receptors connecting the ECM with the cellular cytoskeleton. They provide the cell with important mechanical cues controlling a wide range of cellular responses. We focus on their role in cell cycle regulation. Cellular proliferation is highly altered in cancer. Knowledge of how integrin signaling alters the proliferative state of a cell population will help develop new treatment strategies.

Integrins are transmembrane, heterodimeric receptors. Mammals possess 18 α and 8 β subunits that can form 24 heterodimers. There are several mechanisms as to how integrins promote cellular proliferation. Integrins signal through the same canonical pathways as RTKs, both the PI3K/AKT and Ras/Raf/MEK/ERK axes. Integrins are also capable of regulating RTK expression at the transcriptional level [69]. Within the cell, integrins facilitate the formation of cell adhesion complexes consisting of numerous adaptor proteins (e.g. paxillin, vinculin, talin). Focal adhesion kinase (FAK) is an additional protein recruited to the focal adhesion complex. We use FAK phosphorylation as a means of measuring cellular signaling response to stiffness. Once activated, FAK can engage PI3K, thus activating the AKT signaling pathway. Additionally, FAK is capable of signaling through the ERK axis by forming a complex with GRB2 and activating the Ras family of GTPases [70].

3.2. Mitosis

Activated AKT and ERK are responsible for initiating the cell cycle through the activation of several transcription factors. The scheme and approach for our mitosis module is inspired by the Goldbeter skeleton model, which itself is based off a larger, more comprehensive model that Goldbeter and colleagues have developed [55], [56]. The objective of the model is to reduce the number of variables and parameters necessary to capture the complex phenomena of cell division. Making the model simpler allows for quicker simulation and the ability to search through parameter space. Also, it allows us to begin interpreting the minimal signaling circuitry necessary to capture complex cellular phenomena. The model is constructed around the Cdks that regulate the progression of the mammalian cell cycle through the consecutive phases of mitosis. The cell cycle can be divided into four phases marked by the sequential activation of specific cyclins and their associated kinases.

3.2.1. G1 phase

In the first phase, ERK directly phosphorylates cMyc. The myc family of genes are transcription factors that result in the expression of genes that drive the cell cycle (down regulation of p21 and upregulation of cyclins). They also play an important role in other vital cell processes such as apoptosis, differentiation, and growth. In addition, ERK promotes the formation AP1, which in turn, stimulates the production of cyclin D. In our model, both of these steps are modeled directly with no intermediates. Cyclin D forms an inactive complex with p21/p27, inhibiting the cell cycle. p21/p27 is negatively regulated by FAK activity.

Additionally, activated AKT induces both Myc and cyclin D activation through several mechanisms. First, AKT phosphorylates glycogen synthase kinase 3 (GSK3) preventing the

degradation of cyclin D [71]. AKT also stimulates mTOR signaling, increasing the translation of cyclin D by activating ribosomal protein S6K [72]. The transcription activity of Myc is promoted by AKT-mediated degradation of Mad1. Mad1 is an inhibitor of Myc [73].

Central to the G1/S transition is a bistable switch comprised of Rb and the transcription factor E2F. E2F is another set of transcription factors responsible for both activating and suppressing transcription. In this work, we focus on their role as activators. E2F upregulates the synthesis of cyclins D, E, and A, helping to drive the transition through the cell cycle. Conversely, hypophosphorylated Rb inhibits entry into the cell cycle by binding to E2F, thereby preventing the expression of necessary cell cycle-promoting factors. Rb exists in three forms: hypophosphorylated, monophosphorylated, and hyperphosphorylated states. Each of these serve a specific function within the cell. Our simple model only takes into account two forms of Rb: the non- and phosphorylated (combined mono- and hyper-) states. In our model, cyclin D/Cdk4-6 complex phosphorylates Rb, inhibiting it from binding to E2F. Cyclin E/Cdk2 complex further phosphorylates and inactivates Rb. Biologically, the cyclin D complex is responsible for monophosphorylating Rb and the cyclin E complex hyperphosphorlates Rb, progressing the cell cycle. We model this by increasing the Rb phosphorylation rate from the cyclin E complex in comparison to the cyclin D complex.

3.2.2. S phase

Once cyclin E levels – stimulated by E2F – reach a threshold, the cell moves from G1 into S. Cyclin E/Cdk2 can form an inactive complex with p21/p27. A positive feedback loop controls further activation of the cyclin E/Cdk2 complex through Cdc25. We do not explicitly include Cdc25 in this model, but the effect of its positive feedback is included via self-activation of cyclin

E. Through S-phase kinase-associated protein 2 (Skp2), cyclin E phosphorylates p21/p27, marking it for degradation. Skp2 also degrades cyclin E, helping the cell cycle to progress. Inactivation of cyclin E can occur through Wee1 kinase. In our model, the cyclin E complex is inactivated by the emergence of the cyclin A/Cdk2 complex.

3.2.3. G2 phase

The S/G2 transition occurs with the synthesis of active cyclin A/Cdk2 complex, driven by E2F. Cyclin A supports the transition through S phase by both inactivating E2F and inhibiting cadherin-1 (CDH1). In turn, CDH1 degrades Skp2, which stimulates the degradation of cyclin E. The reduction in cyclin E activity facilitates the proper progression through S phase. Further control is exhibited through p21/p27 binding, sequestering cyclin A out of the available pool. It has been shown that Wee1 also inactivates cyclin A, similar to cyclin E. We do not include this step. Instead, cyclin A is ultimately inactivated by Cdc20, the final component of our mitotic model.

3.2.4. M phase

Cyclin A triggers the activation of the cyclin B/Cdk1 complex, signaling the progression of the cell cycle into M phase. In the full model, cyclin B is degraded by CDH1, with constant synthesis of cyclin B. When CDH1 is inhibited by cyclin A, cyclin B levels can then increase unhindered. In the simple model, cyclin A activates cyclin B directly. Subsequently, cyclin B activates Cdc20, which, in turn, negatively regulates cyclin B and cyclin A. Negative feedback pathways acting through cyclin B and Cdc20 down regulate cyclin A and help set the proper progression through division.

3.3. Apoptosis

We have included an intrinsic apoptosis module to account for cell death in our model by adapting a portion of the Tyson model for apoptosis [74]. ERK, in addition to AKT through forkhead box O3 (FOXO3), negatively regulates BCL-2-like protein 11 (BH3 or BIM). When AKT and ERK are low, BH3 levels increase. B-cell lymphoma 2 (BCL2) initially binds to and removes active BH3 from the activating pool. However, if AKT and ERK remain low, BH3 levels continue to rise, leading to the activation of BCL2-like protein 4 (BAX). Within the cell, BH3 instigates a conformational change within BAX, allowing it to insert into the mitochondrial membrane. BAX oligomerizes with itself to form a pore, allowing the release of cytochrome c and apoptotic factors. Once this occurs, the apoptosis decision within the cell is final. Additionally, BAX negatively regulates BCL2, preventing BCL2 from binding and inhibiting BAX. The relationship allows for a bistable response to low AKT and ERK inputs. In the model, BAX is inactivated at a basal level to prevent accumulation and premature cell death.

Table 3.1 Model equations.

$$\frac{dSkp2}{dt} = s * \left[V_{sskp2} + \left(V_{s2skp2} * FAK\right) - \left(V_{dskp2} * \left(\frac{Skp2}{K_{dskp2} + Skp2} \right) * \left(\frac{Cdh1}{K_{cdh1} + Cdh1} \right) \right) - \left(k_{ddskp2} * Skp2\right) \right] \tag{3.1}$$

$$\frac{dp21}{dt} = s * \left[V_{s1p21} + \left(v_{s2p21} * E2F * \left(\frac{K_{i13}}{K_{i13} + Rb} \right) * \left(\frac{K_{i14}}{K_{i14} + pRb} \right) \right) - (k_{c1} * CycD * p21) + (k_{c2} * CycDp21) - (k_{c3} * CycE * p21) + (k_{c4} * CycEp21) - (k_{c5} * CycA * p21) + (k_{c6} * CycAp21) - (k_{c7} * CycB * p21) + (k_{c8} * CycBp21) - \left(CycE * \left(\frac{V_{1p21} * p21}{K_{1p21} + p21} \right) \right) + \left(\frac{V_{2p21} * p21p}{K_{2p21} + p21} \right) - \left(FAK * \frac{V_{Fp} * p21}{K_{Fp} + p21} \right) - (k_{ddp21} * p21) \right] \tag{3.2}$$

$$\frac{dp21p}{dt} = s * \left[\left(CycE * \left(\frac{V_{1p21} * p21}{K_{1p21} + p21} \right) \right) - \left(\frac{V_{2p21} * p21p}{K_{2p21} + p21} \right) + \left(FAK * \frac{V_{Fp} * p21}{K_{Fp} + p21} \right) - \left(V_{dp21p} * \left(\frac{Skp2}{K_{dp21p} + Skp2} \right) * \left(\frac{p21p}{K_{dp21skp2} + p21p} \right) \right) - (k_{ddp21} * p21p) \right] \tag{3.3}$$

$$\frac{dCdh1}{dt} = s * \left[\left(\left(\frac{V_{1cdh1} * (Cdh1_{tot} - Cdh1)}{K_{1cdh1} + (Cdh1_{tot} - Cdh1)} \right) \right) - \left((CycA + CycB) * \left(\frac{V_{2cdh1} * Cdh1}{K_{2cdh1} + Cdh1} \right) \right) \right] \tag{3.4}$$

$$\frac{dMyc}{dt} = s * \left[\left(\frac{k_M * AKT}{K_S + AKT} \right) + \left(\frac{k_{EMY} * ERK}{K_{MY} + ERK} \right) - (d_M * Myc) \right] \tag{3.5}$$

$$\frac{dCycD}{dt} = s * \left[\left(\frac{k_{CDS} * AKT}{K_S + AKT} \right) - (k_{c1} * CycD * p21) + (k_{c2} * CycDp21) + \left(\frac{k_{ED} * ERK}{K_{MY} + ERK} \right) + \left(\frac{k_{CDF} * FAK}{K_{MY} + FAK} \right) + \left(\frac{k_{CD} * Myc}{K_M + Myc} \right) - (d_{CD} * CycD) \right] \tag{3.6}$$

$$\frac{dCycDp21}{dt} = s * [(k_{c1} * CycD * p21) - (k_{c2} * CycDp21)] \tag{3.7}$$

$$\frac{dRb}{dt} = s * \left[k_R + \left(\frac{k_{DP} * pRb}{K_{RP} + pRb} \right) - (k_{RE} * Rb * E2F) - \left(CycD * \left(\frac{k_{P1} * Rb}{K_{CD} + Rb} \right) \right) - \left(CycE * \left(\frac{k_{P2} * Rb}{K_{CE} + Rb} \right) \right) - (d_R * Rb) \right] \tag{3.8}$$

$$\frac{dpRb}{dt} = s * \left[\left(CycD * \left(\frac{k_{P1} * Rb}{K_{CD} + Rb} \right) \right) + \left(CycE * \left(\frac{k_{P2} * Rb}{K_{CE} + Rb} \right) \right) + \left(CycD * \left(\frac{k_{P1} * RbE2F}{K_{CD} + RbE2F} \right) \right) + \left(CycE * \left(\frac{k_{P2} * RbE2F}{K_{CE} + RbE2F} \right) \right) - \left(\frac{k_{DP} * pRb}{K_{RP} + pRb} \right) - (d_{RP} * pRb) \right] \tag{3.9}$$

$$\frac{dE2F}{dt} = s * \left[\left(k_E * \left(\frac{Myc}{K_M + Myc} \right) * \left(\frac{E2F}{K_E + E2F} \right) \right) + \left(\frac{k_b * Myc}{K_M + Myc} \right) + \left(CycD * \left(\frac{k_{P1} * RbE2F}{K_{CD} + RbE2F} \right) \right) + \left(CycE * \left(\frac{k_{P2} * RbE2F}{K_{CE} + RbE2F} \right) \right) - \left(CycA * \left(\frac{V_{2E2F} * E2F}{K_{2E2F} + E2F} \right) \right) - (k_{RE} * Rb * E2F) - (d_E * E2F) \right] \tag{3.10}$$

$$\frac{dRbE2F}{dt} = s * \left[(k_{RE} * Rb * E2F) - \left(CycD * \left(\frac{k_{P1} * RbE2F}{K_{CD} + RbE2F} \right) \right) - \left(CycE * \left(\frac{k_{P2} * RbE2F}{K_{CE} + RbE2F} \right) \right) - (d_{RE} * RbE2F) \right] \tag{3.11}$$

$$\frac{dCycE}{dt} = s * \left[\left(E2F * (a + b1 * CycE) * \left(\frac{V_{1Me} * (CycE_{tot} - CycEp21 - CycE)}{K_{1Me} + (CycE_{tot} - CycEp21 - CycE)} \right) \right) - \left(CycA * \left(\frac{V_{2Me} * CycE}{K_{2Me} + CycE} \right) \right) - (k_{c3} * CycE * p21) + (k_{c4} * CycEp21) \right] \tag{3.12}$$

$$\frac{dCycEp21}{dt} = s * [(k_{c3} * CycE * p21) - (k_{c4} * CycEp21)] \tag{3.13}$$

$$\frac{dCycA}{dt} = s * \left[\left(E2F * \left(\frac{V_{1Ma} * (CycA_{tot} - CycAp21 - CycA)}{K_{1Ma} + (CycA_{tot} - CycAp21 - CycA)}\right)\right) - \left(Cdc20 * \left(\frac{V_{2Ma} * CycA}{K_{2Ma} + CycA}\right)\right) - (k_{c5} * CycA * p21) + (k_{c6} * CycAp21)\right] \quad (3.14)$$

$$\frac{dCycAp21}{dt} = s * [(k_{c5} * CycA * p21) - (k_{c6} * CycAp21)] \quad (3.15)$$

$$\frac{dCycB}{dt} = s * \left[\left(CycA * (b + b2 * CycB) * \left(\frac{K_{ie}}{K_{ie} + CycE}\right) * \left(\frac{V_{1Mb} * (CycB_{tot} - CycBp21 - CycB)}{K_{1Mb} + (CycB_{tot} - CycBp21 - CycB)}\right)\right) - \left(Cdc20 * \left(\frac{K_{ib}}{K_{ib} + CycE}\right) * \left(\frac{V_{2Mb} * CycB}{K_{2Mb} + CycB}\right)\right) - (k_{c7} * CycB * p21) + (k_{c8} * CycBp21)\right] \quad (3.16)$$

$$\frac{dCycBp21}{dt} = s * [(k_{c7} * CycB * p21) - (k_{c8} * CycBp21)] \quad (3.17)$$

$$\frac{dBAX}{dt} = m * \left[-(k_{f1} + k_{f2} * BAX) * BH3 + kb * (BAXm + BAXmBCL2)\right] \quad (3.18)$$

$$\frac{dBAXmBCL2}{dt} = m * [kasbaxmbcl2 * BAXm * BCL2 - kdsbaxmbcl2 * BAXmBCL2 - kb * BAXm] \quad (3.19)$$

$$\frac{dBH3}{dt} = m * [k1sbh3 + k2sbh3 * Stress - kdbh3 * BH3 - kasbh3bcl2 * BH3 * BCL2 + kdsbh3bcl2 * BH3BCL2] \quad (3.20)$$

$$\frac{dBH3BCL2}{dt} = m * [kasbh3bcl2 * BH3 * BCL2 - kdsbh3bcl2 * BH3BCL2 - kdbh3 * BH3BCL2] \quad (3.21)$$

CHAPTER IV

4. NSCLC cell response to variations in microenvironment

How cells interpret their chemical and mechanical environments to make fate decisions remains an open question at the interface of multiple disciplines. Cancer, generally speaking, is a misappropriation of the cellular machinery responsible for those decisions. Targeted therapy of EGFR has been a successful initial treatment for patients with NSCLC [75]. The disease, however, will frequently progress as the tumors become resistant [76], [77]. One cause of resistance is linked to heterogeneity of cell fates in response to drug treatment. Underlying fate decision, and the resulting population response, is the complex signaling network of the cell. To understand the dynamics of a drug response, it is necessary to determine how an external stimulus is integrated into fate determination and how perturbations in the signaling network alter the decision process.

Previous work has focused on identifying key mutational drivers and quantifying the dynamic changes to cell fate that occur upon treatment, describing the flux between populations of dividing, dying, and quiescent cells [78]–[81]. These changes directly translate into the variation of overall population number and intermitotic time (IMT). Ultimately, cellular regulation of the mitotic and apoptotic pathways dictates overall disease response and subsequently, patient outcome. As the signaling networks of the cell are vast and complex, computational modeling of the cellular circuitry has emerged as a valuable tool offering insight into the behavior of these networks and suggesting novel mechanisms of system control.

We use this chapter to first quantitatively define the single cell response to changes in the microenvironment. The changes are presented as the addition of the small molecule EGFR-

inhibitor erlotinib and by varying the substrate stiffness. We know from past studies that response to erlotinib is heterogenous even for cells that are genetically similar [82]. This type of resistance is distinct from other forms involving receptor amplification, secondary mutations, or activation of separate pathways [67], [83], [84]. Previous work in the Quaranta laboratory has established distinct subclones of the parental PC9 NSCLC cell line based upon the response to erlotinib treatment. Briefly, cells are plated at single-cell density and allowed to grow into small colonies. A perturbation (e.g. drug treatment) is presented to the cells and their change in population is followed through quantity doublings providing a linear fit in log2 scale (Figure 4.1A). What is observed in the case of PC9 is a distribution of response rates (Figures 4.1B and 4.1C). We select the sublines that show the greatest and least response to erlotinib therapy, DS9 and DS8, respectively, to compare further perturbation response.

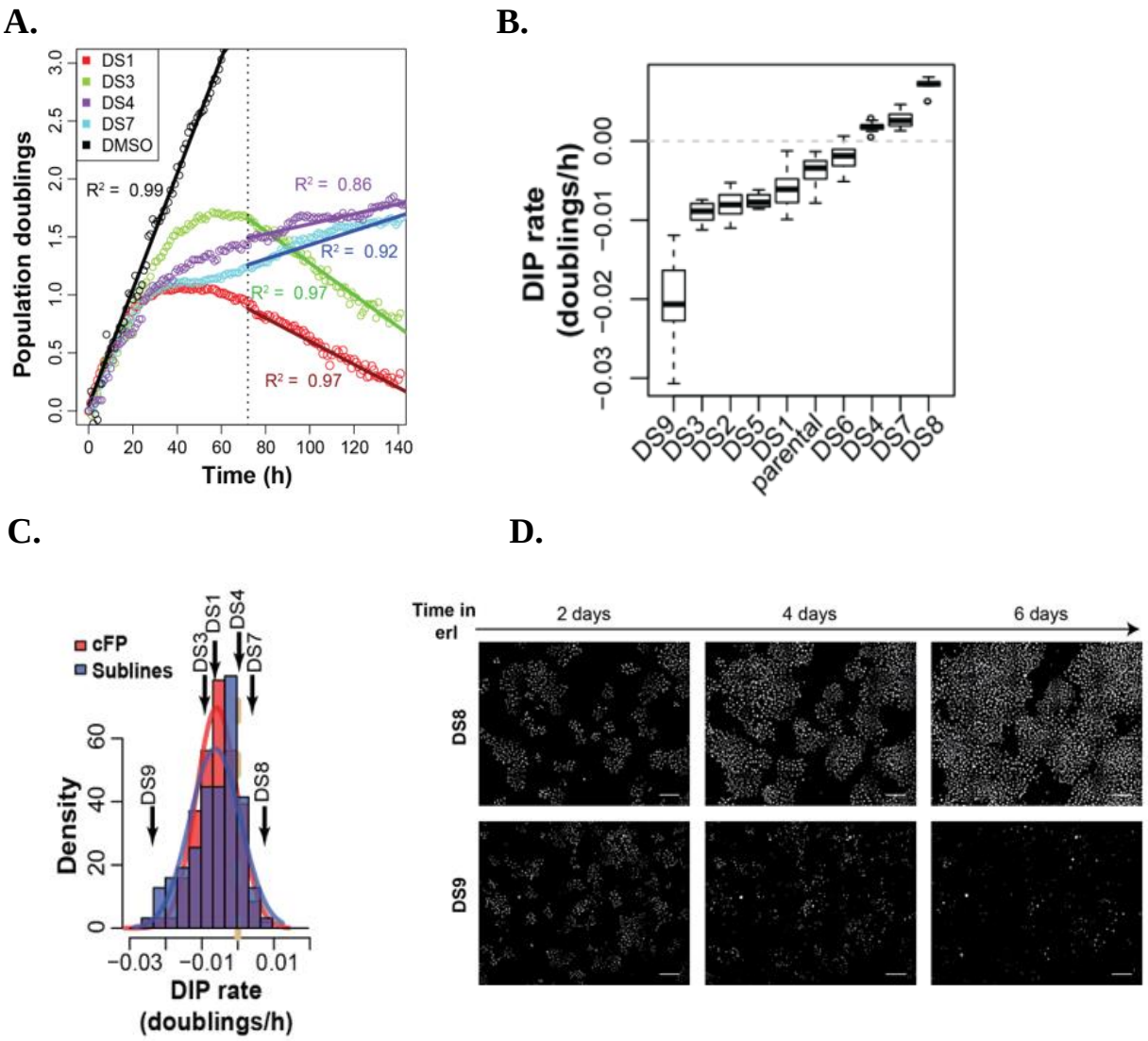


Figure 4.1 PC9 clonal response to erlotinib therapy. **(A)** The drug-induced proliferation rate (DIP) is the linear fit to total cell number in log2 scale. **(B)** Variability in DIP rate response seen within subclones. The DIP rate varies between clones ranging from a strong negative response to little effect to therapy. **(C)** The relative density seen in response. The majority of clones have a negative DIP rate. **(D)** An example time course showing DS8 (resistant) and DS9 (sensitive) response to erlotinib therapy over six days. Scale bars are 200 μm Adapted from [85].

We initially sought to examine the effects of 2D versus 3D culture with the purpose of more closely mimicking an *in vivo* environment. Many studies point to variable cell response depending on how they are cultured with changes in proliferation, morphology, drug metabolism, and

differentiation [86], [87]. We found a large difference between 3D culture techniques with a strong correlation on substrate stiffness, leading us to inquire how substrate stiffness changes NSCLC behavior and confers increased resistance to targeted therapy (Figure 4.2).

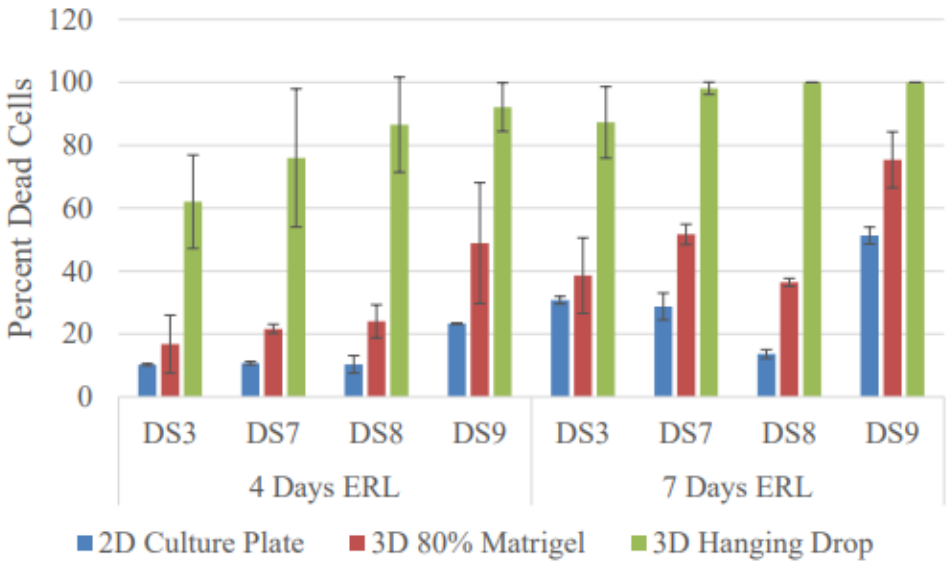


Figure 4.2 Percent dead cells after erlotinib treatment as a function of culture technique. Cells show an increased resistance to erlotinib treatment as substrate stiffness is increased across sublines. Adapted with permission from [88].

4.1. Soft substrate decreases overall cell number similar to erlotinib treatment

To examine the influence of substrate stiffness on NSCLC behavior, cells were grown on polyacrylamide substrates of varying stiffness, modulated from 0.2 kPa to 55 kPa, coated with collagen. For tracking and analysis purposes, cells were labeled with H2B-mRFP and Geminin-mAG. The first reporter allows for continual tracking of individual cells over the course of the experiment, while the latter marks the transition from G1 to S phase, remaining expressed until the cell divides. Cells were imaged every 20 minutes over a period of five days.

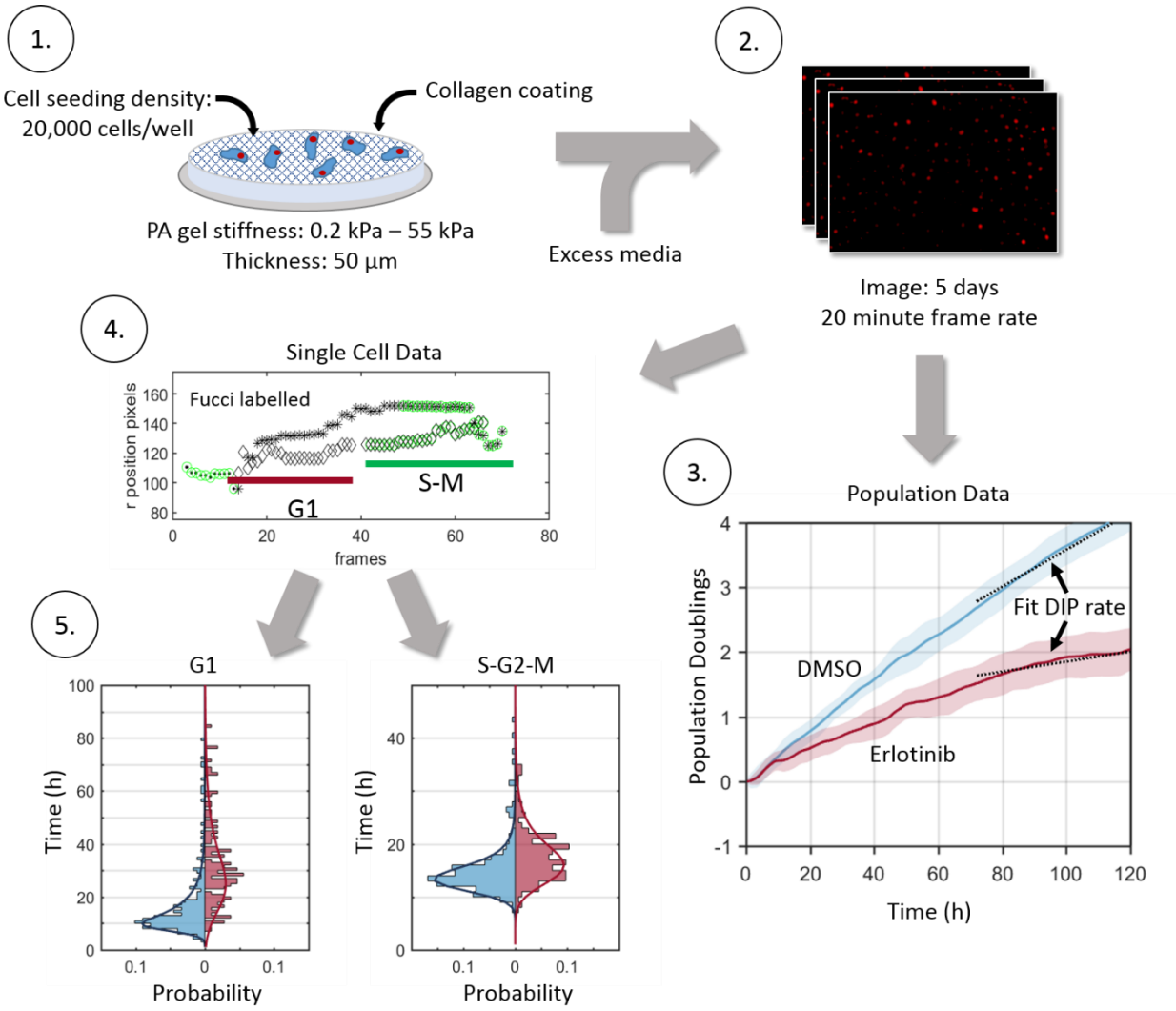


Figure 4.3 Schematic of experimental design on cellular effects of substrate stiffness. **(1)** Cells are seeded on polyacrylamide (PA) gels with stiffnesses varying from 0.2 kPa – 55 kPa. Gels are coated with collagen, allowing cells to adhere at low density. Excess media is added and the media is not changed over the course of the experiment, as it was shown to alter mitotic behavior. **(2)** Cells are imaged every 20 minutes over the course of five days. The frame rate is rapid enough to allow tracking of individual cells over the course of the experiment. **(3)** Population-level analysis is performed using the fractional proliferation framework where the linear response curve is fit to the log2 change in cell number. **(4)** Individual cells can be tracked through the length of the experiment. Cells are histone H2B-mRFP labeled and express Geminin-mAG during S, G2, and M phases. **(5)** The expression of the Geminin label allows for the quantification of cell time spent in G1 and S, G2, and M phases.

We followed the change in overall cell population using the methodology developed by Quaranta et al. called the DIP rate [82]. The DIP rate is the linear fit of cell population doubling

as a function of time after the point of linear response. Using the DIP rate as a growth metric over other fits, such as logistic growth, allows us to compare a single fit quantity across cell populations that are expanding or declining. On the stiffest substrate (55 kPa), the two sublines behave as previously described: with no significant difference between subline DIP rates (Figure 4.4). Upon erlotinib treatment, the underlying heterogeneity of sensitivity is made apparent, with DS8 showing no significant change in DIP rate and DS9 showing a significant negative rate (Figure 4.4).

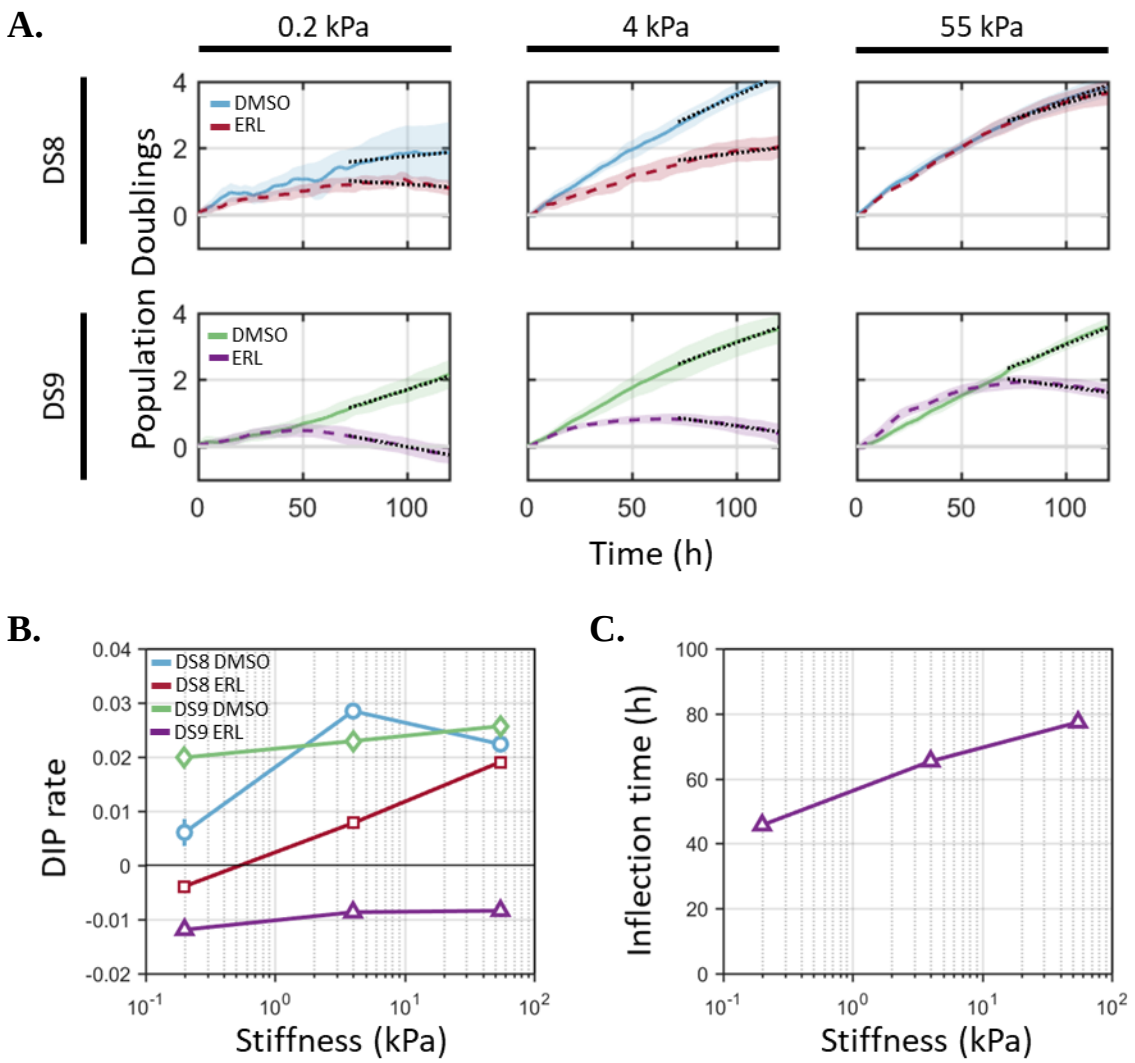


Figure 4.4 DIP rate response to erlotinib and substrate stiffness perturbations. **(A)** Population doubling graphs for the erlotinib-sensitive DS9 and -resistant DS8. The sublines behave similarly when untreated, showing a decrease in overall cell number on soft substrates. On the softest substrate, resistant DS8 has a slightly negative DIP rate under treatment. **(B)** DIP rate as a function of stiffness. DMSO-treated cells show a small decrease in proliferation rate with decreasing stiffness. The sensitive DS9 shows a negative DIP rate across all stiffnesses, while DS8 shows increased proliferation with increasing stiffness. **(C)** The inflection point of sensitive DS9. As stiffness increases, DS9 reaches a greater overall cell number and takes longer to reach an inflection point.

We then investigated the effects of substrate stiffness on DIP rate for both untreated and erlotinib-treated cells. As the substrate becomes softer, cells show a marked decrease in growth rate under dimethyl sulfoxide (DMSO) conditions, still with little difference between sublines. All DIP rates, regardless of stiffness, are positive with no drug treatment. Here, we point out one

limitation of the DIP rate: as the DIP rate expresses the linear fit of the population change over a given window, it does not fully capture all of the dynamics of a cell population. Namely, the initial response and behavior of the cells. On our softest substrate, the DIP rate is significantly different between sublines. In this case, using a logistic growth curve would show more similar behavior between the two experiments. However, it would not be possible to fit a curve that is initially positive, then negative with a logistic curve. We know from previous work in the Quaranta laboratory that DS8 and DS9 are discrete sublines (DS) that respond disparately to EGFR inhibition [82]. In light of our results showing a change in DIP rate with substrate stiffness, we investigated if and how substrate stiffness affects sensitivity to EGFR inhibition. We find that both cell lines behave similarly to previous results on our stiffest substrates. Specifically, DS8 continues to proliferate in the presence of EGFR inhibitor, marked by a positive DIP rate, while DS9 shows a linear, negative DIP rate response after erlotinib treatment. At the intermediate stiffness of 4 kPa, the resistant cell line shows a decreased DIP rate and, intriguingly, at our softest substrate of 0.2 kPa, the resistant cell line exhibits a negative DIP rate. Of note, the erlotinib-treated DS8 grown on 4 kPa substrate responds with the same curve and DIP rate as the non-treated DS8 grown on 0.2 kPa substrate. DS9 shows sensitivity across all substrates with a substantial negative DIP rate. There is a small decrease in DIP rate on the 0.2 kPa substrate compared to the 55 kPa substrate. Additionally, as substrate stiffness increases, the DS9 subline shows a higher cell count before the action of drug treatment is evident and takes a longer time to reach a negative DIP rate.

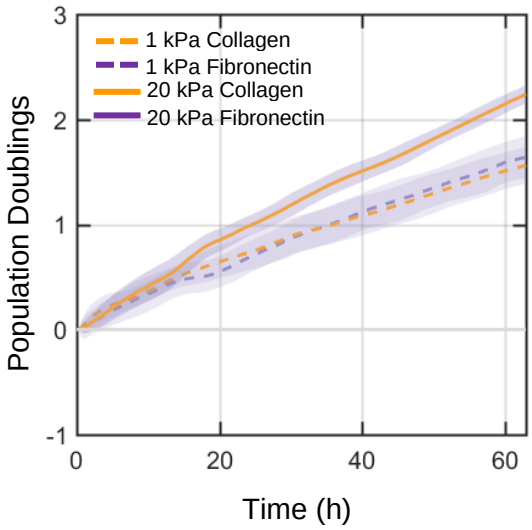


Figure 4.5 DIP rate response to collagen and fibronectin extracellular matrix. Resistant DS8 was cultured on 1 kPa and 20 kPa substrates using collagen and fibronectin as extracellular matrix proteins. There was no discernable difference in the resulting DIP rates due to matrix proteins.

4.2. Sensitivity to substrate stiffness is maintained for different ECM proteins

We questioned whether the differences we saw in cell population responses were a result of substrate stiffness or perhaps another factor. Therefore, we considered if a different extracellular matrix protein would elicit a different cellular response due to disparate cellular adhesion or activation of different signaling pathways. Resistant DS8 was grown on two stiffnesses, 1 kPa and 20 kPa. We found no difference in the resultant DIP rate induced by the two proteins.

4.3. Erlotinib treatment lengthens G1 time in a substrate stiffness-dependent manner

The decreased DIP rate observed in cells grown on soft substrates and in the presence of EGFR inhibitor led us to investigate the origin of the decreasing cell number. A change in DIP rate is produced by a change in fraction of actively dividing, dying, or quiescent cell populations. We investigated the cause of this change in DS8 and DS9 by modulating the stiffness of the substrate in addition to treating with erlotinib. To track variations in cell cycle progression timing, we used a fluorescent marker for G1/S transition, Geminin-mAG. The marker is expressed at the onset of S-phase until the cell completes mitosis. We followed both the time that Geminin was not expressed, labeled G1, and the time that Geminin was expressed, S/G2/M, for individual cells across experiments. We fit an exponentially-modified Gaussian (EMG) distribution to cell cycle events to allow comparison between experiments, noting however, that the fit does not capture all the features of our experimental distributions. In particular there appears to be a small, secondary peak at approximately twice the time of the primary peak (Figure 4.6A).

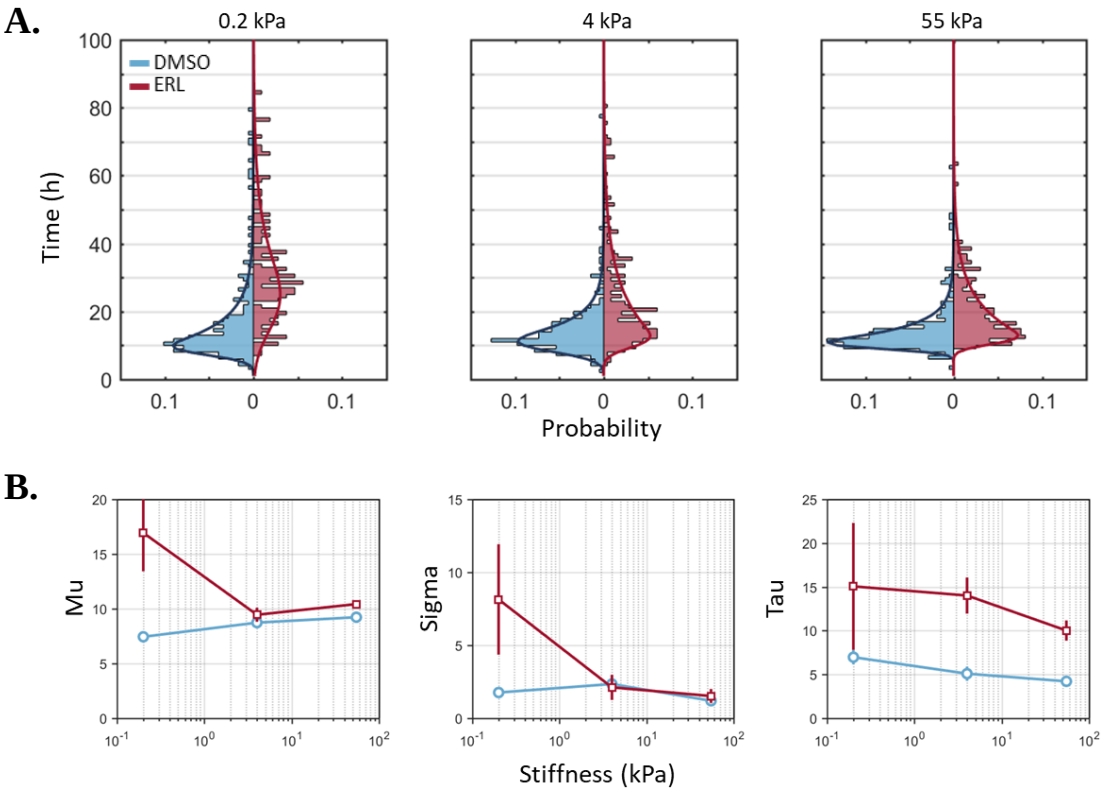


Figure 4.6 G1 time distributions. **(A)** Probability distributions of time spent in G1 for different substrate stiffnesses with and without erlotinib treatment. Curves represent the exponentially-modified Gaussian fit to the distribution. **(B)** Parameters expressing the exponentially-modified Gaussian fit. Increased stiffness results in decreased skew, while erlotinib treatment greatly increases skew for all substrate stiffnesses. The softest substrate shows a large shift in G1 time.

The EMG is a right-skewed distribution that captures G1 variability better than most unimodal models (e.g. log-normal, gamma distributions). The EMG distribution is the result of the convolution of a Gaussian and exponential distribution. The familiar Gaussian probability density function (PDF) with mean μ and standard deviation σ can is represented as follows:

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4.1}$$

and the exponential probability density function with rate τ and $x \geq 0$:

$$h(x;\tau) = \tau e^{-\tau x} \tag{4.2}$$

The resulting EMG has a probability density function:

$$f(x;\mu,\sigma,\tau) = \frac{\tau}{2} e^{\frac{\tau}{2}(-2x+2\mu+\tau\sigma^2)} erfc\left(\frac{-x+\mu+\tau\sigma^2}{\sigma\sqrt{2}}\right) \quad (4.3)$$

It is important to remember that familiar parameters no longer have the same meaning. For example, $\mu + 1/\tau$ represents the mean now instead of μ.

Additionally, we find that upon drug treatment, there is a delay until drug action. For this reason, we capture cell cycle dynamics data for times greater than 24 hours. We first observe that changing substrate stiffness has little apparent effect on G1 length, as exhibited by a slight change in μ and τ. An increase in τ is associated with increased skewness of the distribution (e.g. how long the rightward tail is). Comparing erlotinib-treated cells to their counterparts grown on the equivalent substrate stiffness, we see an increase in G1 time. For erlotinib-treated cells, τ increases two-fold from the non-treated counterparts.

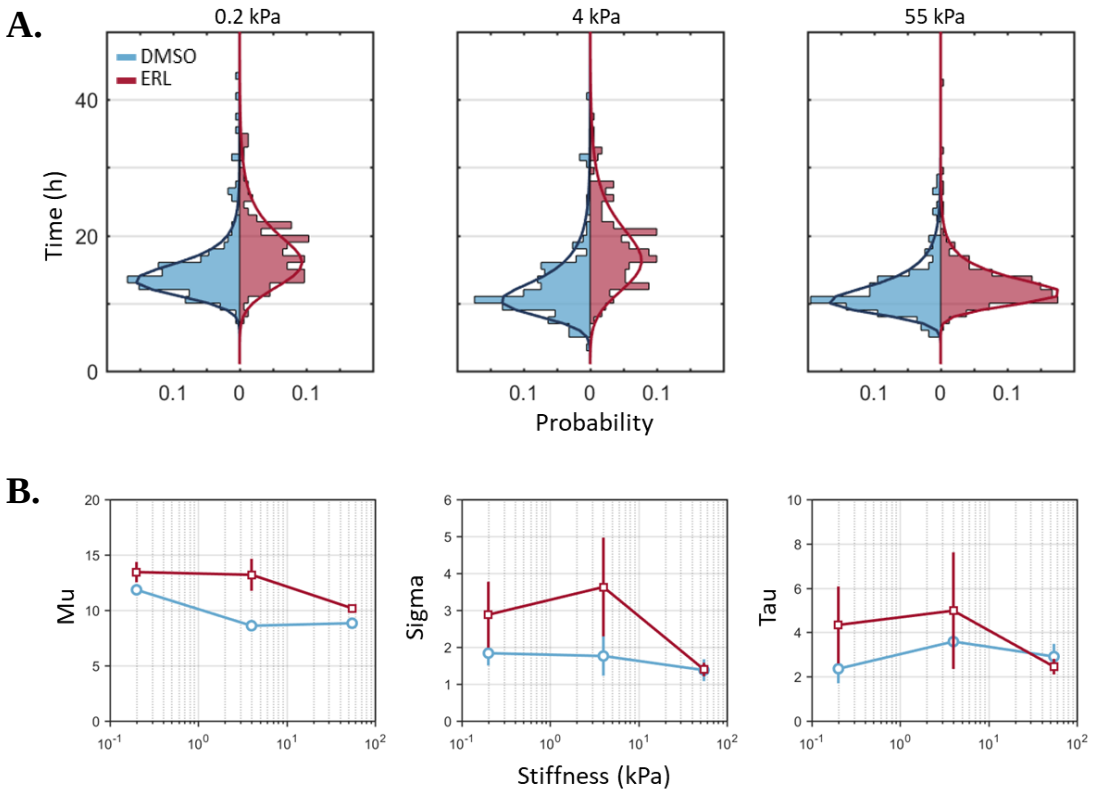


Figure 4.7 S-G2-M time distributions. **(A)** Probability distributions of time spent in S-G2-M for different substrate stiffnesses with and without erlotinib treatment. Curves represent the exponentially-modified Gaussian fit to the distribution. **(B)** Parameters expressing the exponentially-modified Gaussian fit. There is a decrease in mean time spent in S-G2-M as stiffness increases. Treatment with erlotinib cause a slight increase in mean time and increased distribution spread. There is very little skew in S-G2-M in comparison to G1.

4.4. Both substrate stiffness and erlotinib treatment increase S/G2/M times

Next, we compare the effects of modulating substrate stiffness and erlotinib treatment on S-G2-M timing. The first observation we note is that the shape of the distribution is more Gaussian than the G1 distribution. At low substrate stiffness, the mean time spent in S/G2/M is approximately 13 hours compared to 11.5 at 55 kPa stiffness, contributing to the decreased DIP rate. All erlotinib-treated cells have increased S/G2/M time in comparison to their untreated counterparts (Figure 4.7B). There is an increased broadening of the distribution on soft substrate from untreated to treated cells that disappears on the 55 kPa substrate. Decreased DIP rate from

erlotinib treatment is the result of a combination of increased death rate, increased quiescence, and extended time in both G1 and S/G2/M. The major differences observed are an increased mean and variance under erlotinib treatment.

4.5. A subset of cells spends approximately double the mean time in G1

A major factor for the poor fit using either an exponentially-modified Gaussian or log-normal fit for the G1 distributions is the appearance of a secondary peak at approximately double the average G1 time. We have found a small subset of cells take approximately twice as long to complete the G1/S transition (Figure 4.8). The increased intermitotic time seen at lower stiffnesses or under drug treatment are primarily caused by increases in G1 time, as opposed to a change in S-G2-M, evidenced by comparing shifts in distributions in Figure 4.6 and Figure 4.7. Decreased stiffness and erlotinib treatment both increase the fraction of cells in the second peak, with the greatest fraction occurring on the 0.2 kPa substrate with erlotinib treatment. Using a right-skewed fit as a description of the distribution both overestimates intermediate G1 times and underestimates extended G1 times.

Therefore, we compared fits of common models used to describe the cell cycle, most being right-skewed in nature. We used the Akaike information criterion (AIC) to evaluate the fits of our statistical models. The AIC is a means of comparing the quality of model in comparison to a separate model used to describe a distribution [89]. AIC is based within the field of information theory. When a statistical model is used to represent data, the model will generally never return an exact reproduction of the original data. In other words, since we are using a model as a representation, some amount of information is lost about the original data. AIC is an estimation of the information that is lost by using a given statistical model. By comparing the AIC between

statistical models, one can select the model with the least information lost. The AIC also includes a cost penalty that is based on the number of estimated parameters. This has the practical effect of avoiding overfitting.

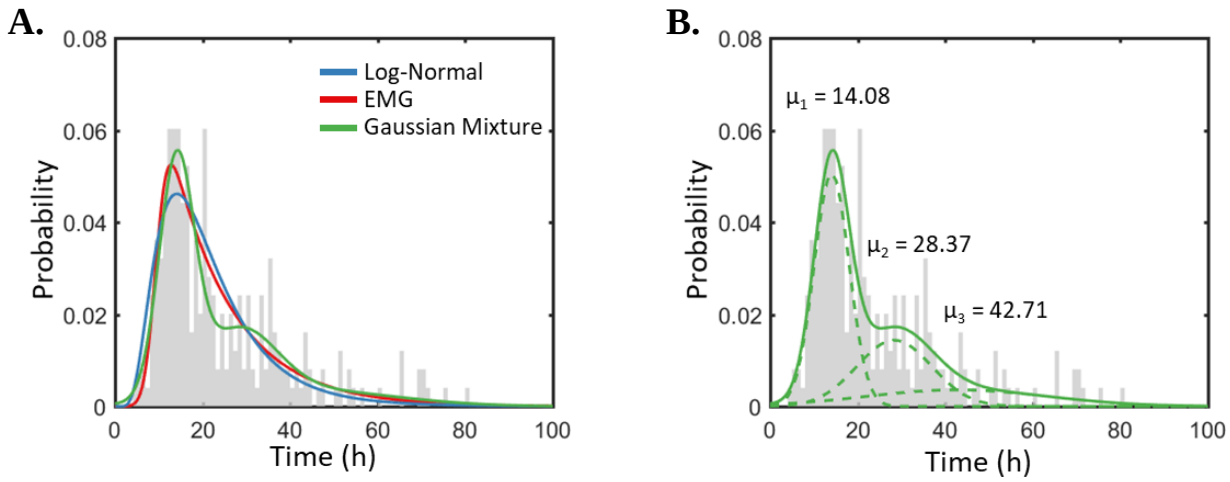


Figure 4.8 G1 statistical model fits. **(A)** Model fits of a log-normal, exponentially-modified Gaussian (EMG), and 3-component Gaussian mixture model. The EMG and Gaussian-mixture model fit the primary peak well. The unimodal models, EMG and log-normal, both overestimate the distribution from approximately 20-30 hours. **(B)** The three Gaussian components that comprise the Gaussian-mixture model. The means for each distribution are approximately all intervals of 14 hours.

Comparing the AIC values, we find that the EMG is the best quality model for the unimodal distributions, scoring consistently better than a Gaussian, log-normal, or gamma distribution (Figure 4.9). The log-normal fits untreated cells fairly well, but serves as a poor statistical model in comparison to the other models examined for erlotinib treated cells. In comparison to the models we fit, the three-component Gaussian mixture model scores the best AIC consistently. Where the secondary peaks are more pronounced (under erlotinib treatment) the two-component Gaussian mixture model also scores well.

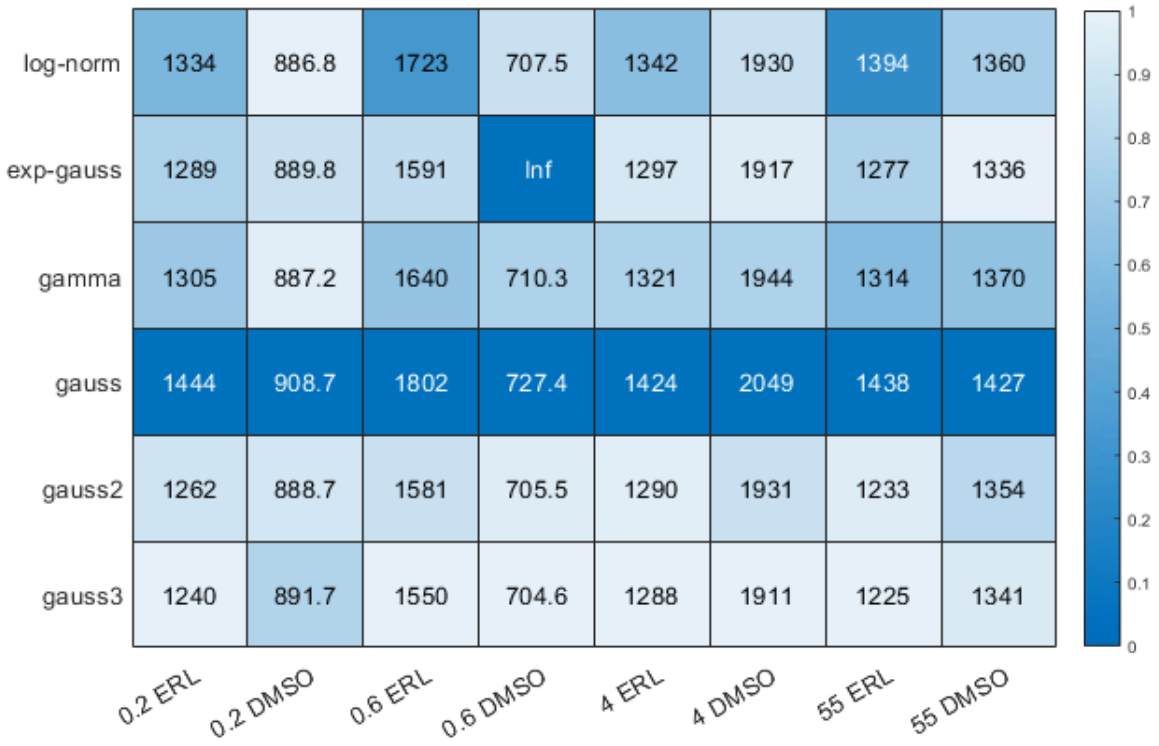


Figure 4.9 AIC scores for fit statistical models. The EMG scores the best out of the unimodal characterizations. The three component Gaussian scores the best of all models most consistently across stiffness and erlotinib perturbations. A lower AIC score represents a better quality model.

4.6. Signaling changes in response to erlotinib treatment and matrix stiffness

After examining how erlotinib treatment and substrate stiffness perturbations affect total cell population and mitotic behavior, we inquired how the same microenvironment alterations modify the cell signaling network. Changes in cell signaling are ultimately manifested in cell behavior (i.e. increased motility, changes in apoptotic, quiescence, and mitotic rates). Here, we seek to associate the changes in key signaling pathways to overall cell population dynamics. The relative signaling strength measured here provides model input in Chapter V. While Western blot analysis does not provide the specific concentration of a molecule (all measurements are relative to a control protein) it does provide detailed enough information to make comparisons between environmental perturbations.

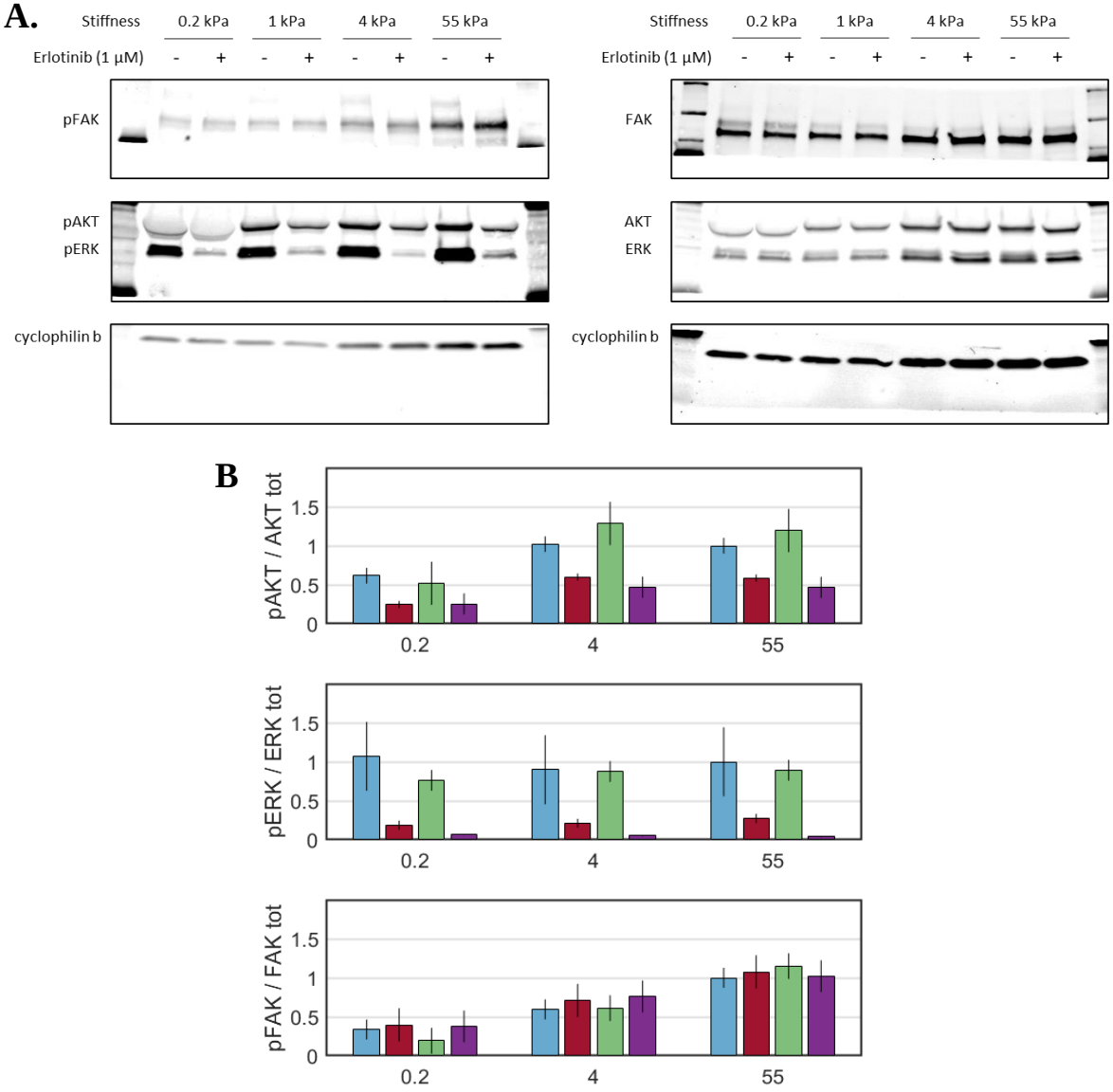


Figure 4.10 Quantitative Western blot analysis of activated fractions of AKT, ERK, and FAK by substrate stiffness with or without erlotinib treatment. **(A)** Representative Western blots of DS8 samples. **(B)** Densitometry of DS8 and DS9 treated with DMSO (control) or erlotinib at 0.2, 4, and 55 kPa. Blue = DS8 DMSO, Red = DS8 erlotinib, Green = DS9 DMSO, Purple = DS9 erlotinib.

We look at three upstream drivers of mitosis, FAK, AKT, and ERK. To determine relative signaling strength, we measure the relative concentration of the phosphorylated state of the protein. Figure 4.10A shows a representative Western blot of the total and phosphorylated state of our

proteins of interest. We examine the same conditions as our imaging analysis. We find a strong stiffness dependence of FAK signaling that is not dependent on treatment or subline. Examining AKT signaling, we find decreased activation at the softest (0.2 kPa) stiffness. Erlotinib treatment diminishes activation approximately two-fold for both sublines. Interestingly, there is no stiffness dependence of ERK activation. Erlotinib treatment greatly reduces ERK signaling, nearly completely ablating activation in DS9.

4.7. Fractional proliferation analysis

We are interested in associating the change in overall cell number to single cell fate decision of mitosis, apoptosis, and quiescence. We use the fractional proliferation framework developed by Quaranta et al. to analyze cell response [78]. The fractional proliferation framework makes use of a quiescence growth model. The model contains two compartments consisting of dividing and nondividing cells (Figure 4.11A). Apoptosis can occur from either compartment. Three parameters determine the rate of cells entering or leaving a compartment: *d*, the division rate, controls the growth of the dividing compartment, *q*, the quiescent rate, describes the number of cells leaving the dividing and entering the quiescent compartment, and *a*, the apoptosis rate, describes the number of cells dying from both compartments.

We examine the fold change in rate values as a function of stiffness and drug treatment for DS8. Interestingly, the division rate does not show a large alteration upon drug treatment. Stiff substrate increases division rate ~50%. There is a large decrease in death rate with increasing substrate stiffness. We find that erlotinib treatment increases the quiescent rate for all stiffnesses.

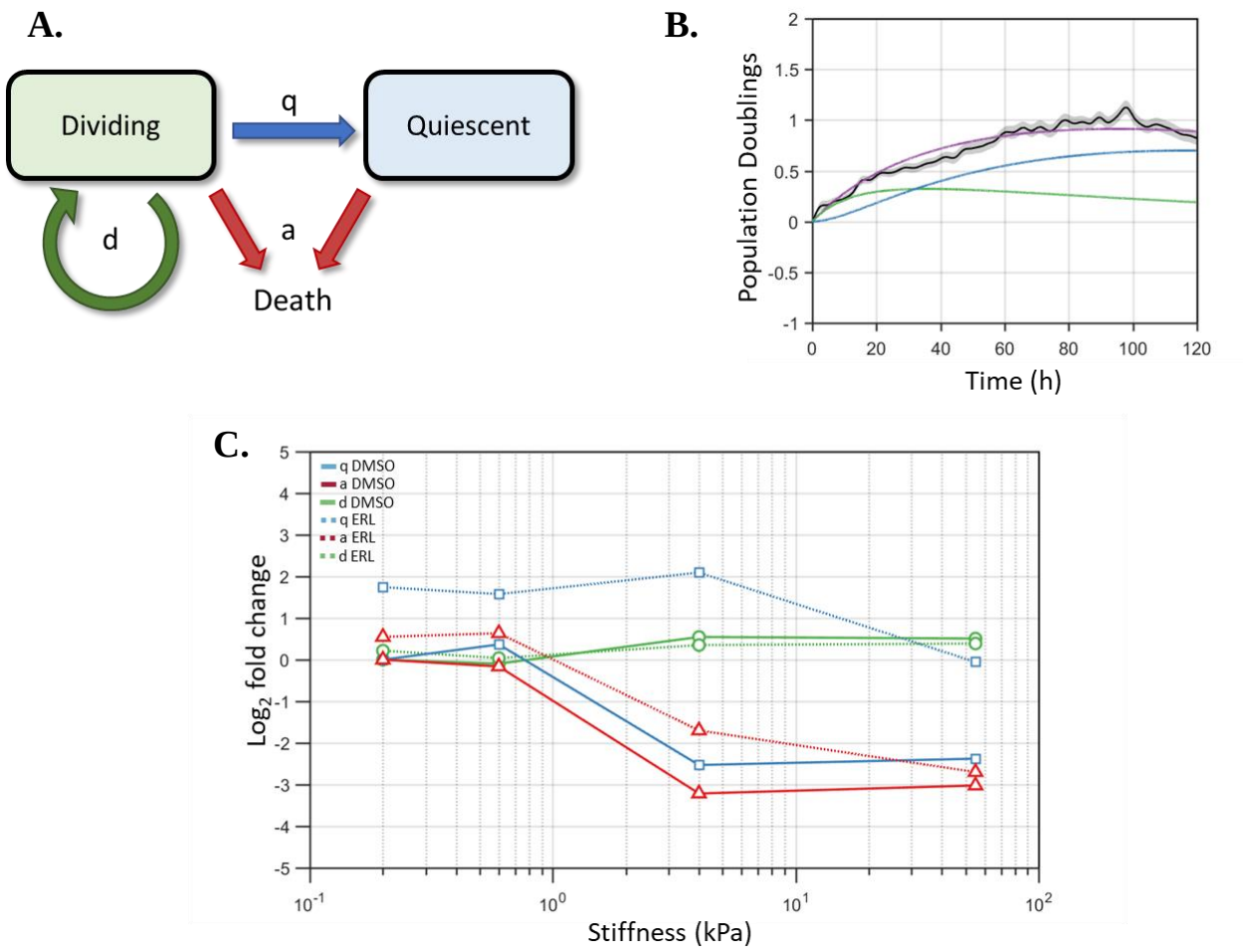


Figure 4.11 Subline fractional proliferation analysis. **(A)** Schematic of the fractional proliferation model consisting of three compartments of dividing, quiescent, and dying cells. d, q, and a define the rates of entry into each compartment, respectively. **(B)** Example fit of the fractional proliferation model on DS8 grown on 0.2 kPa substrate under erlotinib treatment. The purple curve represents the total cell number, blue number of quiescent cells, and green dividing cells. **(C)** The $\log_2$ fold change of fractional proliferation rate parameters as a function of time.

4.8. Discussion and conclusion

In this chapter, we experimentally measured NSCLC single cell and cell population response to changes in stiffness under erlotinib treatment. Understanding what role stiffness can play in cell behavior has important implications in defining disease progression. Lung cancer is known to metastasize to other tissues within the body, and as tumors grow, the tumor cells modify their own

microenvironment, making stiffness an integral factor in treatment response. Single cell studies allow us to investigate heterogenous treatment response in genetically identical subpopulations of a primary tumor. It is important to first define what the expected rates of cell division, quiescence, and death in response to treatment are to then identify differences in cell to cell response.

We established that substrate stiffness can have a profound effect on cell behavior, showing that a stiffer substrate leads to increased proliferation as measured by the DIP rate. These findings fit well into a growing field of literature showing cell stiffness-dependent behavior across many different cell types and applications [25], [90], [91]. Mechanotransduction has been recognized as equivalently important as classic molecule-based signaling. A significant experimental finding we have shown is that matrix stiffness modulates treatment response in an erlotinib-resistant subline. At our softest substrate stiffness of 0.2 kPa, the resistant subline shows a slightly negative DIP rate. This suggests that treatment focusing on disrupting stiffness signaling could be effective, especially in metastatic cases where the new substrate is far stiffer than the lung (e.g. bone).

We observe treatment response heterogeneity between sublines on the stiffest substrate, with DS8 continuing to proliferate under treatment and DS9 quickly dying. We have found in past studies that cMet expression correlates with treatment sensitivity [85]. However, we do not observe any difference in cMet signaling quantified with Western blot densitometry. This suggests there may be some other mechanism of resistance or cMet may be modified in a different capacity. There have been recent studies showing that cMet can replace α5 integrin to form a cMet/β1 integrin complex that has a greater affinity for fibronectin than the α5/β1 integrin complex [92]. Defining the binding partners of surface receptors and how they change in disease states is important to understanding the mechanisms of cellular signaling.

We found that stiffness and erlotinib drug treatment modulate G1 and S-G2-M time. Erlotinib treatment causes a large increase in G1 time, resulting in a reduced DIP rate. Defining the shape of the G1 time distribution can reveal insight surrounding the underlying biological processes and has been the subject of ongoing research. The exponentially-modified Gaussian curve has been successfully implemented in the past and offers a suitable fit for the G1 distribution [78]. The curve is the convolution of an exponential and a Gaussian distribution. These two distributions can be mapped to the processes within the cell cycle. The Gaussian distribution is associated with tasks that occur in sequence corresponding to G2, S, M, and most of G1. The exponential portion is produced by a single rate-limiting event, the G1/S transition [93]. We know underlying cell cycle progression is a series of tightly coupled and interrelated components centering around the CDK network. The cell cycle contains several feedback loops that the EMG distribution does not take into consideration. However, we know that feedback loops are important to the proper function of the cell cycle. Given the tightly coupled nature of the cell cycle, how do feedback loops influence the G1/S transition (exponential portion) of the mitotic time distribution? We build a testable, molecular-level model examine such effects.

We observed experimental evidence that the intermitotic time is not a unimodal distribution. There is a secondary peak that becomes more pronounced under certain experimental conditions (low stiffness and erlotinib treatment). To characterize the complex curve shape, we used a non-equally weighted Gaussian mixture model containing two or three Gaussians. We compared those results against commonly used unimodal distributions, gamma, log-normal, and exponentially-modified Gaussian. Using the AIC as a metric, we found the three Gaussian mixture model to have the best overall fit across treatment and stiffness modulation. Notably, the peaks appear to occur at regular intervals, with peaks at approximately 14, 28, and 42 hours. This phenomenon has been

observed in past studies and referred to as ‘cell cycle quantization’ [94], [95]. We will use a molecular model to recapitulate the phenomenon and discuss it further in the next chapter.

Single cell behavior and fate decision is related to the overall cell population. In the past, the fractional proliferation model, which considers the rate of cells dividing, entering quiescence, and dying, has been used to describe population changes in response to an external perturbation [78]. We use the same methodology here and define the rate constants expressing the dividing, quiescent, and apoptotic populations. Interestingly, the division rate increases approximately 50% on stiffer substrates but shows little dependence on drug treatment. The change in population dynamics can be attributed to increased rates of quiescence and death. Drug treatment increases quiescent rate across all stiffnesses. Unfortunately, an increased quiescent population can be problematic in cancer treatment. Once treatment has ended, relapse can occur, often having effectively selected for therapeutic-resistant cells [96].

One of the primary objectives of this work is to connect molecular level signaling with population level cell response. To this end, we measured the activity level of molecular reporters in major signaling axes connecting cell surface receptors to the cell cycle and apoptosis. We have chosen AKT, ERK, and FAK, a small selection of the multitude of factors that comprise the full signaling network. We find specific perturbations result in disparate responses between these three proteins. FAK activation is highly dependent upon substrate stiffness. AKT shows decreased signaling at low stiffness while ERK shows no stiffness dependence. However, ERK is strongly inhibited by erlotinib treatment with DS9 showing more sensitivity than DS8. We can now use the differentially activated drivers of mitosis and apoptosis to inform our molecular model.

4.9. Future studies

Our findings provide helpful characterization of the difference in substrate-dependent erlotinib therapy behavior. We have some initial findings continuing the above work and potential avenues of investigation outlined below.

4.9.1. Combination drug therapy

We know that NSCLC is a heterogenous disease that recurs in 30-55% of stage I patients and the vast majority of higher staged patients [27]. In an effort to affect more of the cells, combination drug therapy has been developed. When the cancer recurs, it is because some cells have escaped the first-line therapeutic, whether that be resection, chemotherapy, targeted inhibitors, or immunotherapy. Often, this escape is because the remaining cells are resistant to the first-line strategy. Therefore, there are different therapeutics often used in recurrent cases. In the case of our PC9 discrete sublines, we have observed a correlation between erlotinib-treated DIP rate and cMet expression (Figure 4.12), This suggests that cells with high cMet expression are not effectively targeted by erlotinib. Crizotinib is a dual inhibitor of cMet and ALK. It has been investigated with another EGFR-targeted drug, afatanib, in the context of various resistant NSCLC. Only the

combination of afatanib and crizotinib was able to effectively target all three cell lines: EGFR T790M mutation, cMet amplification, and EGFR overexpression [97].

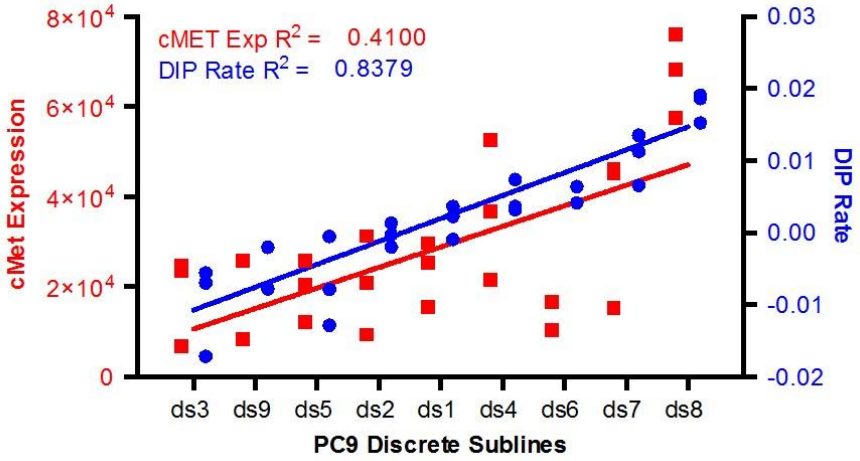


Figure 4.12 Erlotinib-treated DIP rate of PC9 discrete sublines correlates with cMet expression. Shared with permission from [85].

This project has focused on treatment with erlotinib, but we did perform preliminary experiments with crizotinib in our erlotinib-resistant discrete subline, DS8. DS8 had a dose-dependent response to crizotinib treatment (Figure 4.13), suggesting a combination therapy of erlotinib and crizotinib would have a beneficial effect on the original adenocarcinoma cell line, PC9. Therapeutics are used in combination more and more often, either concurrently, sequentially, or both. Understanding these clinically relevant scenarios with extension of our stochastic model has the potential to provide valuable insight into the mechanism and effectiveness of various combinations. In the final stage of a project like this, we would be able to characterize key features of an individual's emerging or recurrent cancer and provide personalized therapeutic strategies.

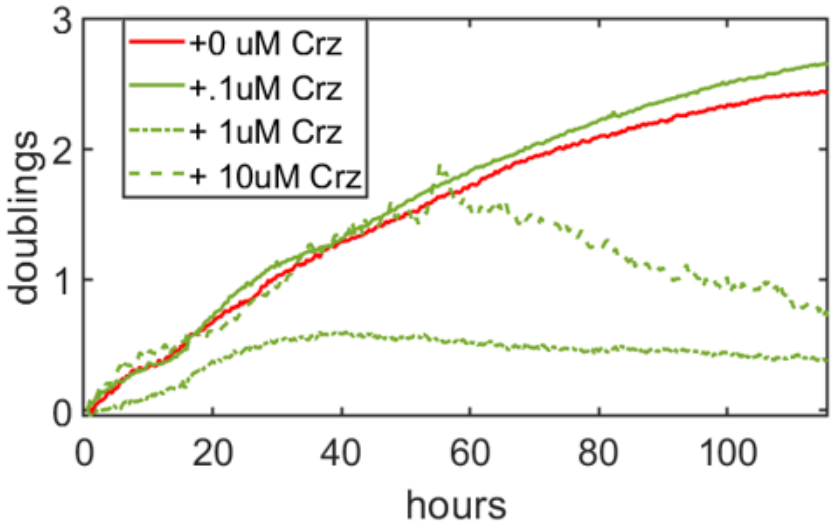


Figure 4.13 DS8 population doublings after treatment with cMet-targeting crizotinib (Crz) shows a dose-dependent inhibitory response.

4.9.2. *In vivo* experiments

Further validation of our work will include studies in *in vivo* models of NSCLC. Preliminary data examining erlotinib treatment of PC9 sublines in a mouse model reflect our *in vitro* results. Briefly, DS8 (resistant) or DS9 (sensitive) cells were injected subcutaneously in Matrigel into the backs of mice and treated daily with DMSO (control) or erlotinib injections. Erlotinib treatment slowed the growth of DS8 cells and was able to kill many DS9 cells, as measured by cell volume quantification via ultrasound (Figure 4.14) [85]. Further *in vivo* studies should incorporate insight gained from our *in vitro* and *in silico* data. Specifically, we would expect even the resistant cells of a heterogenous tumor to respond to erlotinib treatment on soft substrates. However, the brain is a common site of NSCLC metastasis (38%) and is a soft tissue (0.5-7.3 kPa) [10], [11], [23], [24]. Delivery to the CNS is an added challenge, but even high dose erlotinib or gefitinib does not yield a dramatically increased survival time in humans [98]. Extension of our studies to include other relevant cues from the brain microenvironment may reveal why NSCLC brain metastases are

resistant. Similarly, studying NSCLC *in vivo* allows us to observe and eventually model more interactions (e.g. immune response, specific microenvironments, cell biomarkers) to better describe cell fate in a complex environment.

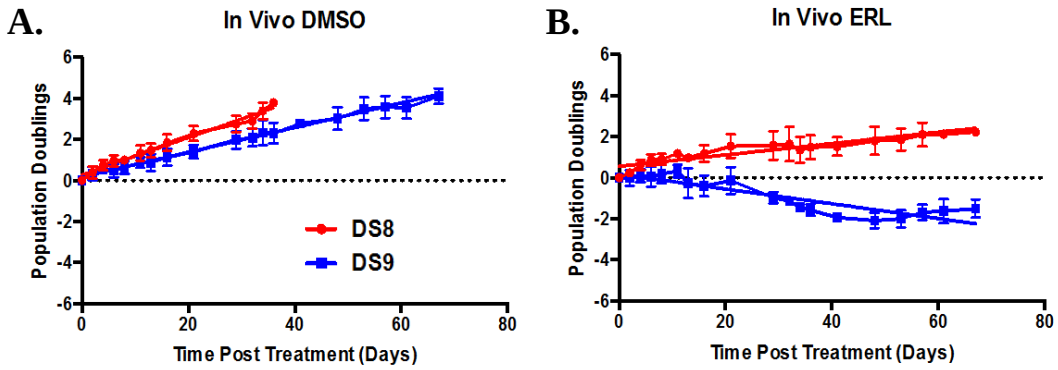


Figure 4.14 Cell volume of PC9 discrete sublines injected in Matrigel subcutaneously on the backs of mice. DMSO or erlotinib (ERL) was injected daily and cell volume was assessed via ultrasound. **(A)** Resistant subline DS8 and sensitive subline DS9 grow at similar rates *in vivo*. **(B)** When treated with erlotinib, DS8 grows at a slower rate and DS9 begins dying within 14 days post treatment. Shared with permission from [88].

4.10. Methods

4.10.1. Cell lines and cell culture

PC9 discrete sublines (DS) were obtained from the laboratory of Vito Quaranta (Vanderbilt University). They were derived as single cell clones from parental PC9 cell line, expanded in drug-free media, and fluorescently labeled with Geminin-mAG (Fucci system) and H2B-mRFP. Cells were cultured in CO_2- (5%), temperature- (at 37°C), and humidity-controlled incubators. Before use, the cells were confirmed to be negative for mycoplasma using MycoAlert PLUS kit (Lonza). Cells were cultured in RPMI 1640 media supplemented with 10% FBS (Atlanta Biologicals) and Antibiotic/Antimycotic mix (Gibco). During imaging, the media was substituted for OPTI-MEM

media (Gibco), supplemented with 2% FBS, and Antibiotic/Antimycotic mix to reduce background auto-fluorescence. The growth rate was confirmed to be the same for these sublines in both types of media.

4.10.2. Collagen-coated polyacrylamide gels

2D polyacrylamide gels are made according to the protocol of the Gardel laboratory [99]. Briefly, circular glass coverslips are activated with 3-aminopropyltrimethoxysilane (3-APTMS) and glutaraldehyde. The coverslips are immobilized on the bottom of a 6-well plate. Polyacrylamide solution of a chosen stiffness is allowed to polymerize between a hydrophobic coverslip and an activated coverslip. The hydrophobic coverslip is gently removed from the well using a razor blade edge. Sulfo-SANPAH is placed upon the gel and exposed to UV light, changing in color from red-orange to brown. UV-treated coverslips are quickly washed in ddH2O and coated with 0.1 mg/ml fibronectin or collagen. Fibronectin or collagen is allowed to bind for 1-2 hours at room temperature or overnight at 4°C. Gels are then washed extensively with PBS (pH 7.4) under sterile conditions and sterilized under a germicidal lamp in a tissue culture hood for 30 minutes.

4.10.3. Western blot

Cells were lysed in RIPA buffer (Sigma) supplemented with protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche). The difficulty in removing cells from the PA gels lead to lysing them straight on the gel. Sample lysates were boiled with Laemmli sample buffer and loaded in a 4-20% gradient gel (Bio-Rad) for electrophoresis. Protein was transferred to polyvinylidene difluoride membranes and then blocked with blocking buffer (LI-COR) and incubated with primary antibodies overnight at 4°C. After 3× washing with TBS, secondary

antibodies were applied for 1 hour and then washed with TBST (0.05% Tween 20) before imaging. Blots were scanned using the Odyssey imager (LI-COR). Densitometry was performed using ImageJ software.

All antibodies are from Cell Signaling Technology and are as follows: phospho-FAK(Tyr397) #8556, FAK #13009, phospho-ERK and ERK (Thr202/Tyr204) antibody duet #8201, phospho-EGFR and EGFR (Tyr1068) antibody duet #11862, phospho-AKT and AKT (Ser473) antibody duet #8200.

4.10.4. Quiescence growth model

The proliferation kinetics model describes the rate at which cells change the overall population given the rates of cells proliferating, dying, and entering quiescence. The model is represented by a pair of ODEs with x characterizing the number of dividing cells, y the number of quiescent cells, and x+y the total cell number. Rate constants for the equations are as follows: d, cell division rate, q, quiescent rate, and a, the death rate.

$$\frac{dx}{dt} = (\mathrm{d} - \mathrm{q} - \mathrm{a})\mathrm{x} \tag{4.4}$$

$$\frac{dy}{dt} = \mathrm{qx} - \mathrm{ay} \tag{4.5}$$

The analytical solution for the equations is:

$$x(t; x_0, \mathrm{d}, \mathrm{q}, \mathrm{a}) = \mathrm{x}_0 e^{(d-q-a)t} \tag{4.6}$$

$$y(t; x_0, \mathrm{y}_0, \mathrm{d}, \mathrm{q}, \mathrm{a}) = \left(y_0 + x_0 \frac{q}{d-q} \left(e^{(d-q)t} - 1 \right) \right) e^{-at} \tag{4.7}$$

If the division rate d is greater than the sum of the quiescence rate q and apoptotic rate a, then the system will approach exponential growth. Including a quiescent population allows for nonlinear proliferation curve.

4.10.5. Drug-induced proliferation (DIP) rate

We use DIP rate as our major quantification of drug response for our cell culture work. DIP rate is the doubling rate of cell proliferation as described by

$$N(t) = \mathrm{N}_0 2^{k_{DIP}t} \tag{4.8}$$

where N(t) is the cell number at time t, and N_0 is the seeding cell number at time t=0, k_{DIP} is the DIP rate of cell proliferation. k_{DIP} can be calculated as:

$$k_{DIP} = \log_2\big(N(\Delta t)\big) - \log_2(N_0)/\Delta t \tag{4.9}$$

4.10.6. Image acquisition

Images are taken on Axio Observer Z1 spinning disk confocal microscope (Zeiss, serial #3834004261). During imaging, the cells are maintained in incubation conditions of 5% CO_2, 37°C temperature, and constant humidity. Images were acquired with 20X objective at the following laser settings: 561λ at 12% laser power and 4ms exposure, 488λ at 1% laser power and 2ms exposure. The imaging consisted of a 12x12 mosaic with 20 μm overlap, each frame was 400 μm x 400 μm. Images were acquired every 20 minutes for a minimum of five days.

CHAPTER V

5. Modeling cell fate response

In this chapter, we relate a molecular-based stochastic model of the cell cycle and apoptosis to experimentally measured population-level cell behavior. We show that a simple, interconnected model is capable of accurately producing cell population behavior and complex IMT distributions observed *in vitro*. Our model is able to connect the effects of variations at the single cell molecular-level to characteristics of the cell population as a whole.

The basis of our model is centered on the system of four Cdks, that control the ordered progression through the phases of the cell cycle: Gap 1 (G1), DNA synthesis (S), Gap 2 (G2), and mitosis (M). A fifth phase, Gap 0 (G0), is used to describe cells that have left the cell cycle and are not actively dividing, or quiescent cells. Recent work has identified the G0 population as an important collection of cells in the development of cancer treatment resistance. We extend a skeleton model of the mammalian cell cycle, a reduced version of a much larger model, both proposed by Goldbeter et al. [55], [56]. Our model is a simplification of an expansive 'bottom-up' model that exchanges many biochemical intricacies for emphasis on the minimum network architecture necessary for complex oscillatory behavior. Identically to the full model, our reduced version is driven by the chronological activation of the cyclin group of proteins and their associated kinases. Each phase of the cell cycle corresponds to the activation of a specific member of this family: cyclin D/Cdk4-6 induction marks the progression of G1, increase in cyclin E/Cdk2 identifies the G1/S transition, cyclin A/Cdk2 activity indicates S and G2 phase, and cyclin B/Cdk1 precipitates M phase.

Additional features of our model include the G1/S restriction point in the form of the Rb-E2F bistable switch. The restriction point in this form provides a sharp threshold for cell cycle activation that is otherwise absent in the reduced model. We also incorporate the proteins Skp2, CDH1, and p21/p27 to allow for substrate stiffness influence on the cell cycle through the activation of the FAK signaling axis. To integrate cell death, we include the initiator module of an apoptosis model proposed by Tyson et al. [74]. The initiator module functions as a switch describing BAX activation. Upon sufficient stress signal, pro-apoptotic BH3 protein level increases. Initially, BCL2 sequesters BH3 by binding to it. However, if the stress signal is maintained, BH3 will continue to accumulate, activating BAX. The activated, mitochondrial-bound form, BAXm, promotes the release of BH3 from BCL2. Once this module is active, irreversible pores are formed within the mitochondrial membrane, augmenting permeability to cytochrome c. A key component of our model is bistability in both apoptosis and mitosis, allowing for signal threshold activation of both elements.

We substantiate our model by comparing its behavior to experimentally measured IMTs and total cell number for oncogene-addicted NSCLC cells. This approach allows us to adjust input strength to the mitotic and apoptotic modules from EGFR signaling, substrate stiffness, and cMet expression through AKT and ERK and to quantify both single cell- and population-level effects of these perturbations. EGFR signaling indicates the presence of drug. Substrate stiffness is relevant in the case of late-stage cancers that metastasize to softer or harder tissues. NSCLC is often not diagnosed until advanced stages and its most common metastatic site is bone, a tissue 500 times stiffer than lung [22]. Variation in cMet signaling allows us to model the response of NSCLC subpopulations that are resistant to current drug therapeutics. We find that decreased substrate stiffness increases sensitivity to EGFR-targeted therapy in drug-resistant sublines. While drug

treatment and stiffness can have the same overall effect on total cell number, division behavior is markedly different between the two environments. Additionally, we show that IMT distributions are bimodal, with a small fraction of cells dwelling in G1/G0 for approximately double the period. Our model is sufficient to capture these phenomena.

We demonstrate the advantages of using a simplified model that still retains the complex behavior found in the comprehensive version. For example, if we deconstruct the model into two oscillating circuits, we lose the period doubling feature observed *in vitro*. Since simplifying the model would lose this key biological characteristic, we know both loops are required for an accurate model. Our model links the period doubling and delayed mitotic progression to several key drivers of the cell cycle including low cyclin E/Cdk2 activity at mitotic exit, proposing a novel mechanism, which demonstrates its usefulness. Our model retains the necessary structure to make insightful predictions while being computationally efficient.

5.1. Ordinary differential equation (ODE) approach of the cell fate model

We begin with the differential equation model that we established in Chapter III. The initial goal is to investigate the behavior of the model under different signal strength (EGFR, cMet, and integrin signaling) through AKT, ERK, and FAK. The relative activation levels of AKT, ERK, and FAK under various perturbations were established in Chapter IV through quantitative Western blotting.

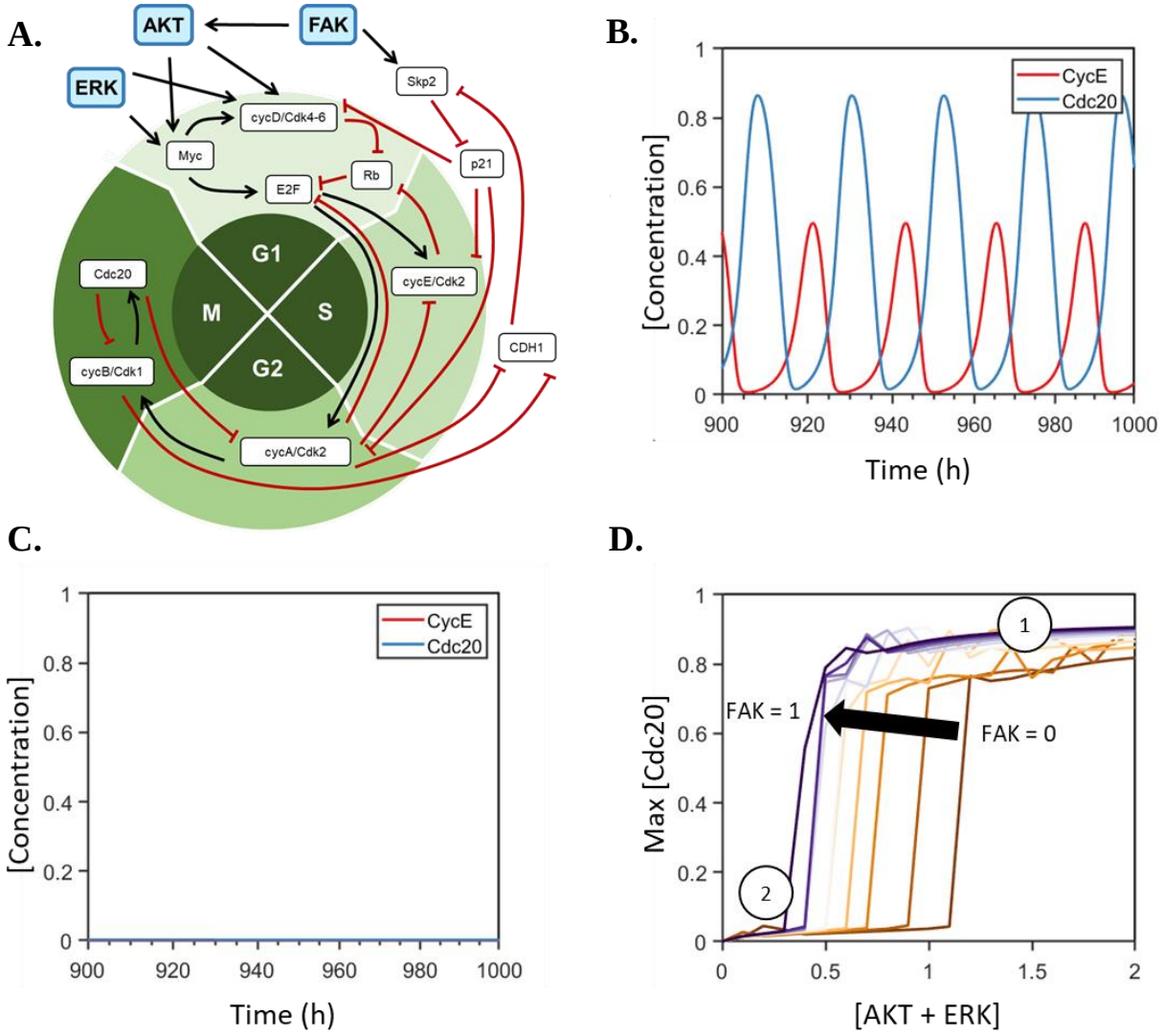


Figure 5.1 Schematic of the mitotic module and system behavior. **(A)** The full wiring diagram of the mitotic module. The circuit is driven by AKT, ERK, and FAK by first engaging the Rb-E2F switch. Once the switch is activated the cell cycle progresses through ordered activation of cyclin E, cyclin A, and cyclin B. **(B)** The ODE solution for the system at AKT = ERK = FAK =1. The time between Cdc20 peak first and cyclin E peak second represents G1 time while the inverse represents S-G2-M. We tune the system manually to match the mean G1 and S-G2-M times found in chapter IV, approximately 12.5 and 11 hours, respectively. **(C)** The steady state solution at AKT = ERK = FAK = 0 showing no oscillations. **(D)** The effect of FAK level on the circuit activation measured by maximum Cdc20 concentration. At high AKT and ERK signal (region 1) Cdc20 is high showing mitotic activity as opposed to low signal (region 2) showing no activity. As FAK level increases the necessary level of AKT and ERK to reach activation is lowered.

The apoptotic and mitotic modules we have established are not yet coupled, so we review each separately. A more detailed schematic including CDH1 and Skp2 is shown in Figure 5.1B. Skp2 is activated by FAK and specifically degrades phosphorylated p21. CDH1 inhibits Skp2 and is itself inhibited by cyclin A and cyclin B. At high AKT and ERK signal, the transcription factor Myc and cyclin D are activated. Myc drives the activation of E2F, and cyclin D inhibits Rb through phosphorylation (the Rb-E2F switch covered in Chapter II). Once E2F reaches a critical threshold, the Rb-E2F switch is activated. Upon activation, the cell cycle can progress by the subsequent activation of cyclin E, cyclin A, and cyclin B. When examining the network topology, several negative feedback loops are evident, suggesting the system is capable of complex behavior.

At high levels of AKT and ERK, the Rb-E2F switch is activated, allowing the cell cycle to progress routinely. The peaks of cyclin E and Cdc20 mark the G1/S transition and the end of mitosis, respectively. Note that the times between peaks are constant, resulting in a static intermitotic time. We manually set the parameters to match our measured experimental results for G1 and S-G2-M from Chapter IV at 12.5 and 11 hours, respectively. At low AKT and ERK activity, the system reaches steady state, indicative of no mitotic oscillations. As FAK level is increased from 0, the necessary signal generated by AKT and ERK to activate the mitotic circuit decreases. Functionally, this is analogous to increasing the stiffness a cell senses from the substrate.

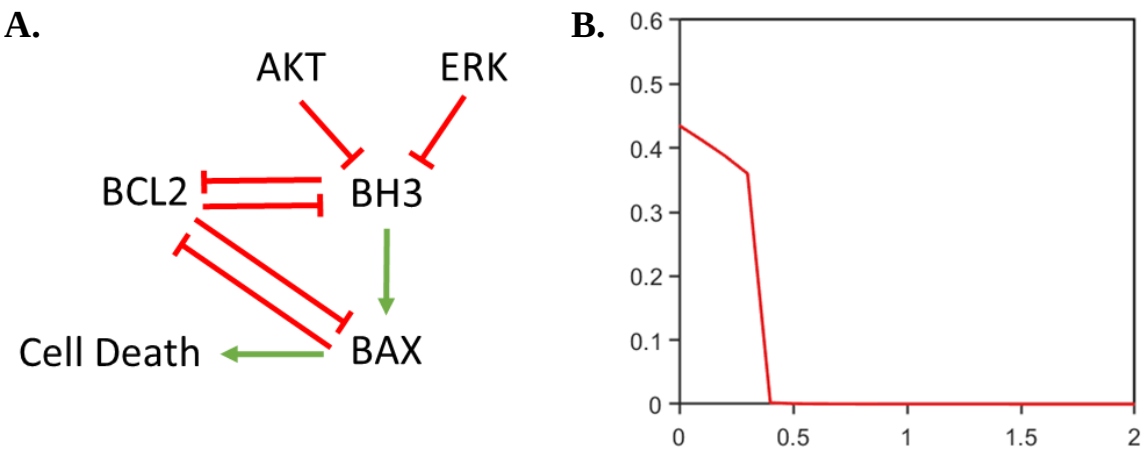


Figure 5.2 Apoptosis schematic and response to input strength. **(A)** The topology of the apoptotic bistable switch. BH3 is activated by low AKT and ERK (high stress) signal. BCL2 can temporally stop BH3 signaling but if a stress signal is maintained, eventually BAX is activated. **(B)** As opposed to the Rb-E2F switch, the apoptotic switch is activated at low input signal.

The apoptotic module is much simpler than its mitotic counterpart. It consists of only three components: BH3, BCL2, and BAX. In our model, we define a stress term that is simply the difference of taking the current AKT and ERK signal strength from the total possible AKT and ERK signal (i.e. 1). If the current signal has strength 0.6, then the stress term would be 0.4. The stress term drives proapoptotic factor BH3. Initially, BCL2 sequesters BH3 but the BCL2 pool is eventually consumed, leading to activation of BAX. BCL2 is also capable of binding to activated BAX. The activated form of BAX is responsible forming a pore within mitochondrial membranes. Pore formation is irreversible, so once the step occurs, the cell is committed to apoptosis.

5.2. Stochastic implementation of the cell fate model

We are interested in capturing the variability observed in our experimental intermitotic time and that is not possible using an ODE solution. While the system of ODEs is capable of complex oscillations resulting in more than one solution for cellular intermitotic time, we would like to

instead introduce small variations around a primary solution, similar to the distributions we see experimentally. There are multiple approaches to solving a system stochastically. Here, we choose to employ the Gillespie or stochastic simulation algorithm (SSA). The approach and algorithm are outlined in section 2.7.4. The approach is the simplest of the stochastic methods but offers an exact solution. We first use the system of differential equations we established for the model to define the set of probabilities associated with each reaction. The number of molecules in the system is set by Ω, the system size. At a large system size, the stochastic solution approaches the results from the ODE solution. In this manner, Ω dictates the amount of noise in the system, with high noise at small Ω and low noise at large values. It is useful to examine the effect of molecular noise on cellular control systems. Understanding how the control systems operate, and importantly, how their regulation can fail, is helpful in developing innovative treatments and therapies.

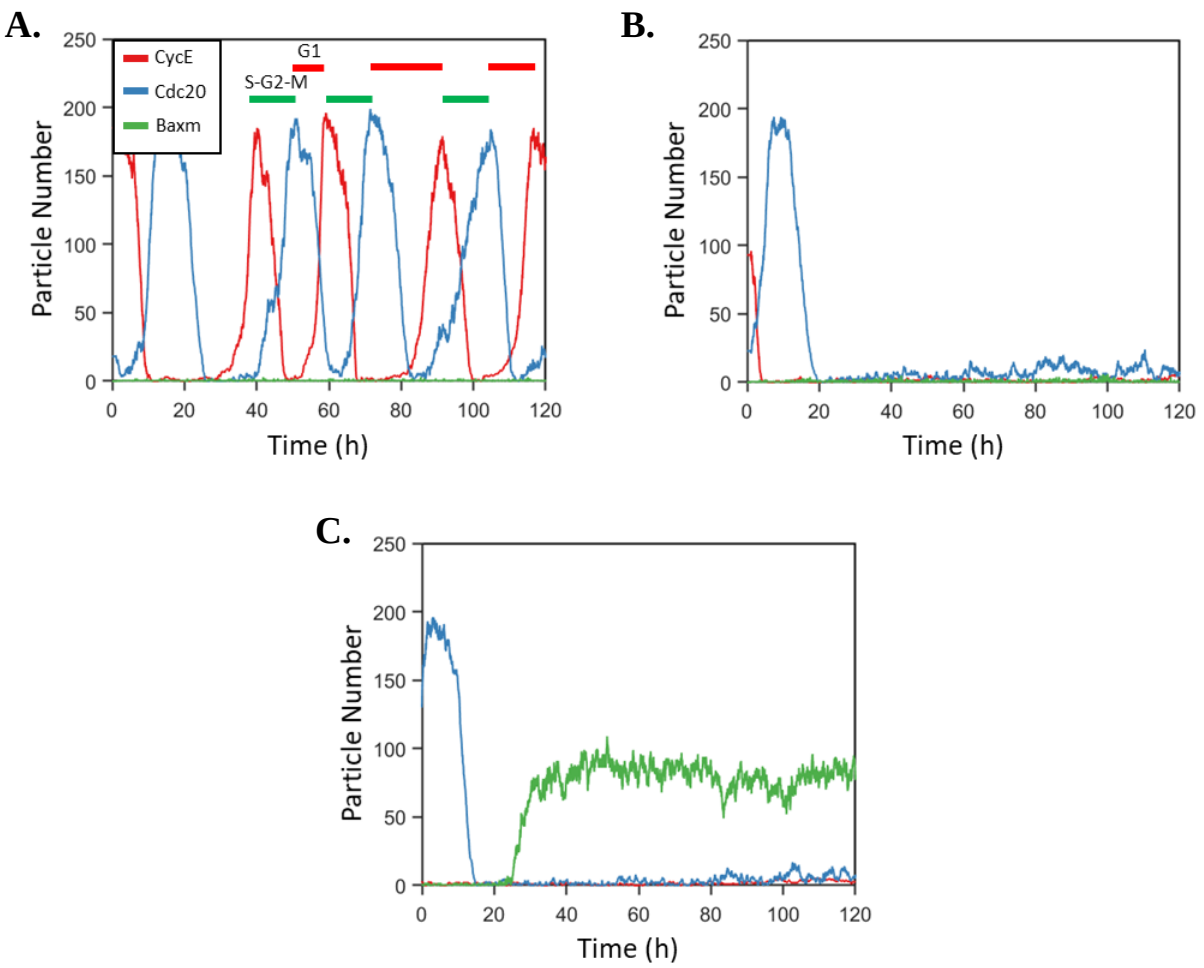


Figure 5.3 Stochastic time course showing a dividing, quiescent, and dying cell. **(A)** Dividing cell. At high model input (high AKT, ERK, and FAK signal), the mitotic module oscillates. Notice the variable time between peaks as opposed to the ODE solution. **(B)** Quiescent cell. At intermediate signal strength, cells exist in a non-cycling and non-apoptotic state. **(C)** Dying cell. At low signal, BAXm is activated, signifying the start of apoptosis.

As with the ODE model, the system can exist in three states: at low, intermediate, and high signal marked by cell death, quiescence, and active cycling, respectively. We first examined how signal strength (through relative AKT, ERK, and FAK levels) affects the behavior of the system. Figure 5.3A shows a representative time course at a high signal (AKT = ERK = FAK =1). The resulting evolution of the system shows regular oscillatory behavior and low BAXm, an apoptotic factor. Notice that the time between peaks, labeled G1 and S-G2-M, are now unequal in contrast

to the ODE solution. The introduced stochastic noise results in a variable intermitotic time. At intermediate signal strength (AKT = 0.4, ERK = 1, FAK = 0.4), shown in Figure 5.3B, the mitotic circuit can cease to oscillate. This is after the mitotic circuit has already completed one cycle. This regime is marked by the system's ability to enter a quiescent state, where there are no oscillations but BAXm has not been activated. The probability that the system will enter such a state is set by the Rb-E2F switch, marking the restriction point at the G1/S transition. At low signal (AKT = ERK = FAK = 0), Figure 5.3C, BAXm signal is high, signifying cell death. The threshold for cell death is set by activation of BH3 by sustained stress signal. As the input strength level is varied, the probability that a cell will exist in one of the particular states varies. At the thresholds between states, where a molecular switch is located, system noise can play an important factor in the overall behavior observed.

5.3. Model – Experiment intermitotic time comparison

An important feature of the stochastic model is the variable intermitotic time shown in Figure 5.3A. Phenomenologically, this matches well with experimentally measured intermitotic times. However, we were interested in how well the cell fate model captures observed intermitotic time distributions. To retrieve intermitotic time distributions, we measured the distance between Cdc20 peaks to CycE peaks for G1, and from CycE peaks to Cdc20 peaks for S-G2-M in the model. Doing so allowed for a direct comparison to the intermitotic times collected experimentally by measuring cellular time with FUCCI on and off.

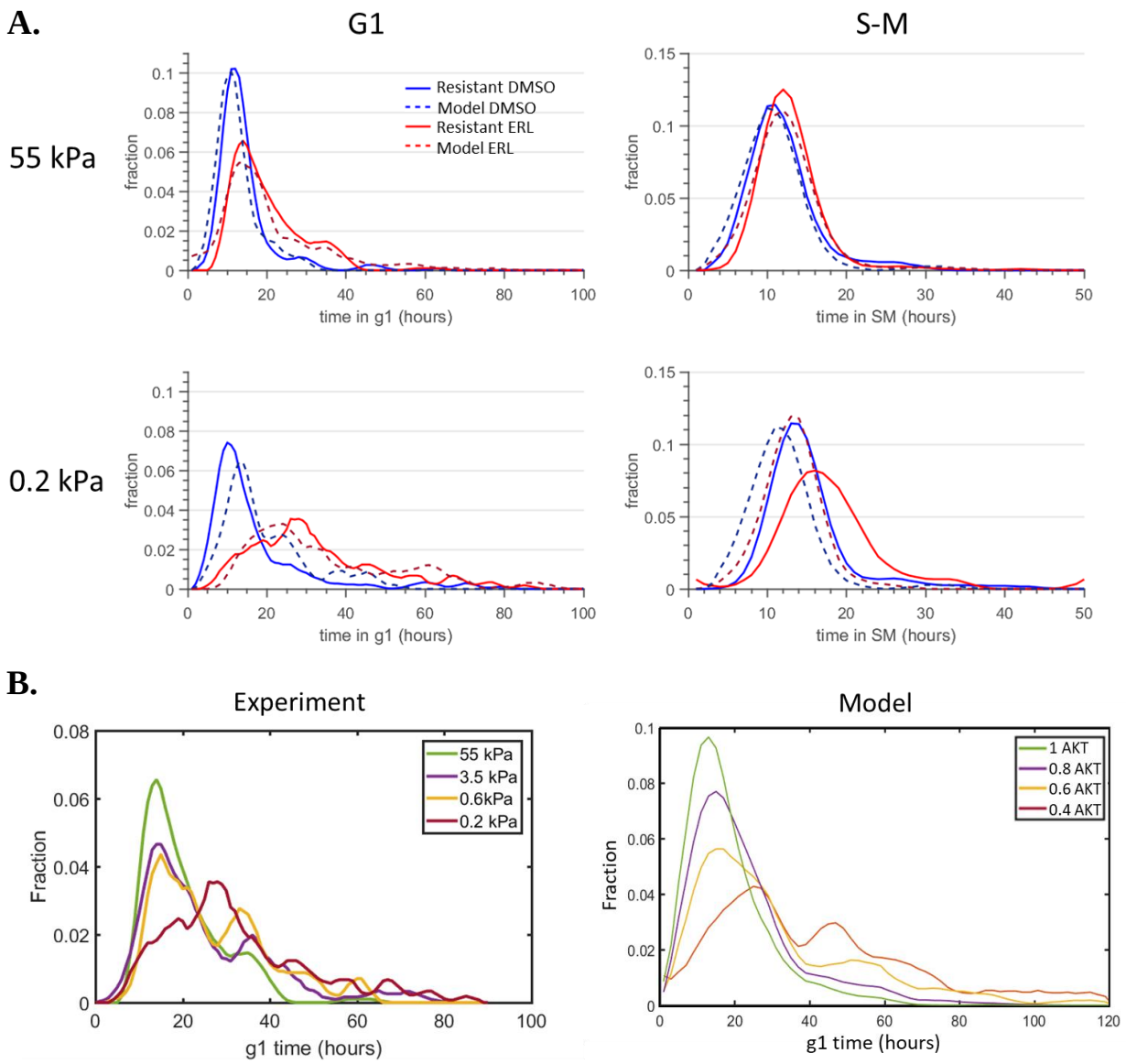


Figure 5.4 Intermitotic time comparison between experiment and model. **(A)** G1 and S-G2-M time comparison between experiment and model for treated and untreated DS8 on 55 kPa and 0.2 kPa stiffness substrate. **(B)** Comparison of DS8 erlotinib treated G1 times with model G1 times.

We find that the model recapitulates the G1 phase well, matching the overall shape observed experimentally. In particular, we note that the model is able to capture the complex bimodal shape of erlotinib-treated cells (Figure 5.4A). In the model, as signal strength is decreased, the number of cells with longer G1 phase increases, similar to experiment (Figure 5.4B). Examining S-G2-M

distribution we find we can initially match experimental observation through the model (Figure 5.4A). However, when signal strength is decreased, analogous to drug treatment or softer substrate, the S-G2-M time does not change. In the model, S-G2-M time varies by only a small magnitude, decreasing slightly with decreased signal.

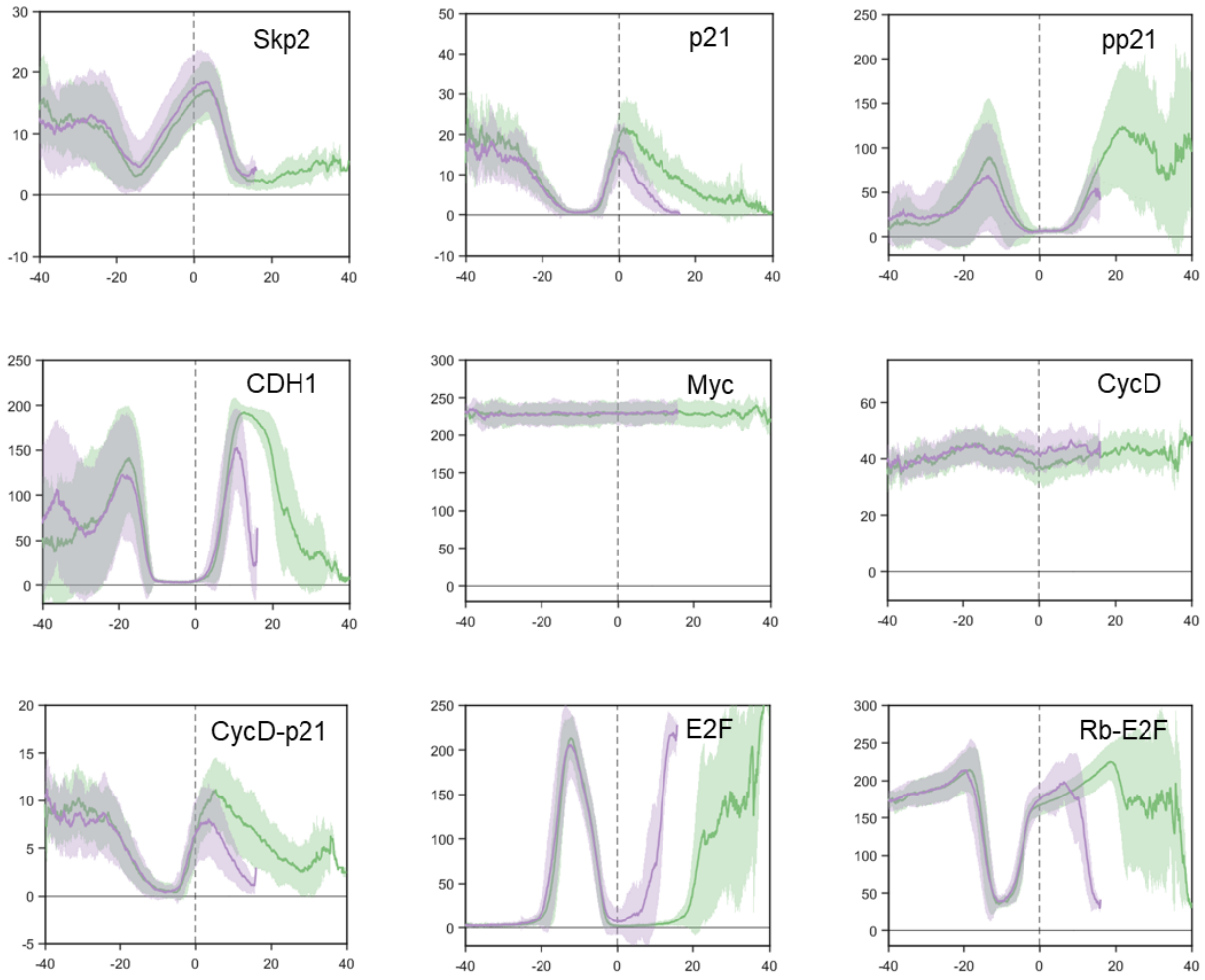
Table 5.1 KS-test results between experiment and model from Figure 5.4A.

	G1	**S-M**
55 kPa DMSO	0.1521	0.1263
55 kPa ERL	0.0760	0.0628
0.2 kPa DMSO	7.4865e-28	6.3821e-17
0.2 kPa ERL	1.5268e-09	4.3079e-17

5.4. Stochastic protein dynamics

We next inquired what the underlying molecular cause was for the mitotic time delay predicted by our model. The model was run at intermediate input strength (AKT = 0.6, ERK = 0.25, FAK = 0.6) and molecular evolution windows were separated based upon mitotic time (G1 < 18 hours and G1 > 24 hours). All traces were then aligned at the initial mitotic event and included 40 hours before and after the event for a total window of 80 hours. There are several interesting observations from the molecular averages. We looked for the first evident difference that linked a change in signaling to a longer mitotic time. Skp2 has consistent elevated concentration prior to short mitotic time. Concomitantly, CDH1 levels are slightly lower for short mitotic time. p21, and its cyclin-bound pool, show identifiable difference between long and short division. p21 and pp21 are elevated in long dividing cells, as are the p21-bound cyclin D, cyclin A, and cyclin B. Cyclin E

stands out from this group as the cyclin E-p21 complex has a higher concentration in quickly dividing cells. Additionally, the cyclin B-p21 complex peaks after mitosis (Cdc20 peak). The Rb-E2F switch components show differential behavior as well. The mitotic inhibitor Rb is increased in long dividing cells and its phosphorylated state shows decreased levels; fractionally, this means an even smaller proportion of inactivated Rb, preventing binding to E2F. E2F has decreased levels in long dividing cells at the mitotic event.



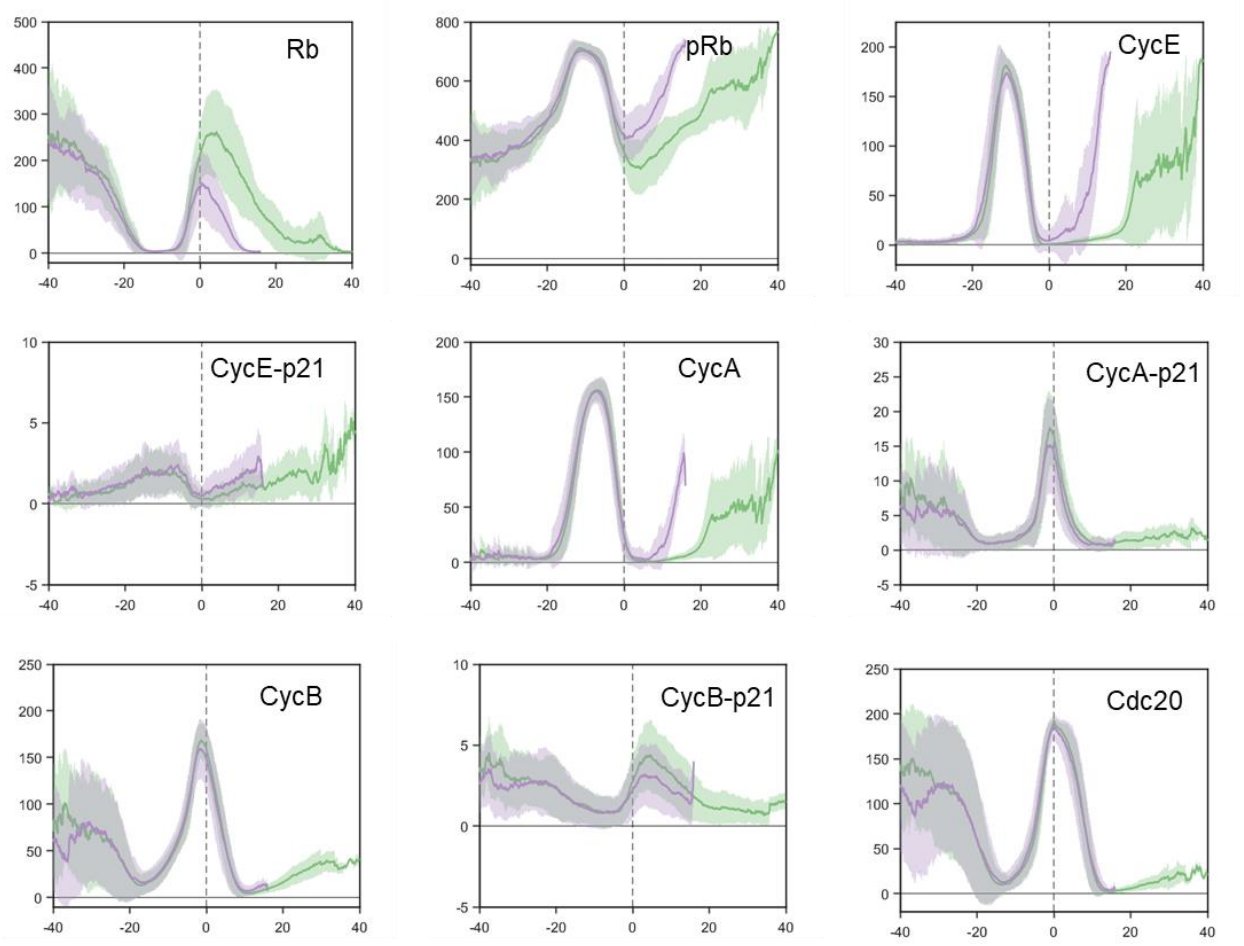


Figure 5.5 Average particle number for all components of the cell cycle with mitotic time greater than 24 hours in green and less than 18 hours in purple. All particle traces are aligned at mitosis, marked by the dashed line corresponding to the peak in Cdc20.

Using cyclin E, cyclin A, and Cdc20 peaks as markers for G1/S transition, S-G2-M time, and the S/M transition respectively, we plotted the limit cycle for the stochastic model and calculated the system dwell time at each phase space coordinate. Comparing the slowly dividing and quickly dividing cells on separate plots allows us to examine where the system is spending the majority of its time and highlights differences in signaling between the two differently behaving states. We notice that the slowly dividing cells spend a large fraction of time in low cyclin E, cyclin A, and Cdc20 state. This would correspond to extended G1 (quiescence). The path to reach the state is

different than the path taken by quickly dividing cells through phase space. For quickly dividing cells, Cdc20 remains high as cyclin E increases. The path from cyclin A to Cdc20 peak remains nearly the same for both conditions. This method of model visualization could be implemented for other complex systems to reveal otherwise less obvious information.

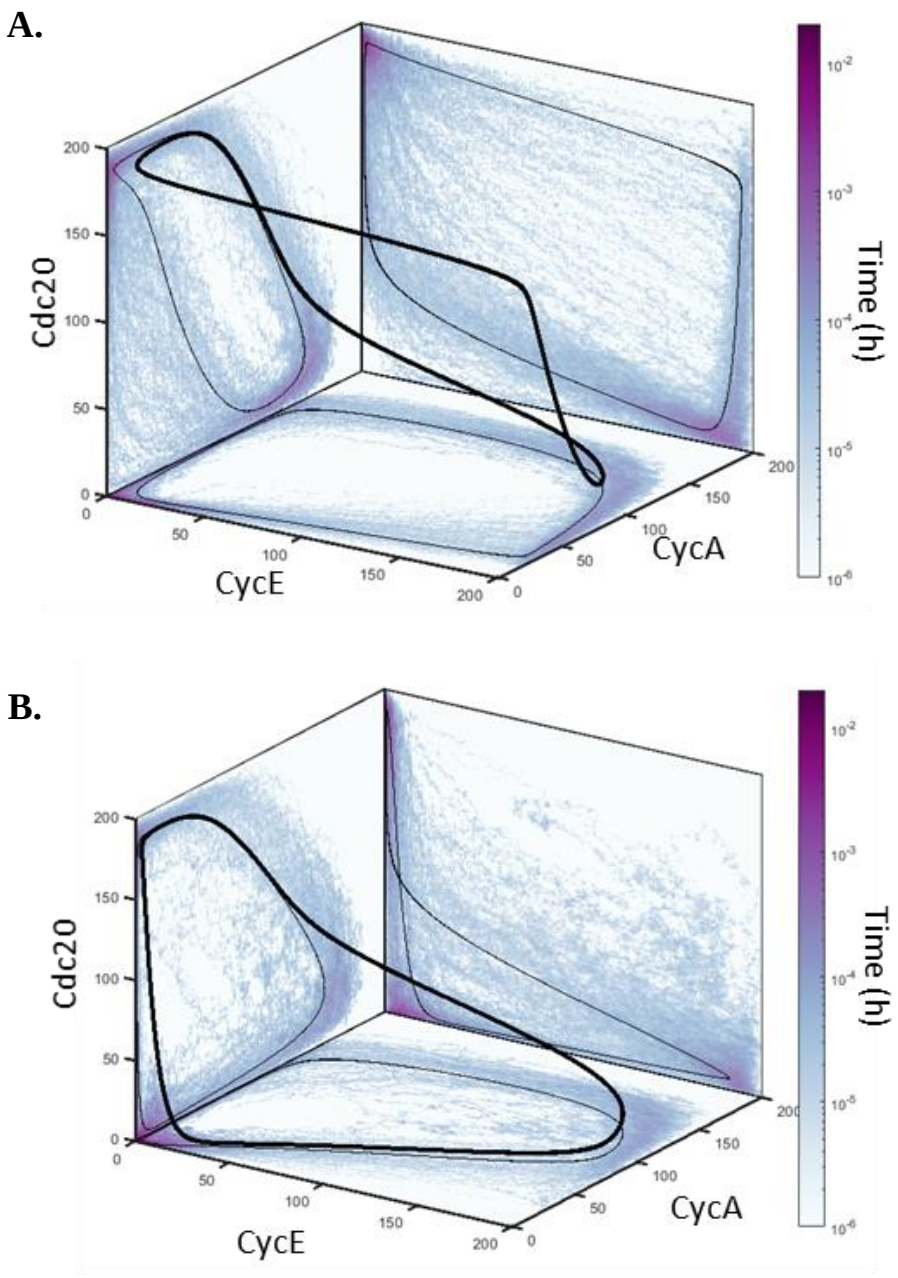


Figure 5.6 Limit cycles of quickly and slowly dividing cells. **(A)** Quickly dividing cells are marked by elevated Cdc20 during activation of CycE. **(B)** Slowly dividing cells show increased dwell time at low CycE, low CycA, and low Cdc20.

5.5. Model p21 inhibition

In our model, p21 and its cyclin-bound species have distinct evolution profiles for quickly and slowly dividing cells. This prompted the hypothesis that mitotic time might depend on p21. We tested this hypothesis by performing an *in silico* knockdown experiment. p21 expression was set to 0, and the stochastic model was run. We then compared the distribution of intermitotic time, time spent in G1, and time spent in S-G2-M. p21 knockdown results in decreased intermitotic time, with a greater proportion of cells populating the primary peak. We found the decreased intermitotic time is caused by a decrease in G1 time, associated with a change in Rb-E2F switch behavior. Surprisingly, there is no change in S-G2-M time. The association of p21 with cyclin A and cyclin B, both components of later stages of the cell cycle, had initially led us to predict S-G2-M variation.

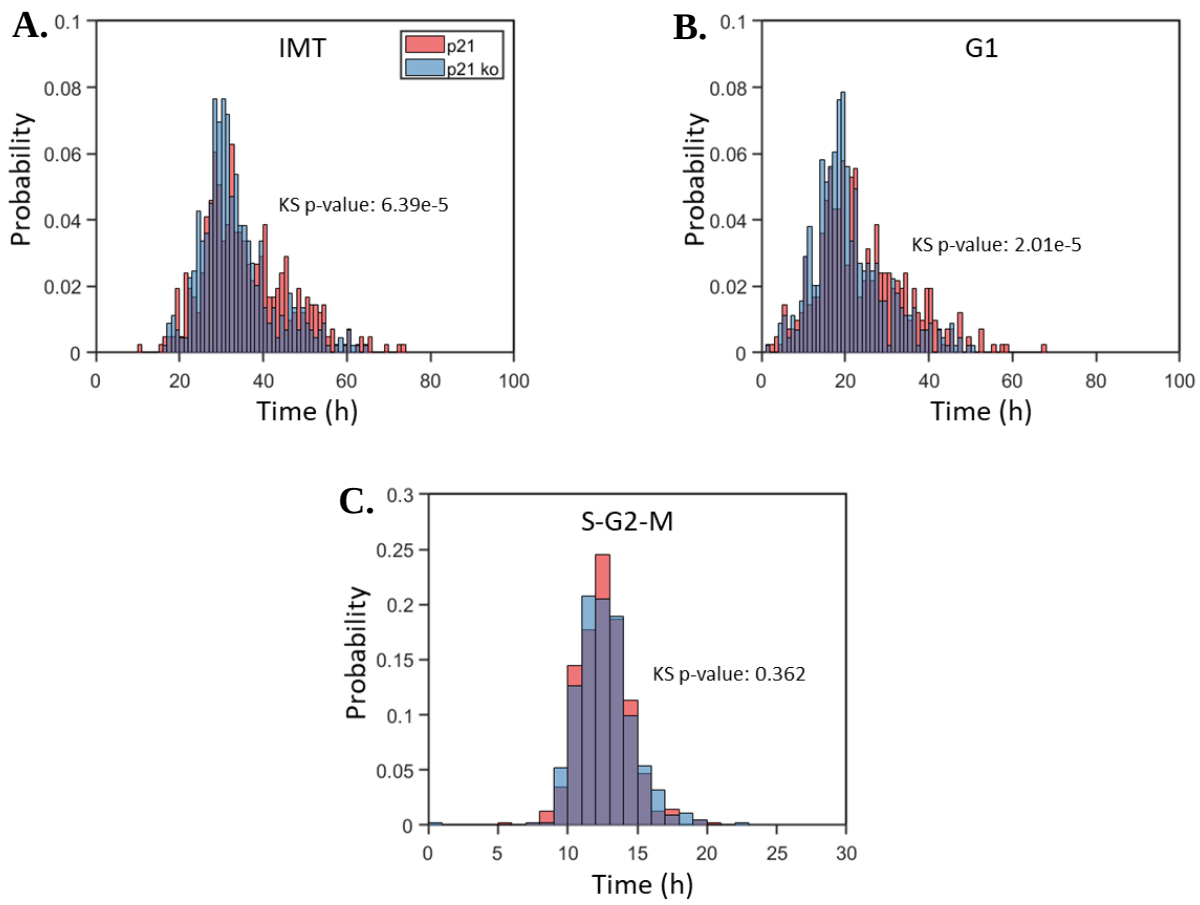


Figure 5.7 Model p21 knockdown. **(A)** Intermitotic time distribution comparison between model p21 wild-type (WT) and p21 knockout (KO). **(B)** G1 time distribution comparison between model p21 WT and p21 KO. **(C)** S-G2-M time distribution comparison between model p21 WT and p21 KO. All distributions include a KS-test value.

5.6. Fractional proliferation analysis of the stochastic cell fate model

By including both a mitotic and apoptotic module, we can perform the same fractional proliferation analysis undertaken for our experimental studies. The model is run to mimic the experimental setup. First, a random starting point along the cell cycle is chosen for each cell based upon the ODE solution. Then, the model is run for each cell for 30 simulation hours and each time course is checked for division or apoptotic events. If a division occurs, a new cell is added at the division time with the same current molecular count as the original cell. If an apoptotic event

occurs, that cell is marked as dead and simulation stops for it. In this manner, we can track the change in overall cell number as function of time. We can also retrieve the intermitotic time distribution and the death rate, all factors used in fractional proliferation.

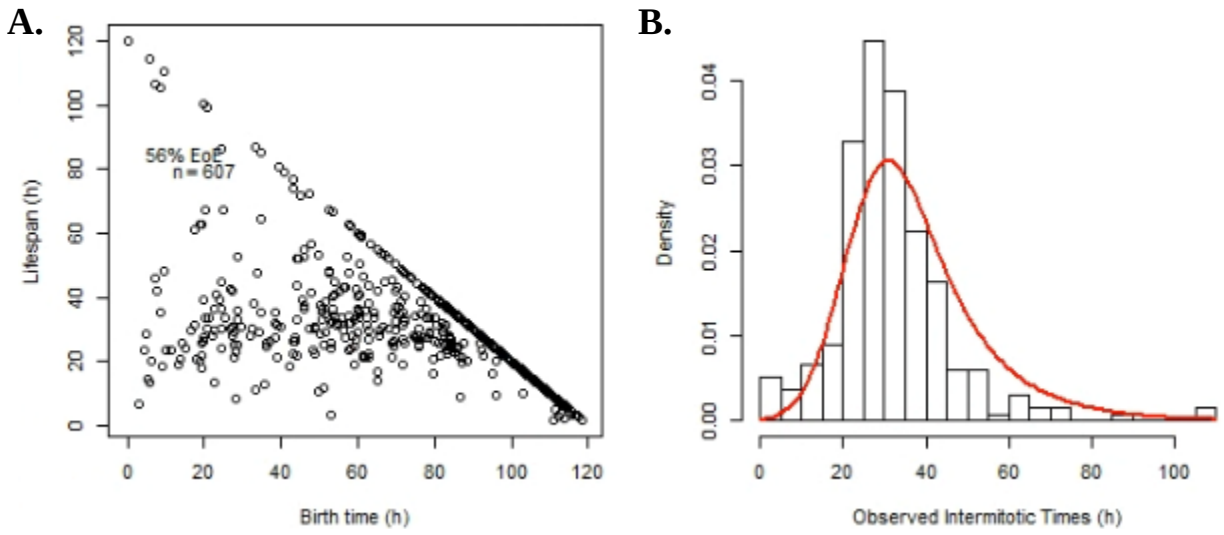


Figure 5.8 Fractional proliferation analysis of model-derived cells. **(A)** Lifespan as a function of birth time. The model is able to generate cells in a quiescent state marked by prolonged IMT time. **(B)** IMT distribution of model-derived cells. Synthetic data was analyzed using the R package *fracprolif*.

To increase the number of 'quiescent' cells in our experiment, we had to decrease the strength of two feedback loops within the model. There exist several positive feedback loops within the cell cycle: the first we modified is the self-activation of cyclin E/Cdk2 through its mutual activation with phosphatase Cdc25 (positive feedback loop at the G1/S transition) and the second is the self-activation of cyclinB/Cdk1 through its mutual activation with phosphatase Cdc25 (positive feedback loop at the G2/M transition). Figure 5.8A shows representative results of a simulated experiment. The model shows several cells reaching the end of the 'experiment', and they are considered quiescent. Using the change in total cell number and intermitotic time distribution shown in Figure 5.8B, we fit the three parameters of the fractional proliferation model. The model

fit shows the majority of cells exist within the dividing population with a small increase in quiescent population. Comparing untreated DS8 cells on 0.2 kPa substrate to the simulated experiment, we observe similar rate terms.

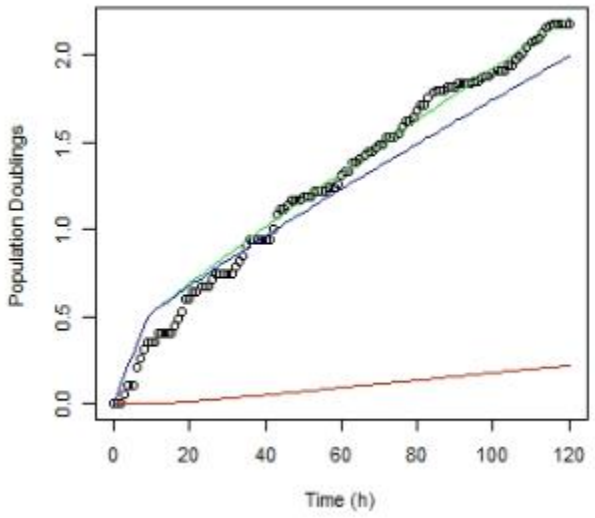


Figure 5.9 Fractional proliferation graph of simulated cell behavior. The green curve is total cell number, the blue represents dividing cells, and the red represents quiescent cells.

Table 5.2 Comparison of fit fractional proliferation rates between model and experiment.

Term	Model	Experiment
d	0.0198	0.0185
q	0.0038	0.0052
a	0.0079	0.0065

5.7. Discussion and conclusion

Fate decision represents an important and common event in cellular behavior. Understanding how that choice is made has important implications for our understanding of cancer progression. Here, we show that it is possible to use a simplified, molecular model to accurately capture single cell behavior and population level dynamics. This model exists as an important middle ground between large, all-encompassing molecular models and simplified mathematical schemes [100][101]. The former can offer powerful insight if properly tuned, but with vast numbers of parameters to estimate, it is difficult to know if the optimal estimation has been achieved. The latter may coarse-grain the biological steps too much, losing important system behaviors. In fact, the level of detail to include in a model to achieve a particular phenomenon answers important questions regarding the signaling topology and overall architecture of the network. We considered these tradeoffs when designing our model, taking cues from the prior and ongoing work of Goldbeter and colleagues. His original work designing a full, descriptive model of the mammalian cell cycle and he subsequently simplified it down to the essential components to achieve phenomenologically similar oscillations. This offers important insight into the control scheme of the cell cycle. Additionally, it makes testing network connections within the model quicker. We use the same approach to begin asking what the necessary components are to reproduce experimentally quantified cellular behavior, intermitotic times, death rates, and drug-induced proliferation rates, given known signaling input. The problem is inherently multiscale in nature, starting with a description of molecular events and eventually describing the change in population.

The first step in the process, achieved over many years of biochemical study, was identifying the components of the cellular signaling system. After, and concurrent with, component identification, many studies have focused on describing how the elements interact. Our work picks

up at this point, using the years of prior work to build a literature-based model. Our model combines the Goldbeter skeleton model that includes positive feedback leading to bistable transitions and combines it with the Rb-E2F bistable switch characterized by Yao et al. [48], [102]. Thereby, we include the necessary components and multiple feedback loops necessary for complex system behavior.

An important control mechanism of treatment response was substrate stiffness. We modeled stiffness input as FAK activity in our model. FAK both helps set AKT level, measured experimentally, and controls the level of p21 through Skp2 activation. p21 in turn binds to cyclin complexes, effectively removing them from the available pool to drive the cell cycle. Cyclin-dependent kinase inhibitors, such as p21, can effectively block the progression of the of the cell cycle by changing the dynamics of the Rb-E2F switch [103]. By including the mechanism of molecular bistable switches, we were able to construct a system that exists in three regimes: a high signal actively oscillating state, an intermediate signal quiescent steady state, and low signal apoptotic steady state.

We characterized the nature of the oscillations and inquired if a simple, molecular model can capture the intermitotic time distributions observed experimentally. The stochastic model accurately recapitulates the trends of our experimental results. Comparing the distributions of experimental G1 and S-G2-M to those of the model, we find the same right-skewed distribution for G1 and more Gaussian distribution for S-G2-M. The model captures the signal-dependent shift of G1, with mean time shifting rightward as signal decreases. Surprisingly, we find the same multimodal structure present in our model results as seen in our experiments.

The existence of such structures has been noted historically as ‘cell cycle quantization’ and been found across eukaryotic species [95], [104], [105]. Prior work has focused on describing these

structures in yeast mutant cells (wee1- cdc25Δ) [94]. Our work builds upon these studies, demonstrating that if positive feedback loops that accelerate the G2/M transition are weak, then cells can revert back to G2, effectively adding a quantized amount of time to the cell cycle. Instead of extended G2-M mitotic time, we find a quantized G1 time that is dependent on signal strength. This leads to questions of the similarity between the circuits and how a change in signal strength leads to complex oscillations. We find the quantization is the result of the coupling of multiple oscillatory circuits within the cell cycle.

We further note that the slowly and quickly dividing states are transient within our model and a result of molecular noise, rather than a specific genetic expression level. We compare the molecular evolution profiles of the two groups and find interesting differences between the two. The first feature we point out is a consistently decreased level of Skp2 signaling in long cycling cells. When Skp2 is low, inhibition of p21 is also low, leading to a quicker cell cycle. Skp2 has been shown to be necessary for the degradation of cyclin-dependent kinase inhibitor (CKI) for proper cell cycle progression [106]. Our model further supports experimental findings by Meyer et al. showing the proliferation-quiescence decision is dependent on Cdk2 activity at mitotic exit [107]. We model cyclin E in complex with Cdk2 and observe increased activity of the complex at mitosis for quickly dividing cells. It is difficult to say our model is dependent on the activity of one protein at a specific point in time to cause a specific outcome. The entire state of the system, the molecular count, at a specific time is what dictates the evolution of our system. However, we can say, given a set of molecular counts, the likelihood of the model behaving a certain way.

As a means of further testing our model, we examine the effects of what would be the experimentally analogous process of p21 knockout. By setting the p21 synthesis to 0, we can compare model dynamics to our ‘wildtype’ experiment. We see that with p21 knockout, there is a

large increase in quickly dividing cells. The change in intermitotic dynamics comes from a shortened time in G1, as opposed to S-G2-M. We find this surprising since p21 interacts with all cyclin complexes. The role of p21 in cell cycle progression can be further explored by changing its association and dissociation constant with the cyclin family. An interesting note is that the cyclin E/p21 complex is the only cyclin/p21 associated complex that shows a greater molecule count that correlates to quickly dividing cells.

With the model established, we can begin to compare directly how changes at the molecular level manifest as overall cell population changes. This is an important link that could enable molecular measurements to be used as predictors of treatment outcome. Additionally, understanding cellular control mechanisms can provide insight into novel mechanisms and drug targets.

5.8. Future studies

Our results suggest our model contains enough detail to be able to reproduce complex cellular behavior. We seek to continue testing the limits of the model as we extend it further by including additional components. We envision the following next steps:

5.8.1. PySB implementation

The first step in future computational studies will be transferring the model and approach to PySB and Python. PySB offers a unified framework for building a rule-based model with the idea of creating reusable, modular signaling fragments [108]. Doing so would allow us to use already established modules and to examine their effects on cell fate decision. It would also be helpful to split the cell cycle into its component oscillating circuits. Additionally, it would allow us to easily

examine the importance of the strength of the coupling between circuits and the effect on population response. It has been previously shown how weak coupling can lead to complex mitotic oscillations[109]. Python also offers easily programmable cost functions, allowing us to quickly match data as we add additional modules.

5.8.2. Additional modules and signaling pathways

There exist a multitude of factors that influence the cell cycle and apoptotic pathway, including factors that directly link the two [110], [111]. It would be insightful to examine how such a link affects both signaling pathways and how the connection can be modulated to favor a particular behavior. Signaling upstream of AKT and ERK are additional modules we would like to examine. There have been recent suggestions of oscillatory behavior in both AKT and ERK [112], [113]. Examining the effects of an oscillatory circuit upstream of the cell cycle could add significant value to our current system. In recent years, there has been an increase in live-cell protein reporters. Incorporating these novel tools would allow for improved kinetic data that could help refine the model. An example of one such tool is the Cdk2 reporter used by Meyer et al. [107].

5.8.3. Drug Synergy

Combination targeted therapy has seen increased efficacy in NSCLC treatment compared single drug treatments [114]. Most patients develop drug resistance to first- generation TKIs (erlotinib, gefitinib). Common modes of resistance include TK domain mutation (T790M), MET amplification, and RAS mutation [115], [116]. Next steps would include how MET amplification and RAS mutation change signaling upstream of the cell cycle machinery and how those changes alter cell fate decision. As the model is capable of mimicking cell population behavior, we see its

possible use as an aid to define drug synergy. Recent work in the Quaranta laboratory has focused on defining the synergistic effects of drug combination therapy [117]. Essentially, an *in silico* dose-response experiment of multiple drug combinations could be run.

5.9. Methods

5.9.1. Statistical methods

To determine whether two distributions were indistinguishable we used the two-sample Kolmogorov-Smirnoff test. This test determines the probability that two sets of data are sampled from the same distribution, and $P < 0.05$ was assumed to be statistically significant evidence that the two data sets are drawn from different distributions.

5.9.2. Cell fate model pseudo-code

All models were written and run in MATLAB. To derive cell population behavior from the stochastic cell cycle and apoptotic modules we perform the following steps:

1. Initialize the system by running the full ODE Cell Fate Model with proper AKT, ERK, and FAK levels as input strength. For drug treated cells, this corresponds to initial seeding and growth prior to treatment.
2. Select a random starting point for each cell along the ODE solution and convert from concentration to molecule number. If cells in the model are drug treated, levels of AKT, ERK, and FAK are changed.
3. Run the Gillespie algorithm on each cell with the end state as the initial condition for 30 computational hours.

4. Check for division or death events within the 30-hour window. Add a cell at the time of division with initial conditions set equal to the original cell at the mitotic point. Stop running for cells that have an apoptotic event.
5. Update the overall cell count and repeat from step 3.

5.9.3. Fractional proliferation

Fractional proliferation analysis was performed as in the previous chapter using the available R package *fracprolif*.

CHAPTER VI

6. Conclusion and Final Remarks

Overall, the broad goal of this work was to examine how external stimuli (presence of drug and substrate stiffness) are integrated by the cell into a fate decision. We built a literature-based model to understand how the decision is made and, ultimately, how a population of cells will behave. This work first focuses on examining experimental studies of the effect of cancer cell microenvironment on chemotherapeutic response and connecting those results with changes in key regulators of the mitotic and apoptotic processes within the cell. We build a literature-based molecular-level model to make the connection. Using the model as a tool, we can query the components and system topology necessary to produce complex cellular phenomena. The work helps establish the mathematical framework for understanding cellular signaling pathways. By understanding the mathematical rules that produce complex cellular responses, novel treatments can be developed more quickly in addition to adding insight into the underlying biology.

We used an EGFR mutated non-small cell lung cancer cell line, PC9, as a model system. It offers an opportunity to practically apply newly developed understanding of cellular signaling regulation to develop improved treatment options. In the past, targeted therapy of EGFR has been a successful initial treatment for patients with non-small cell lung cancer. The disease, however, will generally progress as the tumors become resistant. One cause of resistance is linked to heterogeneity of cell fates in response to drug treatment. Underlying fate decision, and ultimately population response, is the complex signaling network of the cell. To understand the dynamics of a drug response, it is necessary to determine how an external stimulus is integrated into a fate

decision – in our case either division, apoptosis, or quiescence – and how perturbations in the signaling network alter that decision process. Prior work has shown that variable expression of cMet correlates with cell response. In this study, we find that extracellular matrix (ECM) stiffness also modulates proliferation behavior and the distribution of cell fates. Using these results, we have built a stochastic model capable of accurately predicting population dynamics – validated by *in vitro* evaluation of PC9 sublines. The model and its assessment of cell fate decision processes could be a useful tool for quantitative prediction of drug response.

In this work, we experimentally examined the role of microenvironment in cell response to treatment. We focused on characterizing PC9 response to targeted therapy. We used previously described metrics, primarily the drug induced proliferation (DIP) rate and intermitotic time, to define cell response. Experimental results were then used to refine and validate predicted outputs from the molecular model. We have measured the DIP rate of PC9 with varying substrate stiffness, drug treatment, and sensitivity. We found that a soft substrate leads to a decreased DMSO growth rate and increased erlotinib sensitivity in a cell line dependent manner. We have recorded changes in G1 and S-G2-M time distributions with varying substrate stiffness, drug treatment, and cell line. Our results show that erlotinib treatment strongly affects G1 timing, whereas substrate stiffness primarily impacts S-M time. We have recorded phosphorylated and total protein levels through Western blot densitometry of key signaling proteins within the model. Erlotinib drastically decreases ERK signaling and moderately affects AKT. Stiffness impacts AKT signaling with no discernable effect on ERK and strongly modulates FAK activation.

We next surveyed the extent of treatment response that is captured by known minimal protein network models. We have developed a stochastic model of tumor cell population dynamics as a function of experimentally measurable inputs including local drug concentration, cMet expression,

and substrate stiffness. The model is motivated by our experimental results examining the DIP rate of PC9 sublines under erlotinib treatment. The model is built around three modules. The first, based upon the work of Goldbeter et al., governs the activation of cyclin-dependent kinases (CDKs) and subsequently, proliferation. The second module controls apoptosis and is comprised of the BCL-2 family of proteins. Finally, we have included a direct mapping between initial cell surface signal (EGFR, cMet, integrins) and AKT and ERK that drive the previous two modules. The model has three regimes of response including a proliferative state, an apoptotic state, and a quiescent state based upon the initial signal. We have built a model that can integrate external cues (e.g. matrix stiffness, growth factors) into a cell fate decision by combining existing models in the literature. We inform our model with our experimental results. The model recapitulates the overall population dynamics, measured both in DIP rate and intermitotic time distribution. Interestingly, it even captures complex behavior such as period doubling events.

This work demonstrates the capacity of a systems biology approach highlighted in section 2.8. We have refined a molecular model of cell fate decision to capture complex cellular phenomena. A major portion of the power of a systems biology approach comes from developing a mathematical formalism to express complex systems. With such a formalism comes questions of broader applications. Now that we are beginning to understand how the signaling within a cell functions, where else can we apply the knowledge? A clear extension is the application to synthetic biology. Synthetic biology is a broad field that seeks to design novel biological components and systems. Such applications include diagnostic circuits for detection, treatment, and disease prevention [114]–[116]. Another future direction is investigation of the scalability of the molecular system principles. Do the same control rules apply at a cellular or larger scale? Applying the

principles gathered from molecular signaling analyses could yield insight into other biological systems and complex control systems in general.

Appendix

A. Michaelis-Menten parameter values and initial conditions

The initial conditions and parameter values that were used in the ODE and stochastic solution of the Michaelis-Menten equation:

Table A.1 Model parameters

Variable	**Value**
k_f	2
k_r	1
k_{cat}	1.5

Initial conditions: [S] = 8, [E] = 4, [ES] = 0, [P] = 0.

B. Rb-E2F switch ODEs, parameter values, and initial conditions

The variables with biological description, differential equations, parameter values, and initial conditions used to model the Rb-E2F switch adapted from [48]. The model was implemented using MATLAB and solved with ode23 solver.

Table B.1 Model variables.

Variable	Description
S	Serum; growth signals
M	Myc
E	E2F
CD	CycD
CE	CycE
R	RB
RP	Phosphorylated Rb
RE	Rb-E2F complex

Table B.2 Equations for mathematical model of Rb-E2F pathway.

$\frac{d[M]}{dt} = \frac{k_M[S]}{K_S + [S]} - d_M[M]$	(B.1)
$\frac{d[CD]}{dt} = k_E\left(\frac{M}{K_E + [M]}\right)\left(\frac{[E]}{K_E + [E]}\right) + \frac{k_b[M]}{K_M + [M]} + \frac{k_{p1}[CD][RE]}{K_{CD} + [RE]} + \frac{k_{P2}[CE][RE]}{K_{CE} + [RE]} - d_E[E] - k_{RE}[R][E]$	(B.2)
$\frac{d[CD]}{dt} = \frac{k_{CD}[M]}{K_M + [M]} + \frac{k_{CDS}[S]}{K_S + [S]} - d_{CD}[CD]$	(B.3)

$\frac{d[CE]}{dt} = \frac{k_{CE}[E]}{K_E + [E]} - d_{CE}[CE]$	(B.4)
$\frac{d[R]}{dt} = k_R + \frac{k_{DP}[RP]}{K_{RP} + [RP]} - k_{RE}[R][E] - \frac{k_{P1}[CD][R]}{K_{CD} + [R]} - \frac{k_{P2}[CE][R]}{K_{CE} + [R]} - d_R[R]$	(B.5)
$\frac{d[RP]}{dt} = \frac{k_{P1}[CD][R]}{K_{CD} + [R]} + \frac{k_{P2}[CE][R]}{K_{CE} + [R]} + \frac{k_{P1}[CD][RE]}{K_{CD} + [RE]} + \frac{k_{P2}[CE][RE]}{K_{CE} + [RE]} - \frac{k_{DP}[RP]}{K_{RP} + [RP]} - d_{RP}[RP]$	(B.6)
$\frac{d[RE]}{dt} = k_{RE}[R][E] - \frac{k_{P1}[CD][RE]}{K_{CD} + [RE]} - \frac{k_{P2}[CE][RE]}{K_{CE} + [RE]} - d_{RE}[RE]$	(B.7)

Table B.3 Model parameters.

Rate Constants		**Parameter values, sources, and notes**
k_E	0.4 μM/hr	These values were adjusted together so that: (1) The maximum E2F level is higher than the maximum CycD level (based on our experimental observations) (2) E2F activation threshold is about 1% serum (3) The simulated E2F level will be around the corresponding Michaelis-Menten parameter (K_E)
k_M	1.0 μM/hr	
k_{CD}	0.03 μM/hr	
k_{CDS}	0.45 μM/hr	
k_R	0.18 μM/hr	
k_{RE}	180 μM/hr	
k_b	0.003 μM/hr	
K_S	0.5 μM/hr	
k_{CE}	0.35 μM/hr	Assumed to be similar to k_{EF}
d_M	0.7/hr	Myc half-life = 60 min
d_E	0.25/hr	E2F half-life = 2~3 hr
d_{CD}	1.5/hr	CycD half-life = 25~30 min
d_{CE}	1.5/hr	CycE half-life = 30 min
d_R	0.06/hr	Rb half-life = 12 hours
d_{RP}	0.06/hr	Assumed to be the same as d_R
d_{RE}	0.03/hr	Rb-E2F half-life = 6 hours: The Rb-E2F complex assumed to be more stable than Rb alone
k_{P1}, k_{P2}	18/hr	Typical value phosphorylation rate constant is 3600/hr
k_{DP}	3.6 μM/hr	Typical value for dephosphorylation rate assuming a constant phosphatase concentration is 720 μM/hr
K_M	0.15 μM/hr	Estimated based on measured Myc/Max –DNA dissociation constant

K_E	0.15 μM/hr	Assumed to be the same as K_M
K_{CD}	0.92 μM/hr	Experimentally measured
K_{CE}	0.92 μM/hr	Assumed to be the same as CycD
K_{RP}	0.01 μM/hr	Typical value for Michaelis-Menten parameter for dephosphorylation

Initial conditions: [Rb] = 0.55 μM; [Rbp] = [E2F] = [Myc] = [CycD] = [CycE] = 0 μM

C. Mitotic oscillating circuit ODEs, parameter values, and initial conditions

The variables with biological description, differential equations, parameter values, and initial conditions used to model cell cycle adapted from [56]. The model was implemented using MATLAB and solved with ode23 solver.

Table C.1 Model variables.

Variable	Description
Md[a]	Cyclin D/Cdk4-6 complex
E2F	Transcription factor E2F
Me	Cyclin E/Cdk2 complex
Ma	Cyclin A/Cdk2 complex
Mb	Cyclin B/Cdk2 complex
Cdc20	Protein that belongs to the anaphase-promoting complex (APC)

[a]Md reaches steady state quickly, effectively becoming a constant.

Table C.2 Equations for the Goldbeter skeleton model.

$\frac{d[Md]}{dt} = v_{sd}\left(\frac{GF}{K_{gf}+GF}\right) - V_{dd}\left(\frac{Md}{K_{dd}+Md}\right)$	(C.1)
$\frac{E2F}{dt} = V_{1e2f}\left(\frac{(E2F_{tot}-E2F)}{K_{1e2f}+(E2F_{tot}-E2F)}\right)(Md+Me) - V_{2e2f}\left(\frac{E2F}{K_{2e2f}}+E2F\right)Ma$	(C.2)
$\frac{d[Me]}{dt} = v_{se}E2F - V_{de}Ma\left(\frac{Me}{K_{de}+Me}\right)$	(C.3)
$\frac{d[Ma]}{dt} = v_{sa}E2F - V_{da}Cdc20\left(\frac{Ma}{K_{da}+Ma}\right)$	(C.4)
$\frac{d[Mb]}{dt} = v_{sb}Ma - V_{db}Cdc20\left(\frac{Mb}{K_{db}+Mb}\right)$	(C.5)

$$\frac{d[Cdc20]}{dt} = V_{1cdc20}Mb \times \left(\frac{(Cdc20_{tot} - Cdc20)}{K_{1cdc20} + (Cdc20_{tot} - Cdc20)}\right) - V_{2cdc20}\left(\frac{Cdc20}{K_{2cdc20} + Cdc20}\right)$$	(C.6)
$$\mathrm{Md} = \frac{K_{dd}v_{sd}GF/(K_{gf} + GF)}{V_{dd} - (v_{sd}GF/(K_{gf} + GF))}$$	(C.7)

Table C.3 Key model parameters.

Rate Constants		**Parameter values, sources, and notes**
$Cdc20_{tot}$	5 μM	Total concentration of the protein Cdc20
$E2F_{tot}$	3 μM	Total concentration of the transcription factor E2F
GF	1 μM	Growth factor
K_{da}	0.1 μM	Michaelis constant for the degradation, activated by Cdc20, of cyclin A/Cdk2
K_{db}	0.005 μM	Michaelis constant for the degradation, activated by Cdc20, of cyclin B/Cdk1
K_{dd}	0.1 μM	Michaelis constant for the degradation of cyclin D/Cdk4–6
K_{de}	0.1 μM	Michaelis constant for the degradation of cyclin E/Cdk2
K_{gf}	0.1 μM	Michaelis constant for synthesis of the cyclin D/Cdk4–6 complex induced by growth factor
K_{1cdc20}	1 μM	Michaelis constant for Cdc20 activation through phosphorylation by cyclin B/Cdk1
K_{2cdc20}	1 μM	Michaelis constant for Cdc20 inactivation through dephosphorylation
K_{1e2f}	0.01 μM	Michaelis constant for E2F activation by cyclin D/Cdk4–6 and cyclin E/Cdk2 complexes
K_{2e2f}	0.01 μM	Michaelis constant for E2F inactivation by the cyclin A/Cdk2 complex
V_{da}	0.245/hr	Rate constant for the degradation of the cyclin A/Cdk2 complex by the protein Cdc20
V_{db}	0.28/hr	Rate constant for the degradation of the cyclin B/Cdk1 complex by the protein Cdc20
V_{dd}	0.245/hr	Maximum degradation rate of cyclin D/Cdk4–6 complex
V_{de}	0.35/hr	Rate constant for the degradation of cyclin E/Cdk2 by cyclin A/Cdk2
v_{sa}	0.175/hr	Rate constant for synthesis of cyclin A/Cdk2 induced by the transcription factor E2F

v_{sb}	0.21/hr	Rate constant for synthesis of cyclin B/Cdk1 induced by cyclin A/Cdk2
v_{sd}	0.21/hr	Rate constant for synthesis of cyclin D/Cdk4–6 induced by growth factor GF
v_{se}	0.21/hr	Rate constant for synthesis of cyclin E/Cdk2 induced by the transcription factor E2F
V_{1cdc20}	0.21/hr	Rate constant for activation of Cdc20 through phosphorylation by cyclin B/Cdk1
V_{2cdc20}	0.35 µM//hr	Maximum rate constant for inactivation of Cdc20 through dephosphorylation
V_{1e2f}	0.805/hr	Rate constant for activation of E2F by cyclin D/Cdk4–6 and cyclin E/Cdk2 complexes
V_{2e2f}	0.7/hr	Rate constant for inactivation of E2F by the cyclin A/Cdk2 complex

D. Cell fate decision parameter values and initial conditions

The parameter and initial conditions used to for the cell fate model. Parameter were set from the original model and varied to match experimental data.

Table D.1 Cell fate model parameters.

Rate Constants		Description
Ω	200 10^6 L/N_A	System size for stochastic simulations used to convert from concentration in the ODE system to molecule count in the stochastic system
s	0.37	Global scaling parameter
G1 – Rb-E2F switch		
k_E	0.4 /h	Rate constant for E2F synthesis through autoactivation
k_M	0.13 /h	Rate constant for Myc synthesis through AKT pathway
k_{CD}	0.03 /h	Rate constant for cyclin D activation from Myc
k_{CDS}	0.23 /h	Rate constant for cyclin D activation through AKT pathway
k_R	0.18 μM/h	Basal Rb synthesis rate
k_{RE}	180 /μM h	E2F – Rb binding constant
k_b	0.003 μM/h	Rate constant for E2F through Myc
K_S	0.5 μM	Michaelis constant for Myc synthesis through AKT pathway
d_M	0.6 /h	First-order rate constant for non-specific Myc degradation
d_E	0.25 /h	First-order rate constant for non-specific E2F degradation
d_{CD}	2 /h	First-order rate constant for non-specific cyclin D inactivation
d_R	0.06 /h	First-order rate constant for non-specific Rb degradation
d_{RP}	0.06 /h	First-order rate constant for non-specific pRb degradation
d_{RE}	0.03 /h	First-order rate constant for non-specific Rb-E2F degradation
k_{P1}	18 μM/h	Rate constant for Rb phosphorylation through cyclin D
k_{P2}	18 μM/h	Rate constant for Rb phosphorylation through cyclin E
k_{DP}	3.6 μM/h	Rate constant for pRb dephosphorylation
K_M	0.13 μM	Michaelis constant for E2F synthesis through Myc
K_E	0.1 μM	Michaelis constant for E2F synthesis through autoactivation
K_{CD}	0.92 μM	Michaelis constant for Rb-E2F dissociation from cyclin D mediated Rb phosphorylation

K_{CE}	0.92 μM	Michaelis constant for Rb-E2F dissociation from cyclin E mediated Rb phosphorylation
K_{RP}	0.01 μM	Michaelis constant for pRb dephosphorylation
K_{MY}	0.1 μM	Michaelis constant for Myc synthesis through ERK pathway
k_{EMY}	0.15 μM/h	Rate constant for Myc synthesis through ERK pathway
k_{ED}	0.22 μM/h	Rate constant for cyclin D activation through ERK pathway
p21 – CKI control		
V_{dskp2}	1.1 μM/h	Rate constant for Skp2 degradation through Cdh1
V_{1p21}	100 μM/h	Rate constant for p21 phosphorylation from cyclin E
V_{2p21}	0.1 μM/h	Rate constant for pp21 dephosphorylation
V_{dp21p}	5 μM/h	Rate constant for pp21 degradation from Skp2
V_{1cdh1}	1.25 μM/h	Rate constant for Cdh1 activation
V_{2cdh1}	8 μM/h	Rate constant for Cdh1 inactivation through cyclin A and cyclin B
v_{s1p21}	0.5 μM/h	Rate constant for basal p21 synthesis
v_{s2p21}	0.1 μM/h	Rate constant for p21 synthesis through E2F
K_{1cdh1}	0.1 μM	Michaelis constant for Cdh1 activation
K_{2cdh1}	0.1 μM	Michaelis constant for Cdh1 inactivation through cyclin A and cyclin B
K_{dp21p}	0.1 μM	Michaelis constant for pp21 degradation through Skp2
$K_{dp21skp2}$	0.1 μM	Michaelis constant for pp21 degradation through pp21 interaction with Skp2
K_{cdh1}	0.4 μM	Michaelis constant for Skp2 degradation from Cdh1
K_{dskp2}	0.5 μM	Michaelis constant for Skp2 degradation from Skp2 with Cdh1
K_{de}	0.1 μM	Michaelis constant for cyclin E inactivation from Skp2
K_{1p21}	0.5 μM	Michaelis constant for p21 phosphorylation from cyclin E
K_{2p21}	0.5 μM	Michaelis constant pp21 dephosphorylation
k_{ddp21}	0.06 /h	First order rate constant for non-specific p21 degradation
k_{ddskp2}	0.005 /h	First order rate constant for non-specific Skp2 degradation
k_{c1}	3 /μM h	Association constant for cyclin D and p21
k_{c2}	1 /h	Dissociation constant for cyclin D and p21
k_{c3}	8 /μM h	Association constant for cyclin E and p21
k_{c4}	2 /h	Dissociation constant for cyclin E and p21
k_{c5}	5 /μM h	Association constant for cyclin A and p21
k_{c6}	1 /h	Dissociation constant for cyclin A and p21
k_{c7}	0.12 /μM h	Association constant for cyclin B and p21
k_{c8}	0.2 /h	Dissociation constant for cyclin B and p21
V_{Fp}	1.5 μM/h	Rate constant for p21 phosphorylation from FAK
K_{Fp}	0.1 μM	Michaelis constant for p21 phosphorylation from FAK
k_{CDF}	0.21 μM/h	Rate constant for cyclin D activation from FAK

K_{DF}	0.8 μM	Michaelis constant for cyclin D activation from FAK
a	0.8 μM/h	basal term for activation of cyclin E/Cdk2 (without PF loop = 0)
S-G2-M		
$Cdc20_{tot}$	1 μM	Total concentration of Cdc20
$cdh1_{tot}$	1 μM	Total concentration of Cdh1
$CycE_{tot}$	1 μM	Total concentration of cyclinE/Cdk2
$CycA_{tot}$	1 μM	Total concentration of cyclinA/Cdk2
$CycB_{tot}$	1 μM	Total concentration of cyclinB/Cdk1
K_{1E2F}	1 μM	Michaelis constant for E2F activation by cyclinD/Cdk4-6 and cyclinE/Cdk 2 complexes
K_{2E2F}	0.5 μM	Michaelis constant for E2f inactivation by cyclinA/Cdk2 complex
K_{1Me}	0.1 μM	Michaelis constant for cyclinE/Cdk2 activation by E2F
K_{2Me}	0.1 μM	Michaelis constant for cyclinE/Cdk2 inactivation by cyclinA/Cdk2
K_{1Ma}	0.5 μM	Michaelis constant for cyclinA/Cdk2 activation by E2F
K_{2Ma}	0.1 μM	Michaelis constant for cyclinA/Cdk2 inactivation by Cdc20
K_{1Mb}	0.1 μM	Michaelis constant for cyclinB/Cdk1 activation by cyclinA/Cdk2
K_{2Mb}	0.1 μM	Michaelis constant for cyclinB/Cdk1 inactivation by Cdc20
K_{1cdc20}	0.1 μM	Michaelis constant for Cdc20 activation through phosphorylation by cyclinB/Cdk1
K_{2cdc20}	0.1 μM	Michaelis constant for Cdc20 inactivation through dephosphorylation
K_{ie}	0.2 μM	Michaelis constant for cyclinB/Cdk1 inactivation by cyclinE/Cdk2, representing the indirect effect of the DNA replication checkpoint
V_{1cdc20}	1.8 μM/h	Rate constant for activation of Cdc20 through phosphorylation by cyclinB/Cdk1
V_{2cdc20}	0.6 μM/h	Rate for inactivation of Cdc20 through dephosphorylation
V_{1E2F}	0.4 μM/h	Rate constant for activation of E2F by cyclinD/Cdk4-6 and cyclinE/Cdk2 complex
V_{2E2F}	0.7 μM/h	Rate constant for inactivation of E2F by cyclinA/Cdk2 complex
V_{1Me}	0.9 μM/h	Rate for activation of cyclinE/Cdk2 by E2F
V_{2Me}	1.6 μM/h	Rate constant for inactivation for cyclinE/Cdk2 by cyclinA/Cdk2
V_{1Ma}	0.6 μM/h	Rate constant for activation of cyclinA/Cdk2 by E2F
V_{2Ma}	0.6 μM/h	Rate constant for inactivation of cyclinA/Cdk2 by Cdc20
V_{1Mb}	1 μM/h	Rate for activation of cyclinB/Cdk1 by cyclinA/Cdk2
V_{2Mb}	0.8 μM/h	Rate constant for inactivation of cyclinB/Cdk1 by Cdc20

b	0.8 μM/h	basal term for activation of cyclin B/Cdk2 (without PF loop = 0)
Positive feedback loops on the G1/S and G2/M transitions		
b1	1	term representing the self-activation of cyclinE/Cdk2 through the mutual activation between cyclinE/Cdk2 and its phosphatase Cdc25(PF loop in the G1/S transition)
b2	1	term representing the self-activation of cyclinB/Cdk1 through mutual activation between cyclin B/cdk1 and its phosphatase cdk25(PF loop in the G2/M transition)
K_ib	0.5	Inhibition constant for Wee1 inactivation by cyclin B/Cdk1 (mutual inhibition of Wee1 and Cdk1 creates a PF loop in the G2/M transition)

Table D.2 Cell fate starting conditions

Skp2	0.0758
p21	0.01795
pp21	0.0304
Cdh1	0.0868
Myc	0.358
CycD	0.246
CycDp21	0.0114
Rb	0.2096
pRb	0.562
E2F	0.0663
RbE2F	0.659
CycE	0.0823
CycEp21	0.006
CycA	0.0781
CycAp21	0.011
CycB	0.2
CycBp21	0.002
Cdc20	0.367
BAX	0.992
BAXmBCL2	0.0073
BH3	0.004
BH3BCL2	0.199

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA. Cancer J. Clin.*, 2020, doi: 10.3322/caac.21590.

[2] P. Goldstraw *et al.*, "The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM Classification for lung cancer," *J. Thorac. Oncol.*, 2016, doi: 10.1016/j.jtho.2015.09.009.

[3] Y. A. Fouad and C. Aanei, "Revisiting the hallmarks of cancer," *American Journal of Cancer Research*. 2017.

[4] The Global Cancer Observatory, "Cancer fact sheet," *World Heal. Organ.*, 2019, doi: L11\n10.1051/0004-6361/201016331.

[5] N. Duma, R. Santana-Davila, and J. R. Molina, "Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment," *Mayo Clinic Proceedings*. 2019, doi: 10.1016/j.mayocp.2019.01.013.

[6] US National Institutes of Health. National Cancer Institute., "SEER Cancer Stat Facts: Lung and Bronchus Cancer." Bethesda, MD, [Online]. Available: https://seer.cancer.gov/statfacts/html/lungb.html.

[7] US National Institutes of Health. National Cancer Institute., "Financial Burden of Cancer Care." 2020, [Online]. Available: https://progressreport.cancer.gov/after/economic_burden.

[8] Z. Chen, C. M. Fillmore, P. S. Hammerman, C. F. Kim, and K. K. Wong, "Non-small-cell lung cancers: A heterogeneous set of diseases," *Nature Reviews Cancer*. 2014, doi: 10.1038/nrc3775.

[9] American Cancer Society, "What Is Lung Cancer ?" 2019, [Online]. Available: https://www.cancer.org/cancer/lung-cancer/about/what-is.html#written_by.

[10] M. Riihimäki *et al.*, "Metastatic sites and survival in lung cancer," *Lung Cancer*, 2014, doi: 10.1016/j.lungcan.2014.07.020.

[11] H. H. Popper, "Progression and metastasis of lung cancer," *Cancer Metastasis Rev.*, 2016, doi: 10.1007/s10555-016-9618-0.

[12] R. S. Herbst, D. Morgensztern, and C. Boshoff, "The biology and management of non-small cell lung cancer," *Nature*. 2018, doi: 10.1038/nature25183.

[13] US National Institutes of Health. National Library of Medicine., "Genetics Home Reference - Genes." 2020, [Online]. Available: https://ghr.nlm.nih.gov/gene.

[14] C. Gridelli *et al.*, "Non-small-cell lung cancer," *Nature Reviews Disease Primers*. 2015, doi: 10.1038/nrdp.2015.9.

[15] A. J. Alberg and J. M. Samet, "Epidemiology of lung cancer," *Chest*. 2003, doi:

10.1378/chest.123.1_suppl.21S.

[16] N. K. Altorki *et al.*, “The lung microenvironment: an important regulator of tumour growth and metastasis,” *Nature Reviews Cancer*. 2019, doi: 10.1038/s41568-018-0081-9.

[17] N. E. Reticker-Flynn and S. N. Bhatia, “Aberrant glycosylation promotes lung cancer metastasis through adhesion to galectins in the metastatic niche,” *Cancer Discov.*, 2015, doi: 10.1158/2159-8290.CD-13-0760.

[18] G. Burgstaller, B. Oehrle, M. Gerckens, E. S. White, H. B. Schiller, and O. Eickelberg, “The instructive extracellular matrix of the lung: Basic composition and alterations in chronic lung disease,” *European Respiratory Journal*. 2017, doi: 10.1183/13993003.01805-2016.

[19] B. Hinz, “Mechanical aspects of lung fibrosis: A spotlight on the myofibroblast,” in *Proceedings of the American Thoracic Society*, 2012, doi: 10.1513/pats.201202-017AW.

[20] V. C. Shukla, N. Higuita-Castro, P. Nana-Sinkam, and S. N. Ghadiali, “Substrate stiffness modulates lung cancer cell migration but not epithelial to mesenchymal transition,” *J. Biomed. Mater. Res. - Part A*, 2016, doi: 10.1002/jbm.a.35655.

[21] A. E. Forest, Y. Shiozawa, K. J. Pienta, and R. S. Taichman, “The Hematopoietic Stem Cell Niche and Bone Metastasis.” Landes Bioscience, Austin, TX, 2013.

[22] M. Akhmanova, E. Osidak, S. Domogatsky, S. Rodin, and A. Domogatskaya, “Physical, Spatial, and Molecular Aspects of Extracellular Matrix of in Vivo Niches and Artificial Scaffolds Relevant to Stem Cells Research,” *Stem Cells Int.*, 2015, doi: 10.1155/2015/167025.

[23] J. Weickenmeier, R. de Rooij, S. Budday, P. Steinmann, T. C. Ovaert, and E. Kuhl, “Brain stiffness increases with myelin content,” *Acta Biomater.*, 2016, doi: 10.1016/j.actbio.2016.07.040.

[24] D. Chauvet *et al.*, “In Vivo Measurement of Brain Tumor Elasticity Using Intraoperative Shear Wave Elastography,” *Ultraschall der Medizin*, 2016, doi: 10.1055/s-0034-1399152.

[25] R. G. Wells, “The role of matrix stiffness in regulating cell behavior,” *Hepatology*. 2008, doi: 10.1002/hep.22193.

[26] R. Navab *et al.*, “Integrin α11β1 regulates cancer stromal stiffness and promotes tumorigenicity and metastasis in non-small cell lung cancer,” *Oncogene*, 2016, doi: 10.1038/onc.2015.254.

[27] H. Uramoto and F. Tanaka, “Recurrence after surgery in patients with NSCLC,” *Translational Lung Cancer Research*. 2014, doi: 10.3978/j.issn.2218-6751.2013.12.05.

[28] American Society of Clinical Oncology, “Stages of Cancer | Cancer.net.” 2018, [Online]. Available: https://www.cancer.net/navigating-cancer-care/diagnosing-cancer/stages-cancer.

[29] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, “Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship,” in *Mayo Clinic*

Proceedings, 2008, doi: 10.4065/83.5.584.

[30] L. C. Cheung, H. A. Katki, A. K. Chaturvedi, A. Jemal, and C. D. Berg, "Preventing lung cancer mortality by computed tomography screening: The effect of risk-based versus U.S. Preventive services task force eligibility criteria, 2005–2015," *Annals of Internal Medicine*. 2018, doi: 10.7326/M17-2067.

[31] D. J. Heineman, J. M. Daniels, and W. H. Schreurs, "Clinical staging of NSCLC: Current evidence and implications for adjuvant chemotherapy," *Therapeutic Advances in Medical Oncology*. 2017, doi: 10.1177/1758834017722746.

[32] US National Institutes of Health. National Library of Medicine., "ClinicalTrials.gov." 2020, [Online]. Available: https://clinicaltrials.gov/ct2/home.

[33] E. A. Ponomarenko *et al.*, "The Size of the Human Proteome: The Width and Depth," *International Journal of Analytical Chemistry*. 2016, doi: 10.1155/2016/7436849.

[34] S. H. Strogatz, "Exploring complex networks," *Nature*. 2001, doi: 10.1038/35065725.

[35] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology*. 2010.

[36] B. N. Kholodenko, "Cell-signalling dynamics in time and space," *Nature Reviews Molecular Cell Biology*. 2006, doi: 10.1038/nrm1838.

[37] B. B. Aldridge, J. M. Burke, D. A. Lauffenburger, and P. K. Sorger, "Physicochemical modelling of cell signalling pathways," *Nature Cell Biology*. 2006, doi: 10.1038/ncb1497.

[38] N. J. Eungdamrong and R. Iyengar, "Modeling cell signaling networks," *Biology of the Cell*. 2004, doi: 10.1016/j.biolcel.2004.03.004.

[39] L. Michaelis, M. L. Menten, R. S. Goody, and K. A. Johnson, "Die Kinetik der Invertinwirkung/ The kinetics of invertase action," *Biochemistry*, 1913, doi: 10.1021/bi201284u.

[40] A. Goldbeter and D. E. Koshland, "An amplified sensitivity arising from covalent modification in biological systems," *Proc. Natl. Acad. Sci. U. S. A.*, 1981, doi: 10.1073/pnas.78.11.6840.

[41] S. Hooshangi, S. Thiberge, and R. Weiss, "Ultrasensitivity and noise propagation in a synthetic transcriptional cascade," *Proc. Natl. Acad. Sci. U. S. A.*, 2005, doi: 10.1073/pnas.0408507102.

[42] N. B. Trunnell, A. C. Poon, S. Y. Kim, and J. E. Ferrell, "Ultrasensitivity in the Regulation of Cdc25C by Cdk1," *Mol. Cell*, 2011, doi: 10.1016/j.molcel.2011.01.012.

[43] T. Y. C. Tsai, J. A. Theriot, and J. E. Ferrell, "Changes in Oscillatory Dynamics in the Cell Cycle of Early Xenopus laevis Embryos," *PLoS Biol.*, 2014, doi: 10.1371/journal.pbio.1001788.

[44] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," in *Journal of Physical Chemistry*, 1977, doi: 10.1021/j100540a008.

[45] B. D. Aguda and Y. Tang, "The kinetic origins of the restriction point in the mammalian

cell cycle," *Cell Prolif.*, 1999, doi: 10.1046/j.1365-2184.1999.3250321.x.

[46] B. Novák and J. J. Tyson, "A model for restriction point control of the mammalian cell cycle," *J. Theor. Biol.*, 2004, doi: 10.1016/j.jtbi.2004.04.039.

[47] G. Yao, T. J. Lee, S. Mori, J. R. Nevins, and L. You, "A bistable Rb-E2F switch underlies the restriction point," *Nat. Cell Biol.*, 2008, doi: 10.1038/ncb1711.

[48] G. Yao, T. J. Lee, S. Mori, J. R. Nevins, and L. You, "A bistable Rb-E2F switch underlies the restriction point.," *Nat. Cell Biol.*, vol. 10, no. 4, pp. 476–482, 2008, doi: 10.1038/ncb1711.

[49] M. V. Frolov and N. J. Dyson, "Molecular mechanisms of E2F-dependent activation and pRB-mediated repression," *Journal of Cell Science*. 2004, doi: 10.1242/jcs.01227.

[50] C. Attwooll, E. L. Denchi, and K. Helin, "The E2F family: Specific functions and overlapping interests," *EMBO Journal*. 2004, doi: 10.1038/sj.emboj.7600481.

[51] L. Wu *et al.*, "The E2F1-3 transcription factors are essential for cellular proliferation," *Nature*, 2001, doi: 10.1038/35106593.

[52] J. Y. Leung, G. L. Ehmann, P. H. Giangrande, and J. R. Nevins, "A role for Myc in facilitating transcription activation by E2F1," *Oncogene*, 2008, doi: 10.1038/onc.2008.55.

[53] D. O. Morgan, "Principles of CDK regulation," *Nature*, 1995, doi: 10.1038/374131a0.

[54] G. Cooper, *The Cell: A Molecular Approach. 2nd edition.* Sunderland (MA), 2000.

[55] C. Gérard and A. Goldbeter, "Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 51, pp. 21643–21648, 2009, doi: 10.1073/pnas.0903827106.

[56] C. Gerard and A. Goldbeter, "A skeleton model for the network of cyclin-dependent kinases driving the mammalian cell cycle," *Interface Focus*, vol. 1, no. 1, pp. 24–35, 2011, doi: 10.1098/rsfs.2010.0008.

[57] A. Goldbeter and J. Keizer, "Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour," *Phys. Today*, 1998, doi: 10.1063/1.882194.

[58] J. C. Leloup and A. Goldbeter, "Chaos and birhythmicity in a model for circadian oscillations of the PER and TIM proteins in Drosophila," *J. Theor. Biol.*, 1999, doi: 10.1006/jtbi.1999.0924.

[59] B. Novak and J. J. Tyson, "Modeling the control of DNA replication in fission yeast," *Proc. Natl. Acad. Sci. U. S. A.*, 1997, doi: 10.1073/pnas.94.17.9147.

[60] A. Csikász-Nagy, D. Battogtokh, K. C. Chen, B. Novák, and J. J. Tyson, "Analysis of a generic model of eukaryotic cell-cycle regulation," *Biophys. J.*, 2006, doi: 10.1529/biophysj.106.081240.

[61] H. P. Fischer, "Mathematical modeling of complex biological systems: From parts lists to understanding systems behavior," *Alcohol Research and Health*. 2008.

[62] T. Friedlander, A. E. Mayo, T. Tlusty, and U. Alon, “Evolution of Bow-Tie Architectures in Biology,” *PLoS Comput. Biol.*, 2015, doi: 10.1371/journal.pcbi.1004055.

[63] M. A. Lemmon and J. Schlessinger, “Cell signaling by receptor tyrosine kinases,” *Cell*. 2010, doi: 10.1016/j.cell.2010.06.011.

[64] S. Morandell, T. Stasyk, S. Skvortsov, S. Ascher, and L. A. Huber, “Quantitative proteomics and phosphoproteomics reveal novel insights into complexity and dynamics of the EGFR signaling network,” *Proteomics*. 2008, doi: 10.1002/pmic.200800204.

[65] G. Levkowitz *et al.*, “Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1,” *Mol. Cell*, 1999, doi: 10.1016/S1097-2765(00)80231-2.

[66] T. Mitsudomi and Y. Yatabe, “Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer,” *FEBS Journal*. 2010, doi: 10.1111/j.1742-4658.2009.07448.x.

[67] J. A. Engelman *et al.*, “MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling,” *Science (80-.).*, 2007, doi: 10.1126/science.1141478.

[68] F. McCormick, “How receptors turn Ras on,” *Nature*. 1993, doi: 10.1038/363015a0.

[69] A. R. Grassian, Z. T. Schafer, and J. S. Brugge, “ErbB2 stabilizes Epidermal Growth Factor Receptor (EGFR) expression via Erk and sprouty2 in extracellular matrix-detached cells,” *J. Biol. Chem.*, 2011, doi: 10.1074/jbc.M110.169821.

[70] V. J. Fincham, “Active ERK/MAP kinase is targeted to newly forming cell-matrix adhesions by integrin engagement and v-Src,” *EMBO J.*, 2000, doi: 10.1093/emboj/19.12.2911.

[71] J. P. Alao, “The regulation of cyclin D1 degradation: Roles in cancer development and the potential for therapeutic invention,” *Molecular Cancer*. 2007, doi: 10.1186/1476-4598-6-24.

[72] N. Hay and N. Sonenberg, “Upstream and downstream of mTOR,” *Genes and Development*. 2004, doi: 10.1101/gad.1212704.

[73] J. Zhu, J. Blenis, and J. Yuan, “Activation of PI3K/Akt and MAPK pathways regulates Myc-mediated transcription by phosphorylating and promoting the degradation of Mad1,” *Proc. Natl. Acad. Sci.*, 2008, doi: 10.1073/pnas.0802785105.

[74] T. Zhang, P. Brazhnik, and J. J. Tyson, “Computational analysis of dynamical responses to the intrinsic pathway of programmed cell death,” *Biophys. J.*, vol. 97, no. 2, pp. 415–434, 2009, doi: 10.1016/j.bpj.2009.04.053.

[75] B. A. Chan and B. G. M. Hughes, “Targeted therapy for non-small cell lung cancer: Current standards and the promise of the future,” *Translational Lung Cancer Research*. 2015, doi: 10.3978/j.issn.2218-6751.2014.05.01.

[76] S. M. Lim, N. L. Syn, B. C. Cho, and R. A. Soo, “Acquired resistance to EGFR targeted therapy in non-small cell lung cancer: Mechanisms and therapeutic strategies,” *Cancer*

Treatment Reviews. 2018, doi: 10.1016/j.ctrv.2018.02.006.

[77] S. Zimmermann and S. Peters, "Present standards and future perspectives in the treatment of metastatic non-small cell lung cancer," *Cancer Metastasis Rev.*, 2015, doi: 10.1007/s10555-015-9560-6.

[78] D. R. Tyson, S. P. Garbett, P. L. Frick, and V. Quaranta, "Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data," *Nat Methods*, vol. 9, no. 9, pp. 923–928, 2012, doi: 10.1038/nmeth.2138.

[79] S. Vyse and P. H. Huang, "Targeting EGFR exon 20 insertion mutations in non-small cell lung cancer," *Signal Transduction and Targeted Therapy*. 2019, doi: 10.1038/s41392-019-0038-9.

[80] R. Vijayalakshmi and A. Krishnamurthy, "Targetable 'Driver' Mutations in Non Small Cell Lung Cancer," *Indian Journal of Surgical Oncology*. 2011, doi: 10.1007/s13193-011-0108-0.

[81] L. Cheng *et al.*, "The landscape of EGFR pathways and personalized management of non-small-cell lung cancer," *Future Oncology*. 2011, doi: 10.2217/fon.11.25.

[82] P. L. Frick, B. B. Paudel, D. R. Tyson, and V. Quaranta, "Quantifying heterogeneity and dynamics of clonal fitness in response to perturbation," *J. Cell. Physiol.*, 2015, doi: 10.1002/jcp.24888.

[83] W. Pao *et al.*, "Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain," *PLoS Med.*, 2005, doi: 10.1371/journal.pmed.0020073.

[84] Z. Zhang *et al.*, "Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer," *Nat. Genet.*, 2012, doi: 10.1038/ng.2330.

[85] P. L. Frick, "Quantifying cancer heterogeneity to predict and improve targeted therapy outcomes," Vanderbilt University, 2014.

[86] D. Anton, H. Burckel, E. Josset, and G. Noel, "Three-dimensional cell culture: A breakthrough in vivo," *International Journal of Molecular Sciences*. 2015, doi: 10.3390/ijms16035517.

[87] E. C. Costa, A. F. Moreira, D. de Melo-Diogo, V. M. Gaspar, M. P. Carvalho, and I. J. Correia, "3D tumor spheroids: an overview on the tools and techniques used for their analysis," *Biotechnology Advances*. 2016, doi: 10.1016/j.biotechadv.2016.11.002.

[88] J. Zhao, "Harnessing Non-invasive Imaging and Modeling towards Early Detection of Tumor Resistance to Targeted Therapy," Vanderbilt University, 2017.

[89] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, 1974, doi: 10.1109/TAC.1974.1100705.

[90] V. Gkretsi and T. Stylianopoulos, "Cell adhesion and matrix stiffness: Coordinating cancer cell invasion and metastasis," *Frontiers in Oncology*. 2018, doi: 10.3389/fonc.2018.00145.

[91] H. Gerardo *et al.*, “Soft culture substrates favor stem-like cellular phenotype and facilitate reprogramming of human mesenchymal stem/stromal cells (hMSCs) through mechanotransduction,” *Sci. Rep.*, 2019, doi: 10.1038/s41598-019-45352-3.

[92] A. Jahangiri *et al.*, “Cross-activating c-Met/β1 integrin complex drives metastasis and invasive resistance in cancer,” *Proc. Natl. Acad. Sci. U. S. A.*, 2017, doi: 10.1073/pnas.1701821114.

[93] A. Golubev, “Genes at work in random bouts: Stochastically discontinuous gene activity makes cell cycle duration and cell fate decisions variable, thus providing for stem cells plasticity,” *BioEssays*, 2012, doi: 10.1002/bies.201100119.

[94] A. Sveiczer, A. Csikasz-Nagy, B. Gyorffy, J. J. Tyson, and B. Novak, “Modeling the fission yeast cell cycle: Quantized cycle times in wee1- cdc25Δ mutant cells,” *Proc. Natl. Acad. Sci. U. S. A.*, 2000, doi: 10.1073/pnas.97.14.7865.

[95] R. R. Klevecz, “Quantized generation time in mammalian cells as an expression of the cellular clock,” *Proc. Natl. Acad. Sci. U. S. A.*, 1976, doi: 10.1073/pnas.73.11.4012.

[96] L. Ding *et al.*, “Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing,” *Nature*, 2012, doi: 10.1038/nature10738.

[97] S. Nanjo *et al.*, “Ability of the Met kinase inhibitor crizotinib and new generation EGFR inhibitors to overcome resistance to EGFR inhibitors,” *PLoS One*, 2013, doi: 10.1371/journal.pone.0084700.

[98] C. Grommes *et al.*, “Pulsatile high-dose weekly erlotinib for CNS metastases from EGFR mutant non-small cell lung cancer,” *Neuro. Oncol.*, 2011, doi: 10.1093/neuonc/nor121.

[99] P. W. Oakes, S. P. Winter, J. Stricker, M. L. Gardel, and Y. Aratyn-Schaus, “Preparation of Complaint Matrices for Quantifying Cellular Contraction,” *J. Vis. Exp.*, 2010, doi: 10.3791/2173.

[100] M. Bouhaddou *et al.*, “A mechanistic pan-cancer pathway model informed by multi-omics data interprets stochastic cell fate responses to drugs and mitogens,” *PLoS Comput. Biol.*, 2018, doi: 10.1371/journal.pcbi.1005985.

[101] T. S. Weber, I. Jaehnert, C. Schichor, M. Or-Guil, and J. Carneiro, “Quantifying the Length and Variance of the Eukaryotic Cell Cycle Phases by a Stochastic Model and Dual Nucleoside Pulse Labelling,” *PLoS Comput. Biol.*, 2014, doi: 10.1371/journal.pcbi.1003616.

[102] C. Gérard, D. Gonze, and A. Goldbeter, “Effect of positive feedback loops on the robustness of oscillations in the network of cyclin-dependent kinases driving the mammalian cell cycle,” *FEBS J.*, vol. 279, no. 18, pp. 3411–3431, 2012, doi: 10.1111/j.1742-4658.2012.08585.x.

[103] C. Giacinti and A. Giordano, “RB and cell cycle progression,” *Oncogene*. 2006, doi: 10.1038/sj.onc.1209615.

[104] P. Wang, S. Hayden, and Y. Masui, “Transition of the blastomere cell cycle from cell size-independent to size-dependent control at the midblastula stage in Xenopus laevis,” *J.*

Exp. Zool., 2000, doi: 10.1002/1097-010X(20000701)287:2<128::AID-JEZ3>3.0.CO;2-G.

[105] A. Sveiczer, B. Novak, and J. M. Mitchison, "The size control of fission yeast revisited," *J. Cell Sci.*, 1996.

[106] U. Kossatz, N. Dietrich, L. Zender, J. Buer, M. P. Manns, and N. P. Malek, "Skp2-dependent degradation of p27kip1 is essential for cell cycle progression," *Genes Dev.*, 2004, doi: 10.1101/gad.321004.

[107] S. L. Spencer, S. D. Cappell, F. C. Tsai, K. W. Overton, C. L. Wang, and T. Meyer, "XThe proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit," *Cell*, 2013, doi: 10.1016/j.cell.2013.08.062.

[108] C. F. Lopez, J. L. Muhlich, J. A. Bachman, and P. K. Sorger, "Programming biological models in Python using PySB," *Mol. Syst. Biol.*, 2013, doi: 10.1038/msb.2013.1.

[109] C. Gérard and A. Goldbeter, "From simple to complex patterns of oscillatory behavior in a model for the mammalian cell cycle containing multiple oscillatory circuits," *Chaos*, 2010, doi: 10.1063/1.3527998.

[110] B. Pucci, M. Kasten, and A. Giordano, "Cell cycle and apoptosis," *Neoplasia*. 2000, doi: 10.1038/sj.neo.7900101.

[111] F. Q. B. Alenzi, "Links between apoptosis, proliferation and the cell cycle," *British Journal of Biomedical Science*. 2004, doi: 10.1080/09674845.2004.11732652.

[112] N. Obatake, A. Shiu, X. Tang, and A. Torres, "Oscillations and bistability in a model of ERK regulation," *J. Math. Biol.*, 2019, doi: 10.1007/s00285-019-01402-y.

[113] D. M. Norris, P. Yang, J. R. Krycer, D. J. Fazakerley, D. E. James, and J. G. Burchfield, "An improved Akt reporter reveals intra- and inter-cellular heterogeneity and oscillations in signal transduction," *J. Cell Sci.*, 2017, doi: 10.1242/jcs.205369.

[114] K. Pardee *et al.*, "Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components," *Cell*, 2016, doi: 10.1016/j.cell.2016.04.059.

[115] C. C. Kloss, M. Condomines, M. Cartellieri, M. Bachmann, and M. Sadelain, "Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells," *Nat. Biotechnol.*, 2013, doi: 10.1038/nbt.2459.

[116] L. Schukur, B. Geering, G. Charpin-El Hamri, and M. Fussenegger, "Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis," *Sci. Transl. Med.*, 2015, doi: 10.1126/scitranslmed.aac4964.