

**The Dissertation Committee for Sandeep Sanga Certifies that this is the approved
version of the following dissertation:**

**A Computational Systems Biology Approach to Predictive Oncology: A
Computer Modeling and Bioinformatics Study Predicting Tumor
Response to Therapy and Cancer Phenotypes**

Committee:

Vittorio Cristini, Supervisor

Mary E. Edgerton

Paul T. Macklin

Yin Liu

Ennio Tasciotti

**A Computational Systems Biology Approach to Predictive Oncology: A
Computer Modeling and Bioinformatics Study Predicting Tumor
Response to Therapy and Cancer Phenotypes**

by

Sandeep Sanga, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2009

Dedication

I dedicate this dissertation to my mother. She is the driving force in my life. For better or for worse, I am who I am today because of her. I love you mom.

Acknowledgements

I would like to mention the individuals who have been amazing scientific mentors and have supported me over the years towards the culmination of this research dissertation:

- | Vittorio Cristini, Ph.D.
- | Mary E. Edgerton, M.D., Ph.D.
- | Paul T. Macklin, Ph.D.
- | John P. Sinek, Ph.D.
- | Mauro Ferrari, Ph.D.
- | Hermann B. Frieboes, Ph.D.
- | John Lowengrub, Ph.D.
- | Xiaoming Zheng, Ph.D.
- | Steven M. Wise, Ph.D.
- | Sangeeta N. Bhatia, M.D., Ph.D.
- | Vicki I. Chin, Ph.D.

**A Computational Systems Biology Approach to Predictive Oncology: A
Computer Modeling and Bioinformatics Study Predicting Tumor
Response to Therapy and Cancer Phenotypes**

Publication No. _____

Sandeep Sanga, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Vittorio Cristini

Technological advances in the recent decades have enabled cancer researchers to probe the disease at multiple resolutions. This wealth of experimental data combined with computational systems biology methods is now leading to predictive models of cancer progression and response to therapy. We begin by presenting our research group's multi-scale *in silico* framework for modeling cancer, whose core is a tissue-scale computational model capable of tracking the progression of tumors from a diffusion-limited avascular phase through angiogenesis, and into invasive lesions with realistic, complex morphologies. We adapt this core model to consider the delivery of systemically-administered anticancer agents and their effect on lesions once they reach their intended nuclear target. We calibrate the model parameters using *in vitro* data from the literature, and demonstrate through simulation that transport limitations affecting drug and oxygen distributions play a significant role in hampering the efficacy of chemotherapy; a result that has since been validated by *in vitro* experimentation. While this study demonstrates the capability of our

adapted core model to predict distributions (*e.g.*, cell density, pressure, oxygen, nutrient, drug) within lesions and consequent tumor morphology, nevertheless, the underlying factors driving tumor-scale behavior occur at finer scales. What is needed in our multi-scale approach is to parallel reality, where molecular signaling models predict cellular behavior, and ultimately drive what is seen at the tumor level. Models of signaling pathways linked to cell models are already beginning to surface in the literature. We next transition our research to the molecular level, where we employ data mining and bioinformatics methods to infer signaling relationships underlying a subset of breast cancer that might benefit from targeted therapy of Androgen Receptor and associated pathways. Defining the architecture of signaling pathways is a critical first step towards development of pathways models underlying tumor models, while also providing valuable insight for drug discovery. Finally, we develop an agent-based, cell-scale model focused on predicting motility in response to chemical signals in the microenvironment, generally accepted to be a necessary feature of cancer invasion and metastasis. This research demonstrates the use of signaling models to predict emergent cell behavior, such as motility.

The research studies presented in this dissertation are critical steps towards developing a predictive, *in silico* computational model for cancer progression and response to therapy. Our Laboratory for Computational & Predictive Oncology, in collaboration with research groups throughout in the United States and Europe are following a computational systems biology paradigm where model development is fueled by biological knowledge, and model predictions are refining experimental focus. The ultimate objective is a virtual cancer simulator capable of accurately simulating cancer progression and response to therapy on a patient-specific basis.

Table of Contents

Chapter 1: A Brief Introduction to Cancer Biology	1
The Central Dogma of Molecular Biology	2
The Cell Cycle: Normal vs. Cancer	4
What Are the Key Features of Cancers?	6
Carcinogenesis Occurs at the DNA Level	8
Cancer Metastasis	9
How Are Cancers Classified?	10
How Are Cancers Currently Treated?	11
Surgery	12
Radiation Therapy	12
Chemotherapy	13
Cytotoxic Chemotherapy	13
Targeted Chemotherapy	14
Anti-Angiogenic Chemotherapy	14
Hormonal Therapy	15
Chapter 2: Introduction	16
Preface	16
The Multifaceted Nature of Cancer	17
The Push for Personalized Cancer Medicine: Predictive Oncology	18
The Role of Computational Systems Biology in Cancer Research and Therapy Design	21
Multi-Scale Cancer Simulation Paradigm	23
Research Purpose	24

Chapter 3: Techniques and Methods used in Predictive Oncology	27
Mechanistic Cancer Modeling Approaches	28
Modular Development of a Multi-scale Cancer Simulation	30
Tumor Growth	31
Tumor-Induced Angiogenesis.....	32
Pharmacology and Drug Efficacy	36
Core Tumor Model Used in Chapter Five	40
Empirical Cancer Modeling Approaches: Microarray Gene Expression Analysis	44
Unsupervised Analysis – Exploring Gene Expression Levels to Identify Co-Expressing Genes Within and Across Samples	45
Supervised Analysis.....	46
Identifying Differentially Expressed Genes (Comparative Analysis): Discovery of Predictive Gene Signatures and Biomarkers.....	47
Classifier Prediction Methods.....	48
Modeling Intracellular Pathways Driving Tumor Phenotypes.....	49
Reconstruction of Signaling Maps.....	51
Chapter 4: Predictive Oncology: A Review of Multi-disciplinary, Multi-scale <i>In Silico</i> Modeling Linking Phenotype, Morphology and Growth.....	53
Preface.....	53
Abstract.....	53
The Role of Predictive Scientific Computation as “ <i>In Silico</i> ” Cancer Modeling	54
Cancer Progression and Invasion: Current Understanding.....	55
Links Between Cellular- and Tumor-scale	55
A Novel, <i>In Silico</i> Approach to Cancer Modeling.....	56
Incorporation of Patient Data: Predictive Modeling.....	58
Multi-Scale Modeling and Simulation of Tumor Morphology and Invasion.....	61
Model Development Goals and Choices.....	61
Significance of <i>In Silico</i> Modeling: A Novel Hypothesis-Generation Tool	63
Determination of Functional Relationships and Parameter Values.....	63

Computational Modeling: A Framework for Linking Phenotype, Morphology, and Cancer Invasion.....	65
Effects of Cell-Cell and Cell-Matrix Interactions.....	69
Effects of Cell Motility.....	71
Effects of Micro Vessel Density and Acidosis.....	72
Effects of Cell Substrate Concentration.....	73
Continuum-Based Parameter-Sensitivity Studies of FCCMU.....	73
Morphologic Instability as a Mechanism of Tumor Invasion	73
Clinical Relevance.....	77
Effects of Phenotype on Morphology and Growth.....	80
Conclusions and Future Work.....	84
Acknowledgements.....	86
Chapter 5: Predicting Drug Pharmacokinetics and Effect in Vascularized Tumors Using Computer Simulation.....	87
Preface.....	87
Abstract.....	88
Introduction.....	90
Materials and Methods.....	94
Model Description.....	94
Pharmacokinetics Model Parameters.....	99
Pharmacodynamics Model Parameters.....	104
<i>In Silico</i> Experiments.....	106
Results.....	108
Discussion.....	114
Acknowledgments.....	121
Chapter 6: Gene Expression Meta-Analysis Supports Existence of Molecular Apocrine Breast Cancer with a Role for Androgen Receptor and Implies Interactions with ErbB Family.....	122
Preface.....	122
Abstract.....	124
Introduction.....	125

Data.....	127
Results.....	128
Data Normalization.....	128
Comparison for Molecular Equivalence.....	132
Functional Analysis of the “Molecular Apocrine” Phenotype Using LeFEminer.....	134
Network Inference Analysis of “Molecular Apocrine” Phenotype Using Backward Chaining Rule Induction and MetaCore.....	137
Persistence of Molecularly-defined Phenotype in Larger Dataset....	139
Network Inference Analysis of “Molecular Apocrine” Phenotype Using Gene Shaving & Robust Bayesian Network Analysis.....	141
Analysis of the Interacting Gene Clusters.....	143
Discussion & Conclusion.....	144
Observations of Normalization Strategies to Remove Institutional Bias in Meta-Analysis of Gene Expression Array Data.....	144
Molecular Equivalence of the “ER ⁻ Subclass A” with “Molecular Apocrine” Breast Cancer.....	144
Role of AR Signaling in Molecular Apocrine Tumors.....	145
Pathways that Interact with AR in Molecular Apocrine Breast Cancer	146
Clinical and Therapeutic Implications for Molecular Apocrine Breast Cancer.....	148
Learning the Systems Biology of Cancer Using Network Inference Methods to Analyze Gene Expression Data.....	149
Materials & Methods.....	150
Data Collection.....	150
Microarray Normalization: Removing Systematic and Institutional Bias	150
Significance Analysis of Microarrays: Modified T-Test.....	151
Hierarchical Clustering and Principal Components Analysis.....	151
Two-Dimensional Kolmogorov-Smirnov Test.....	152
Statistical Significance of Overlap Between Gene Signatures.....	152
Backward Chaining Rule Induction.....	152
Gene Shaving.....	153

Robust Bayesian Network Analysis.....	153
List of Key Abbreviations.....	154
Acknowledgements.....	155
Chapter 7: From Receptor Dynamics to Directed Cell Motion: A Predictive Agent- Based Model for Cell Motility in Complex Microenvironments.....	156
Preface.....	156
Abstract.....	157
Introduction.....	158
Mathematical-Computational Modeling Can Link the Multiple Scales	158
Signal Transduction Research Reveals Signaling Relationships Governing Cellular Behavior.....	159
Towards the Development of Predictive, Multi-Scale, Computer Representations of Cancer by Incorporating Pathways Knowledge into Each Cell.....	160
Cell Motility Modeling Using an Agent-based Approach with Pathways Knowledge.....	163
The Model Building Cycle.....	164
Model Development.....	165
Key Observables of Directed Cell Motility: A Priori Knowledge for Model Development.....	165
A Cell Modeled as a “Shape Machine”.....	167
Cell Model Details.....	168
Model of Cell Shape as a Function of a Protruding and Retracting Membrane.....	168
Model of the Receptor-Ligand Dynamics Leading to Intracellular Signaling.....	175
Other Modeling Considerations: Agent-Based Framework, Cell-Cell Sensing, and Proliferation.....	178
Simulation of In Vitro Microenvironments.....	181
Prototype Model Performance: Comparison to In Vitro Experiments from Literature.....	182
Calibration of Receptor Occupancy Theory-based Signaling Model	184
Validation: Breast Carcinoma Cells Responding to EGF Gradients	190

Using a Protein Signaling Model to Justify the Functional Link Between Receptor-Ligand Binding and Signal Generation Leading to Motility	192
Discussion and Conclusion	194
Materials & Methods.....	196
Acknowledgements	197
Appendix	198
Chapter 6 Supplementary Figures and Tables.....	198
Figures	198
Tables	201
Close Network Interactions Identified by MetaCore for the Select Interacting Gene Clusters	216
References	221

Chapter 1: A Brief Introduction to Cancer Biology

Worldwide, cancer claims the lives of two out of every three diagnosed individuals approximating to seven million deaths annually, and millions more living abridged lifestyles as a consequence of treatment side-effects (Kamangar, Dores, & Anderson, 2006). Currently, there are approximately 25 million persons alive with cancer, with lung, breast, and colon cancers the most commonly diagnosed (Parkin, Bray, Ferlay, & Pisani, 2005).

Normal and healthy animal physiology can be thought of as a utopian society of cells functioning in harmony with one another, where self-sacrifice for the greater good is the covenant; individual cells play a role in this society by proliferating when needed, organize in multi-cellular patterns, tissues, and organs to serve a biological function, and die when their time has come to maintain an appropriate population balance. Cancer is a complex disease whose origins lay in individual cells experiencing a series of genetic mutations which allow them to reproduce and mobilize in defiance of normal constraints. The selective advantage attained by cancer cells gives them a selfish nature that disrupts the otherwise harmonious society. Cancer cells operate and prosper at the expense of their neighboring cells, growing into large cellular masses called tumors that burden the normal functioning society to the verge of collapse, ultimately causing organ failures and death.

The purpose of this chapter is meant to briefly introduce the reader to the nature of this complex, multifaceted disease and set the stage for the subsequent chapters in this dissertation describing computational systems biology-based investigations of cancer, how multi-scale modeling can advance our understanding of the disease, and therefore improve our ability to treat it. Specifically, this chapter surveys the basic principles of

molecular biology, how cancer initiates, its variability between cell types, its progression towards metastasis, and current treatment options.

THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology states that a cell's DNA (genome) encodes all of the RNA (transcriptome), which in turn encodes the protein molecules (proteome) required to construct cells and coordinate their behaviors (Figure 1.1). While DNA serves as the metaphorical collection of musical notes, the RNA and proteins can be thought of as the combination of sounds that produce musical masterpieces. Just as each musical piece differs based on how musical notes are constructed together, each cell type in an organism differs in its appearance, construction, and behavior despite access to the same library. Fundamentally, cell types in multi-cellular organisms differ from one another because they accumulate different sets of RNA and protein molecules, while conserving the nuclear DNA sequences that encode them. Many cellular functions and behaviors are common across cell types, and as such, rely on many of the same proteins. While human cells are thought to utilize approximately 21,000 proteins (Mishra, *et al.*, 2006), differences in only several hundred protein interactions are believed to enable the vast differences in morphology and behavior between cell types (Alberts, *et al.*, 1994).

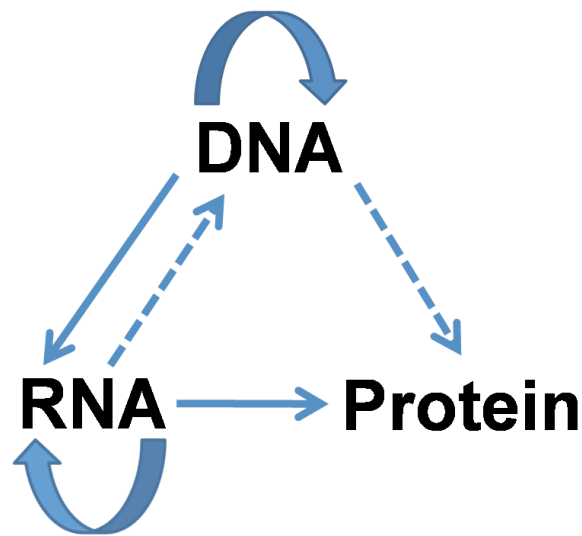


Figure 1.1: The Dogma of Molecular Biology.

The schematic represents the transfer of sequence information amongst DNA, RNA, and Proteins. In the most general cases, transfer of sequence information occurs when DNA replicates itself preceding mitosis, DNA segments undergo transcription to form RNA, and RNA is translated into Proteins. In some special cases occurring in some viruses, RNA can be used as a template for synthesizing DNA in a process called reverse transcription and to copy itself in a process called RNA replication. DNA has been shown to be able to directly translate into Proteins without the usual intermediary (RNA) under in vitro experimental conditions.

Gene expression refers to the dynamics of cellular control over their phenotypes, or observable behaviors, by regulating the pathway between DNA to RNA to Proteins. Cells typically alter the expression of their genes in response to external signals, which in turn drives their phenotype. This process can be regulated at many steps along the pathway. While the specifics of this regulation is beyond the scope of this brief introductory chapter, it is important to note that the transmission of signals between the genome, transcriptome, proteome, and external cues (*e.g.*, levels of oxygen, levels of nutrient, presence of growth promoters, presence of motility promoters, presence of a drug, etc.) form the basis of cell signaling networks that drive cell phenotype, their social behavior in multi-cellular patterns, and their response to therapy. In the case of cancer, alterations in the signaling dynamics between the genome, transcriptome, and proteome

can have a profound, altering effect on how cells respond to their microenvironment, which drives these aberrant cell phenotypes to grow into drug-resisting, invasive, malignant tumor lesions.

THE CELL CYCLE: NORMAL VS. CANCER

Each cell undergoes a series of events between replication events. The duration of this so called cell cycle varies between cell types (usually on the order of 12-24 hours for proliferating mammalian cells), but can be generalized into several distinct phases (Figure 1.2):

- | G1 Phase: The interval following a successful cell division and prior to entering the S Phase.
- | S Phase: The interval during which cells synthesize a copy of their DNA in a process referred to as *DNA Replication*, and necessary prior to every cell division. Once cells commit to this phase, they decondense their nuclear DNA, the double-helix structure of the DNA is split into two single strands, and each strand serves as a template upon which biochemical reactions assisted by DNA polymerase enzymes synthesize double-stranded DNA that are exact (or nearly exact) copies of one another.
- | G2 Phase: The interval following the completion of DNA Synthesis (S Phase) and prior to mitosis (M Phase).
- | M Phase: The cell undergoes mitosis, or the active process by which cells traffic their components including each duplicate set of DNA to polar opposite ends, and subsequently cleaves to produce two daughter cells with nearly identical DNA to each other and their parent cell. The most

significant potential for differences between the cells is the result of uncaught errors during the DNA Synthesis process.

- | G0Phase: An optional phase of the cell cycle where a cell might choose to leave the G1 Phase and enter a state of arrest or quiescence. This phase is enables cells to regulate proliferation rates.

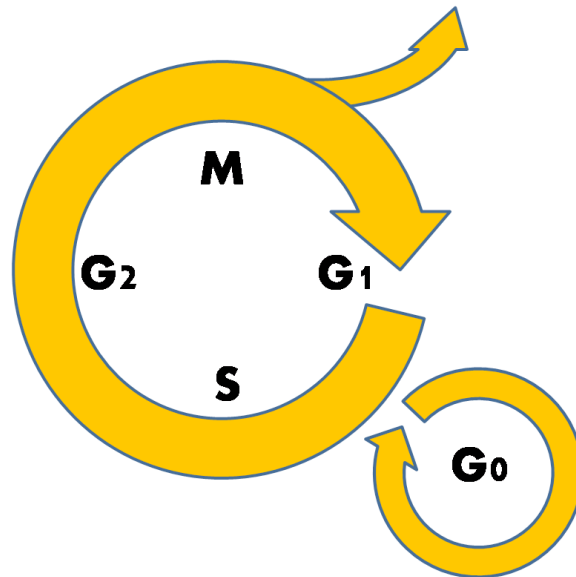


Figure 1.2: The Cell Cycle.

A cell's progression through the cell cycle is a well-coordinated process that is regulated by a family of proteins called Cyclin-Dependent Kinases (CDKs). These proteins are thought to undergo structural changes that advance the cell along the cell cycle phases. Thus, progression through the cell cycle is regulated by the activation and deactivation of CDKs, which are tied directly to the transcription of CDK genes into proteins, their degradation, and phosphorylation events that modify their structure. Built into the complex machinery governing cell cycle regulation are negative controls acting as “checkpoints” that primarily work to arrest the cell cycle if the integrity of the genome

has been compromised. These checkpoints are responsible for keeping healthy, normal cells from progressing through the cell cycle and passing genetic flaws on to their progeny. During checkpoint-initiated cell cycle arrests, normal cells have the opportunity to correct the compromise to the DNA integrity (*e.g.*, DNA repair of missing or mismatched base-pairs) before proceeding back along the cycle (Hartwell & Kastan, 1994).

The most prominent feature of a healthy cell cycle is its vigilant surveillance of the integrity of the genome before it is passed on to the following generations of cells. Disruptions to this surveillance can result in genetic instabilities characteristic of cancer cells. This is supported by experimental evidence pointing to a number of cancer types overcoming checkpoints due to mutations and abnormalities to the genes and corresponding proteins involved in the normal, healthy regulation of the cell cycle (Hartwell & Kastan, 1994).

WHAT ARE THE KEY FEATURES OF CANCERS?

The cancer research community has embraced six key features as necessary for normal cells to transition into invasive cancerous cells (Hanahan & Weinberg, 2000):

1. Self-sufficiency in growth signals: Whereas normal cells are dependent upon external stimuli to induce a proliferative state, cancer cells are autonomous in this regard by acquiring the ability to induce proliferation in the absence of externally derived growth signals.
2. Insensitivity to growth inhibitory signals: Normal tissues maintain a homeostasis with anti-proliferative signals that keep cells from leaving a quiescent state into a proliferative state. In order to evade this boundary, cancer cells acquire a phenotype that is insensitive to growth inhibitory signals.

3. Unbounded proliferative potential: Normal cells have an intrinsic program that restricts their multiplication. This program appears to operate independently of proliferation signaling pathways responding to growth and inhibitory factors. Cancer cells acquire the ability to overcome this finite limit on replication.
4. Ability to evade cell-initiated death (apoptosis): The ability for cancer cells to expand into tumors is not only due to their increased potential to proliferate, but also their resistance to commit suicide via programmed cell death, or apoptosis.
5. Ability to recruit vasculature via angiogenesis to feed the burgeoning tumor: Without another source of nutrients and oxygen to grow, cancer cells proliferate into tumors of a size limited by the local availability of nutrients and oxygen. In order to grow beyond this limit, tumors are able to recruit blood vessels via a process called angiogenesis.
6. Ability to invade local host tissue and metastasize to alternate locations in the body: Almost all deaths related to cancer are the result of tumors invading local tissue and metastasizing to other locations in the body. Changes in cells' adhesive and migratory properties enable cells to escape their primary tumor site. This ability to invade and metastasize is an important distinction between *benign* and *malignant* tumors. Cells that are just relentlessly growing into a *benign* tumor mass, remain clumped together, and can be cured by surgical resection are not considered cancer. However, cells that have achieved the ability to invade and form metastases are difficult to eradicate because of their capacity to spread throughout the body, and are considered cancer due to this malignant nature.

CARCINOGENESIS OCCURS AT THE DNA LEVEL

For an abnormal cell to pass on its altered traits to its descendants, these traits must be heritable. Altering a cell's DNA sequence, or genetic changes, is one method to pass on traits to the next generation of cells. A second option is to pass on a change in how the DNA is expressed by the cell, or what is referred to as epigenetic changes. In the last decade, epigenetic changes have been reported to be a cause in many cancers, and are considered to be as causative of cancer phenotypes as genetic changes (Miranda, *et al.*, 2006). Traditionally, carcinogenesis has been associated with genetic changes, while epigenetic changes are associated with promoting its progression. This is supported by the correlation between carcinogenesis (the initialization of normal cells transitioning to cancerous behaviors) and mutagenesis (the generation of genetic mutations) caused by accidental errors during the DNA replication process prior to mitosis or stimulated by chemical carcinogens, ionizing radiation, and viruses that are all capable of altering the DNA of a cell. For example, cigarette smoking is considered to be a chemical carcinogen because of the well-known association between smokers and the development of lung cancer. However, preliminary evidence has pointed to the potential for normal cells to be made susceptible to developing cancer phenotypes by epigenetic changes upon acquisition of a genetic mutation, suggesting a more direct role for epigenetic changes in carcinogenesis (Feinberg, 2004). Regardless of strength the role it plays in carcinogenesis, epigenetic changes and its interplay with genetic changes is well-recognized in the cancer community as a fundamental cause of cancer.

Cancer is not the result of a single mutation. Rather, it is thought to be the result of a series of mutations whose additive effects along with epigenetic changes enable a cell to meet the six key features of cancer listed above (Hanahan & Weinberg, 2000). These conditions are thought to cause cancers to be genetically unstable in comparison to

normal cells, artificially accelerating their evolution process by increasing the rate at which their DNA mutates and selecting for species that overwhelm the normal cell population and develop phenotypes that are highly proliferative, motile, and drug-resistant. For example, normal cells have the ability to repair damage or errors to the DNA such as mismatches between base-pairs and missing/excised base-pairs; mutations to the genes controlling these repair functions predisposes cells to developing cancer by promoting genetic instability and effectively increasing the normal rate of mutagenesis, thereby speeding up cellular evolution.

The number of cell divisions a person experiences increases with age, and thus more likely to compile enough mutations that lead to cancer. This is thought to be the underlying cause for why cancer incidence appears to be a function of age. However, uncaught mutations by the cell-cycle checkpoints or mutagens having stimulated DNA damage artificially alter these natural dynamics and increase the likelihood for a cell to pick up a mutation at each cell division. Mutagens effectively increase the probability of a cell to pick up mutations, and the likelihood of it eventually becoming cancer.

CANCER METASTASIS

The ability of cancer cells to invade local host tissue and metastasize to alternate sites is partially responsible for the difficulty associated with eradicating cancers with current treatment options. To successfully metastasize, cells must be able to loosen their adhesion to their neighbors, escape from the primary tumor, burrow through local tissues until they reach an escape route such as a blood or lymphatic vessel, cross the basal lamina and endothelial lining of the vessel so as to reach the circulation, exit from the vessel at an alternate site in the body, and then survive and proliferate in the new environment. In order to achieve each of these steps, cells must acquire properties such as

decreased adhesion, secretion of matrix-degrading enzymes, and directed motility; at the tumor scale, recruitment of blood vessels via angiogenesis is also a signature marker for a tumor's metastatic capacity. Experimental evidence has shown that many carcinomas exhibit decreased expression of cell-to-cell surface adhesion molecules and the ability to secrete proteolytic enzymes from the invading leading edge of metastatic cells. Furthermore, it has been suggested that though directed cell motility may not be a sufficient condition for metastasis, it is necessary (Bracke, *et al.*, 2008). This has been experimentally confirmed by the *in vitro* motile behavior of cell lines considered to be models of metastasis.

HOW ARE CANCERS CLASSIFIED?

Cancers are broadly classified by the tissue and cell type from which they derive. Those that arise from epithelial cells are referred to as *carcinomas* while those that arise from connective tissue or muscle cells are referred to as *sarcomas*. Cancers that are derived from blood cells and the immune system are referred to as *leukemias* and *lymphomas*, respectively. About 90% of all diagnosed human cancers are carcinomas, which is because epithelial cells are the most proliferative cells in the body and they are more exposed to damage from the environment, thus increasing their likelihood of acquiring mutations that lead to cancer. Cancer cells often retain characteristics of the tissue from which they derived. For example, skin cancer cells (melanoma) typically continue to produce pigment granules. This property makes it possible for medical professionals to derive the origin of biopsied cancer cells to determine whether or not the tumor in question is a metastasis or primary tumor.

HOW ARE CANCERS CURRENTLY TREATED?

The major classes of treatment available to cancer patients are surgery, chemotherapy (cytotoxic, anti-angiogenic, and targeted), hormone therapy, and radiation therapy. Most common cancers can be treated with one or more of these classes of therapy in some form. Clinicians devise treatment plans by considering the information available regarding their patients' cancer, such as its type, location, grade, stage, size, morphology, molecular characteristics, patient's medical history of treatment, patient's general health, vascularization status, whether or not it has metastasized, and general treatment response profiles of patients having suffered from the same disease in the past. In the process of devising their plan, Clinicians must determine which combination of treatment options will provide the patient with the best outcome, and in doing so, must also consider the intention of each treatment option:

- | Curative: The treatment option is intended to cure the patient of the disease
- | Adjuvant: The treatment is intended to be used in combination with another treatment option concurrently
- | Neoadjuvant: The treatment is intended to be administered prior to another treatment option (usually surgery or radiation therapy) so as to reduce the tumor size and extent, and maximize the chance for the subsequent therapy(ies) to succeed
- | Therapeutic: The treatment can be curative and benefits the patient's expected survival; it is expected to halt, delay, reduce, and or reverse tumor progression with subsequent improvements in patient health
- | Palliative: The treatment is not expected to be curative or therapeutic; rather its purpose is to minimize the severity of disease symptoms, relieve patient suffering, and improve the patient's quality of life

Surgery

Surgery is often a curative treatment option for patients whose tumors have not yet metastasized, the clinicians are confident that they can define the boundaries of the tumor's extent in the patient, and the proposed surgical excision of the tumor will not knowingly fatally damage organ systems. Surgery can also be a first line of therapy for patients even if the tumor has metastasized when the clinicians determine that the patient's best prospects involve removal of as much of the primary and secondary tumors as possible via surgery, followed by some combination of radiation therapy, chemotherapy, and hormonal therapy in hopes of eradicating any remnant cancer cells. If the cancer has been diagnosed at a point in the cancer progression where the primary tumor has excessively invaded the local host tissue and metastasized to the lymph nodes (an indication that the cancer has spread to multiple sites throughout the body), surgery is usually not recommended because it would place far too much burden on the patient, and likely be more detrimental to the patient's health than other treatment options.

Radiation Therapy

Radiation therapy is the medical use of ionizing radiation to induce damage to cells' DNA. This damage to the DNA can induce apoptosis or be passed on to further generations of the cancer and cause them to reproduce at a slower pace. Availability of oxygen in the tumor microenvironment is a consideration for the clinician when devising a treatment plan to use radiation therapy because oxygen helps form DNA-damaging free radicals. The response of cancer cells to radiation therapy also depends on the tumor type, its size, and the corresponding radiosensitivity. Neoadjuvant chemotherapies can help increase radiosensitivity of a tumor by shrinking it prior to radiation treatment. Some adjuvant chemotherapies are known to enhance radiosensitivity of the tumor as well. While the radiation therapy has its advantages in regards to inducing DNA-damage to

cancer cells, it can also cause such damage to healthy cells which can result in new cancer cells in the patient's future. In other words, radiation therapy itself is a mutagen, thus limiting a patient's exposure to ionizing radiation is critical concern of clinicians. Furthermore, radiation therapy is often limited to localized treatments because it would be harmful, as well as impractical, to subject a patient's whole body to ionizing radiation; thus it is primarily used to treat non-metastatic cancer.

Chemotherapy

Chemotherapies are drugs that are administered intravenously or orally depending on the class of chemotherapy, the specific type, and the clinician designing the dosing cycle. These drugs reach their targets through the blood stream. Ironically, the tumor's own recruitment of blood vessels to sustain growth acts as a route for drugs to reach their intended targets. However, these drugs typically have harsh side effects and must be used rationally and cautiously.

Cytotoxic Chemotherapy

Cytotoxic chemotherapies are chemical drugs that kill cells. Often, this class of chemotherapy targets cells that rapidly divide due to the cell-cycle specific nature of their mechanism of action. For example, the drugs doxorubicin and cisplatin target the DNA of actively cycling cells, and are most effective during S-phase when the DNA is most susceptible to drug-DNA interactions. However, cytotoxic chemotherapy is often viewed as a double-edge sword; because of its non-specific nature, cytotoxic chemotherapy drugs also target normal cells that are actively replication such as in the bone marrow, digestive tract, and hair follicles resulting in decreased production of blood cells, inflammation of the digestive tract, and hair loss, respectively.

Targeted Chemotherapy

Targeted chemotherapies are a newer class of drugs meant to specifically interact with proteins involved in the signaling pathways promoting cancerous behaviors (*e.g.*, proliferation, motility, lack of adhesion). These drugs are rationally designed compared to cytotoxic therapies which are relatively non-specific. For example, endothelial growth factor receptor (EGFR) is overexpressed in many cancer types, and is the target of many small molecule-based targeted therapies. In addition to small molecules, another class of targeted therapy is monoclonal antibody-based therapies (also sometimes referred to as Biologics or Biotherapeutics), where antibodies that specifically target surface receptors on cancer cells are produced in bulk in bioreactors. In some cases, the antibody itself can induce apoptosis once it binds to its targeted receptor (such as Rituximab which targets CD20 receptors). In other cases, the antibody binds to its target and interferes with the activity of the receptor, such as Herceptin targeting Her2/neu receptors expressed in about 25% of breast cancers and is responsible for stimulating proliferation. Radioimmunotherapies take advantage of the targeted nature of monoclonal antibodies to deliver an attached radioactive isotope to the local vicinity of cancer cells to induce cell death. For example, Ibritumomab tiuxetan targets CD20 receptors and kills cells by radioactive damage induced by an attached radioactive isotope.

Anti-Angiogenic Chemotherapy

Anti-angiogenic treatments are meant to inhibit further growth of tumors by blocking the formation of new blood vessels that supply growing tumors with oxygen and nutrients. Bevacizumab was the first clinically approved anti-angiogenic drug available on the market. It inhibits angiogenesis by targeting and inhibiting the function of vascular endothelial growth factor, which is a key promoter of new vessel formation.

Hormonal Therapy

Several cancers, most prominently prostate and breast cancer, are well-known to have their growth linked to the expression of steroid hormone receptors. Hormone therapies operate through administration of specific hormones or hormone antagonists to manipulate cancer cells hormone receptor activity. For example, tamoxifen is the standard treatment given to breast cancer patients whose tumors express Estrogen Receptor, which is about 70% of cases. Tamoxifen is an antagonist of the Estrogen Receptor in breast cancer, thus inhibiting estrogen from binding to the receptor and activating proliferative signaling pathways. Similarly, antagonists of the Androgen Receptor are used to treat prostate cancer in order to inhibit androgens from binding to the receptors.

Chapter 2: Introduction

PREFACE

Segments of this introduction chapter are based on a published article, which was a collaborative effort between Sandeep Sanga, Dr. John Sinek, Dr. Hermann B. Frieboes, Dr. Mauro Ferrari, Dr. John P. Fruehauf, and Dr. Vittorio Cristini. Sanga took the lead role in preparing the manuscript. Sinek and Frieboes assisted in the manuscript's preparation. Ferrari and Fruehauf provided related expertise in nanotechnology and oncology. Cristini conceived of the article's premise and directed its progress. The modeling approach towards developing a modular, multi-scale virtual simulator of cancer for predicting tumor progression and response to therapy shaped the theme for Sanga's computational systems biology research presented in this dissertation.

Note: Portions of this chapter are based on an article originally published as (Sanga, *et al.*, 2006) in *Expert Review of Anticancer Therapy*, which can be accessed at <http://dx.doi.org/10.1586/14737140.6.10.1361>, and have been included in this dissertation with permission from Future Drugs, Ltd.

THE MULTIFACETED NATURE OF CANCER

The physiological processes underlying cancer are highly complex, spanning a wide range of interrelated temporal and spatial scales. The fundamental causes are believed to reside at the molecular scale where gene mutations and epigenetic changes lead to aberrant signaling pathways, enabling cells to develop a selective advantage that allows them to reproduce in defiance of normal constraints. In time, these cells form avascular masses limited to approximately a few millimeters in diameter due to the transport limitations of oxygen and nutrients into tissue (Folkman, 1971). As inner layers of the nascent tumor begin to necrose, tumor angiogenic regulators (*e.g.*, VEGF) are released by the avascular tumor mass, which diffuse through the surrounding tissue and trigger a cascade of events upon arrival at local vasculature, culminating in the recruitment of vessels that supply blood to the burgeoning tumor (*i.e.*, angiogenesis). At this point, the vascularized tumor may remain compact and non-invasive, *i.e.*, benign, in which case it can usually be successfully removed by surgical resection or treated with radiation. Conversely, upon receiving infusion of nutrient from its newly formed vasculature, a tumor may become malignant and rapidly invade local tissue, usually acquiring mutations that lend its cells the ability to navigate through the bloodstream and lymphatics to metastasize to non-native locations in the body (Folkman, 1971). The non-localized nature of metastatic cancer limits the success of surgical and radiation treatment approaches, thus systemically administered chemotherapy continues to be the standard option in spite of marginal results (Tanaka, *et al.*, 2009).

Tumor neovasculature plays an integral role in the administration of such treatment. It is the first tumor-level barrier, which an administered drug molecule must navigate on its journey to its intended intracellular target, and its anatomical and

functional irregularities are thought to significantly impair drug distribution to lesion tissue (Jain, 2001a; Pries, *et al.*, 2009). For standard therapy, once drug molecules extravasate through vasculature, they must diffuse through interstitial space, permeate cell membranes, survive a gauntlet of intracellular drug resistance mechanisms designed to detoxify cells (Wayne Materi & David S. Wishart, 2007; W. Materi & D. S. Wishart, 2007), and finally bind to subcellular targets at sufficient cytotoxic concentrations (Jain, 2001a). This multi-scale series of barriers parallels the multifaceted nature of the disease, and combines to produce an overall reduction in the efficacy of many unrelated anticancer drugs (Krishna & Mayer, 2000) and cannot be overcome simply by administering more drug, as toxicity to host tissue presents a formidable challenge.

THE PUSH FOR PERSONALIZED CANCER MEDICINE: PREDICTIVE ONCOLOGY

Traditionally, cancer therapy has been planned according to the organ or tissue from which the cancer originated. With the advance of experimental techniques to probe the molecular nature of the disease, this approach has become antiquated as our understanding of the molecular origins of cancer and the underlying intracellular signaling pathways that drive the cancer phenotypes has improved. Knowledge that cancer is the product of a series of random genetic and epigenetic changes has partially explained why patients with cancers of similar organ/tissue origins can respond with variable sensitivity to anticancer agents, and identifies an important obstacle that must be overcome towards improving our success in treating cancer. This heterogeneity in the patient population of any one particular cancer type has led to a paradigm shift in cancer treatment from a 'one size fits all' to a more personalized approach.

In order to devise personalized treatment plans, clinicians use patient-specific information called *biomarkers* that are indicators of the particular genetic defects

underlying the patient's tumor. Currently, biomarkers are derived from analysis of blood serum, biopsied tissue, or surgically excised tissue. Biomarkers assist clinicians in managing their patient's care by predicting a patient's clinical outcome and their response to different classes of therapies. Clinicians can weigh the severity of the outcome prediction against the combinations of therapy options (*e.g.*, surgery, radiation therapy, chemotherapy, hormonal therapy) and rationally devise a plan.

Personalizing cancer medicine is important paradigm shift as it helps to classify patients who are more likely to respond to a therapy regimen based on molecular scale knowledge of a patient's tumor. Just as important is predictive knowledge of whether a patient is unlikely to respond so as to avoid giving patients unnecessary adjuvant chemotherapy. The side effects of radiation therapy and chemotherapy are harsh and can immensely reduce the quality of life for a patient (Mols, Vingerhoets, Coebergh, & van de Poll-Franse, 2005). This is particularly a problem in the United States where it has been shown that approximately 55-75% of women with early stage breast cancer undergo unnecessary adjuvant chemotherapy (van't Veer & Bernards, 2008).

With the completion of the Human Genome Project and the corresponding advance of gene expression microarray chips in the last decade, scientists and clinicians now have access to technologies capable of simultaneously monitoring the gene expression of the full human transcriptome. This has led to studies comparing the gene expression between diseased and healthy human tissues to identify the genetic basis for a particular diseased tissue such as cancer. Similarly, the field of pharmacogenomics investigates the genetic basis underlying the variability in drug response between patients. This technology has led to a new class of molecular diagnostic tests for cancer: gene expression-based molecular signatures. Such signatures have already led to commercially-marketed diagnostic tests, such as MammaPrint® by Agendia, which

assesses risk of metastasis in breast cancer patients and is based on a 70-gene signature (van de Vijver, *et al.*, 2002). Furthermore, because gene expression is regulated by transcription factors (proteins that are themselves gene products); statistical associations between transcript abundance (the quantity measured by gene expression microarrays) can infer gene regulatory mechanisms and reconstruct signaling relationships in datasets of microarray data with large enough samples (Margolin, *et al.*, 2006; Soinov, 2003). Elucidating the multitude of gene interactions involved in a cancer's progression can help devise rational strategies for overcoming drug resistance and define molecular targets for therapy (Edgerton, Fisher, Tang, Frey, & Chen, 2007).

However, despite our increased understanding of the molecular origins of cancer and subsequent improvements in the design of treatment plans, chemotherapy still experiences limited success in many patients, which suggests that success of systemically-delivered therapies may also be associated with their overcoming biobarriers at the courser scales (*i.e.*, cellular, tissue, organ). An anticancer agent faces a multitude of barriers along its route from the point of administration that prevent it from reaching its intracellular targets at the appropriate levels, and thus reducing the efficacy. The body recognizes drugs in circulation as toxins, and consequently acts to filter the drug from the blood stream. In the case of nanoparticle-mediated drug delivery, the reticuloendothelial system (RES), a system of macrophages and specialized cells lining the liver, spleen, bone marrow, and lymphatic tissues, acts to sequester and remove particles from circulation (ElBayoumi & Torchilin, 2009). The irregular hemodynamics of tumor vasculature compromises blood circulation in lesions, and consequently impairs transport of circulating anticancer agents to the tissue interstitium (Jain, 2001a; Pries, *et al.*, 2009), where they must diffuse and permeate cell membranes, survive intracellular

drug resistance mechanisms, and finally reach their intracellular target at sufficient concentrations to induce cell inhibition (death or quiescence).

THE ROLE OF COMPUTATIONAL SYSTEMS BIOLOGY IN CANCER RESEARCH AND THERAPY DESIGN

Developing a detailed understanding of the underlying pathophysiology of cancer, its progression, mechanisms of drug resistance at various scales, as well as the optimization of drug dosing protocols, is the subject of vast amounts of research directed towards the development of effective treatment and prevention strategies. Because of the complexity of this disease, it has proven difficult to assign quantitative weights to each component. This may be due in part to the often reductionist nature of experimental investigation where mechanisms are often studied in an isolated context. It has been suggested that a conceptual framework is necessary to fully understand the data produced in quantity by tumor biologists and clinical oncologists (A. R. Anderson & Quaranta, 2008; Araujo & McElwain, 2004; R.A. Gatenby, 1998; R. A. Gatenby & Gawlinski, 2003b; Wayne Materi & David S. Wishart, 2007; W. Materi & D. S. Wishart, 2007). The challenges of better understanding the overall cancer phenomenon and its treatment can benefit from a computational systems biology approach, where the disciplines of medicine, biology, physics, chemistry, computer science, and engineering are integrated to study cancer on a multi-scale, system-wide basis. In particular, this approach can be integrated with the vast amounts of data already available from biological experiments and clinical trials to calibrate and test predictive models, and eventually to rationalize their focus.

Traditional clinical and biological experiments require costly investments in both time and materials and are often limited by equipment precision, human error, and the inability to distinguish between various underlying mechanisms governing tumor growth

(Araujo & McElwain, 2004; Kunz-Schughart, Kreutz, & Knuechel, 1998). In parallel, a critical weakness of theoretical models is their plasticity in uncritically recapitulating training data, without regard to the models' actual validity and predictive capability. Nevertheless, modeling can provide investigators with tools to run computational experiments that would otherwise be very difficult or impossible to recreate in an experimental setting; accordingly, modeling can provide valuable information to plan effective biological experiments for testing theoretical hypotheses. Data from biological experiments provides necessary constraints for choosing appropriate model structure and parameters. Therefore, although pure theoretical or experimental investigations alone have inherent flaws and limitations, an ideal synergy between the two can be approached by using a circular, recursive work flow methodology characteristic of computational systems biology (Figure 2.1).

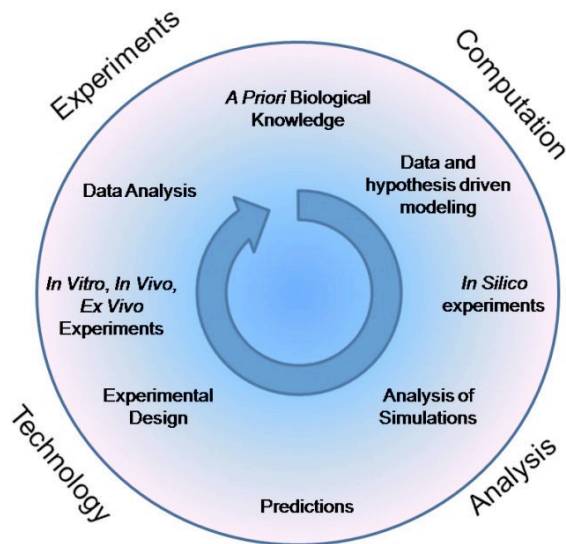


Figure 2.1: Systems Biology Paradigm: Hypothesis-driven research in systems biology

A cycle of research begins with the selection of contradictory issues of biological significance and the creation of a model representing the phenomenon. Models can be created either automatically or manually. The model represents a computable set of assumptions and hypotheses that need to be tested or supported experimentally. Computational "dry" experiments, such as simulation, on models reveal computational adequacy of the assumptions and hypotheses embedded in each model. Inadequate models would expose inconsistencies with established experimental facts, and thus need to be rejected or modified. Models that pass this test become subjects of a thorough system analysis where a number of predictions may be made. A set of predictions that can distinguish a correct model among competing models is selected for "wet" experiments. Successful experiments are those that eliminate inadequate models. Models that survive this cycle are deemed to be consistent with existing experimental evidence. While this is an idealized process of systems biology research, the hope is that advancement of research in computational science, analytical methods, technologies for measurements, and genomics will gradually transform biological research to fit this cycle for a more systematic and hypothesis-driven science. This image and figure caption have been adapted from (Kitano, 2002) with permission from AAAS.

MULTI-SCALE CANCER SIMULATION PARADIGM

As the saying goes, "Rome wasn't built in a day." Likewise, the effort to develop a comprehensive, dynamic simulator of the cancer disease, let alone the entire mammalian system, is an intricate, arduous process. Nevertheless, its development has important implications towards predicting tumor progression and response to therapy on a patient-specific basis (a research field known as *Predictive Oncology*).

Our research group is taking an integrative, multi-scale approach towards developing a cancer simulator with the capacity to predict *in vivo* tumor growth and response to therapy, where the long-term goal is to provide clinicians with a tool to plan surgical interventions, radiation therapy, and chemotherapy based on patient-specific tumor information, thereby maximizing benefit to both patients and health-care providers. The endeavor to develop software packages capable of sophisticated, *in vivo*-like tumor simulation requires a modular, multi-scale development process where individual components are built upon models simulating disease progression at multiple resolutions, anticancer drug pharmacology, and drug resistance mechanisms. These models are then linked to simulate the disease and possible therapies through a wide temporal range.

RESEARCH PURPOSE

The purpose of this research dissertation is to emphasize the benefit of computational systems biology approaches in studying cancer at multiple, inter-related scales, and its role in personalizing cancer medicine through advancing a multi-scale, predictive model(s) of solid tumor growth and progression towards invasion and metastasis. In the next chapter, we review methods and techniques used by predictive oncologists most relevant to the research performed in this dissertation. In the subsequent chapter, we describe the multi-scale cancer modeling framework in perpetual development by our research group, which forms the basis of the research conducted in the following chapters.

In Chapter 5, we take a significant step towards a predictive model for tumor response to therapy by formulating drug delivery (pharmacokinetics) and drug effect (pharmacodynamics) models whose parameters we calibrate from *in vitro* experiments reported in the literature. The drug delivery model considers drug extravasation from tumor vasculature, its penetration into cells, drug resistance mechanisms, and drug

binding to intracellular targets. The drug effect model takes into account the local availability of oxygen and nutrients and their impact on chemotherapy efficacy. We model the delivery and effect of cytotoxic chemotherapy under a best-case scenario, where the cells comprising the tumor are molecularly homogenous and drug-sensitive, which is far from the heterogeneous reality. The simulations presented in this study demonstrate that the limited efficacy of drugs is at least in part due to barriers beyond the molecular difference between tumors.

In Chapter 6, we leverage publicly available breast cancer gene expression data performed on microarray technology to infer cell signaling networks and predict breast cancer phenotypes. We present a gene expression meta-analysis studying the molecular characteristics of a subtype of breast cancer patients that may benefit from targeted therapy. While this molecularly-defined subtype was first identified by two previous gene expression studies, we use contemporary normalization methods to confirm that the two studies identified the same subtype, and that this subtype appears to persist in the breast cancer. In addition to defining a more robust gene signature for predicting this breast cancer subtype, we use gene regulation network inference methods to hypothesize signaling interactions, some of which have already been established through experimentation with cell lines and others which can guide future experiments. By generating hypotheses as to the network interactions defining this breast cancer phenotype, we are identifying potential targets for targeted therapy. One of the key molecules associated with this breast cancer subset is the Androgen Receptor, which is a promoter of proliferation in prostate cancer and is targeted for prostate cancer therapy. Furthermore, our network reconstruction process is important in a multi-scale cancer modeling framework by hypothesizing the “connection map” of interacting signals that once validated experimentally, can form the basis of signaling models driving the behavior of cellular-scale models.

In Chapter 7, we present a two-dimensional cell-scale model for chemotaxis whose motility is an emergent property of remodeling of the cytoskeletal structure in response to receptor-ligand dynamics at the cell perimeter. Chemotaxis is generally accepted as a model for cancer metastasis (Bracke, *et al.*, 2008; J. Condeelis, Singer, & Segall, 2005) because of the necessity of cells to have the ability to mobilize away from the primary tumor mass. We apply receptor occupancy theory to model the generation of signals promoting protrusion activity at the cell periphery as a function of the percentage of receptors occupied by ligand at the cell surface. We calibrate and test the model with experimental data from the literature of cells responding to controlled chemoattractant gradients. Our simulations support that the underlying mechanisms of directed motility are conserved across cell-types, and help to elucidate how motile cells sense gradients in their microenvironment. This model is a critical step towards linking information derived from cell signaling models into a multi-scale cancer modeling framework to predict cellular behaviors, such as cell motility.

Chapter 3: Techniques and Methods used in Predictive Oncology

Note: Portions of this chapter are based on an article originally published as (Sanga, *et al.*, 2006) in Expert Review of Anticancer Therapy, which can be accessed at <http://dx.doi.org/10.1586/14737140.6.10.1361>, and have been included in this dissertation with permission from Future Drugs, Ltd.

The purpose of a model is to describe a large class of observations and make definitive predictions about the results of future observations. Numerous models have been reported in the literature predicting behavior for a wide range of phenomena, each asserting to provide additional insight about the phenomenon in question. The extent to which models describe the underlying mechanisms of a process determines the model's level of empiricism. There are two schools of thought to predictive modeling: 1) empirical, data-based models are thought to have more practical value because they can overcome the complexity and unknown mechanisms of the underlying phenomena, whereas 2) theoretical, mechanistic-based models are considered to have more predictive potential because they are built upon models of underlying mechanisms (either by *a priori* knowledge or by hypothesis) and are able to generate new knowledge (Nestorov, Hadjitodorov, Petrov, & Rowland, 1999). Predictive oncology research reflects both schools of thought, and the purpose of this chapter is to present a comprehensive overview of the approaches used in predictive oncology today, and extend upon those that are most relevant to the research presented in this dissertation. As we have seen thus far, the dynamics of cancer span a wide range of scales ranging from events at the molecular scale initiating carcinogenesis all the way to tissue scale where the bulk tumor behavior can invade, metastasize, and overwhelm the human body. This range of scale is reflected in the wide variety of modeling techniques used to study the disease.

MECHANISTIC CANCER MODELING APPROACHES

“Mathematical models can provide biologists and clinicians with tools that might guide their efforts to elucidate fundamental mechanisms of cancer progression and either improve current treatment strategies or stimulate the development of new ones” (Alarcon, Byrne, & Maini, 2003). Many mechanistic cancer models have been proposed that focus on one or more phases of cancer progression (*i.e.*, avascular, angiogenesis, vascular), and can typically be categorized as either a *continuum*, *discrete*, or *hybrid* approach (Alarcon, *et al.*, 2003; Araujo & McElwain, 2004; Moreira & Deutsch, 2002). Continuum-based models draw upon principles from fluid and continuum mechanics, and describe cancer-related variables, such as cell population, nutrient concentration, oxygen distribution, and growth factor concentrations as continuous fields by means of differential equations (Alarcon, *et al.*, 2003; Zhang, Wang, Sagotsky, & Deisboeck, 2009). Alternatively, discrete-based approaches describe the dynamics of discrete elements (*e.g.*, tumor cells), whose states are governed by a set of deterministic and/or probabilistic rules. These rules can be translated from known biological processes or hypothesized by the investigating researcher in a straightforward manner. The state evolution of these elements can be tracked through both space and time. In the case of cellular-automaton (CA) models, the behavior of the discrete elements and their microenvironment is constrained to a two- or three-dimensional lattice of grid points. Agent-based models (ABMs) are similar to CA models, and can be considered a more general form of CA models; the primary difference is that ABMs are not constrained to a pre-determined geometric arrangement. Both CA models and ABMs enable researchers to enclose a complex set of biologically-based rules into each “agent” and simulate emergent multi-cellular behavior. A recent trend has been to couple ABMs with continuum approaches to model intracellular cell signaling events and trigger the rules driving the agent phenotypes, as well as to model species in the microenvironment such as diffusing oxygen from local vasculature. The multi-scale complexity of tumors and their microenvironment make it difficult to describe all the relevant processes

sufficiently with using solely a continuum-based or discrete-based approach. Rather, hybrid models are becoming increasingly popular because of their ability to simulate cancer through a wide temporal and spatial range, while optimizing computational expense. Hybrid cancer modeling approaches combine continuum fields with discrete descriptions. In particular, substances such as oxygen, nutrient, drug, and growth factors can be described as a continuum in the tumor microenvironment, while individual CA or ABM elements dynamically evolve in response to local substance concentrations.

The scientific literature devoted to the theoretical investigation of tumor growth and progression using continuum and CA modeling approaches has been reviewed in depth (Araujo & McElwain, 2004; Mantzaris, Webb, & Othmer, 2004; Moreira & Deutsch, 2002). While most of the models described within these reviews are able to provide useful insight into cancer-related processes occurring at a particular length and time scale, they do not “address the fundamental problem of how phenomena at different scales are coupled (Alarcon, *et al.*, 2003).” The multi-scale complexity of cancer progression warrants a multi-scale modeling approach to be taken to produce truly predictive tumor simulators. Processes occurring at various length and time scales must be coupled appropriately in order to capture all the dynamics involved. Previous works have developed multi-scale systems modeling complex biological processes, such as cancer (Alarcon, *et al.*, 2003; Alarcon, Byrne, & Maini, 2005; Jiang, Pjesivac-Grbovic, Cantrell, & Freyer, 2005; Ribba, Colin, & Schnell, 2006), the heart & lung (Crampin, *et al.*, 2004; Hunter, Kohl, & Noble, 2001; Lagana, *et al.*, 2005; Smith, *et al.*, 2002), and various phenomena related to developmental biology (Chaturvedi, *et al.*, 2005; Sharp, Reinitz, & Mjolsness, 1993). In particular, Jiang *et al.* (Jiang, *et al.*, 2005) and Alarcon *et al.* (Alarcon, *et al.*, 2003, 2005) present frameworks for building multi-scale cancer progression models capable of integrating a hierarchy of processes at varying length and time scales. Most cancer models and multi-scale systems (Alarcon, *et al.*, 2003, 2005; Araujo & McElwain, 2004; Burton, 1966; H. M. Byrne & Chaplain, 1995; H. M. Byrne & Chaplin, 1996; H.P. Greenspan, 1972; Jiang, *et al.*, 2005; Maggelakis & Adam, 1990; Ribba, *et al.*, 2006; Thomlinson &

Gray, 1955) primarily produce 1D and 2D simulations that are limited in their ability to capture complex *in vivo* tumor morphologies and microenvironment.

Modular Development of a Multi-scale Cancer Simulation

The present section provides a brief overview of a select group of tumor growth, angiogenesis, and pharmacology models serving as the inspiration and foundation for efforts to develop a multi-scale cancer simulator by our research group (Figure 3.1), and the core model adapted in the research study presented in Chapter 5 to investigate drug delivery and effect to tumor lesions.

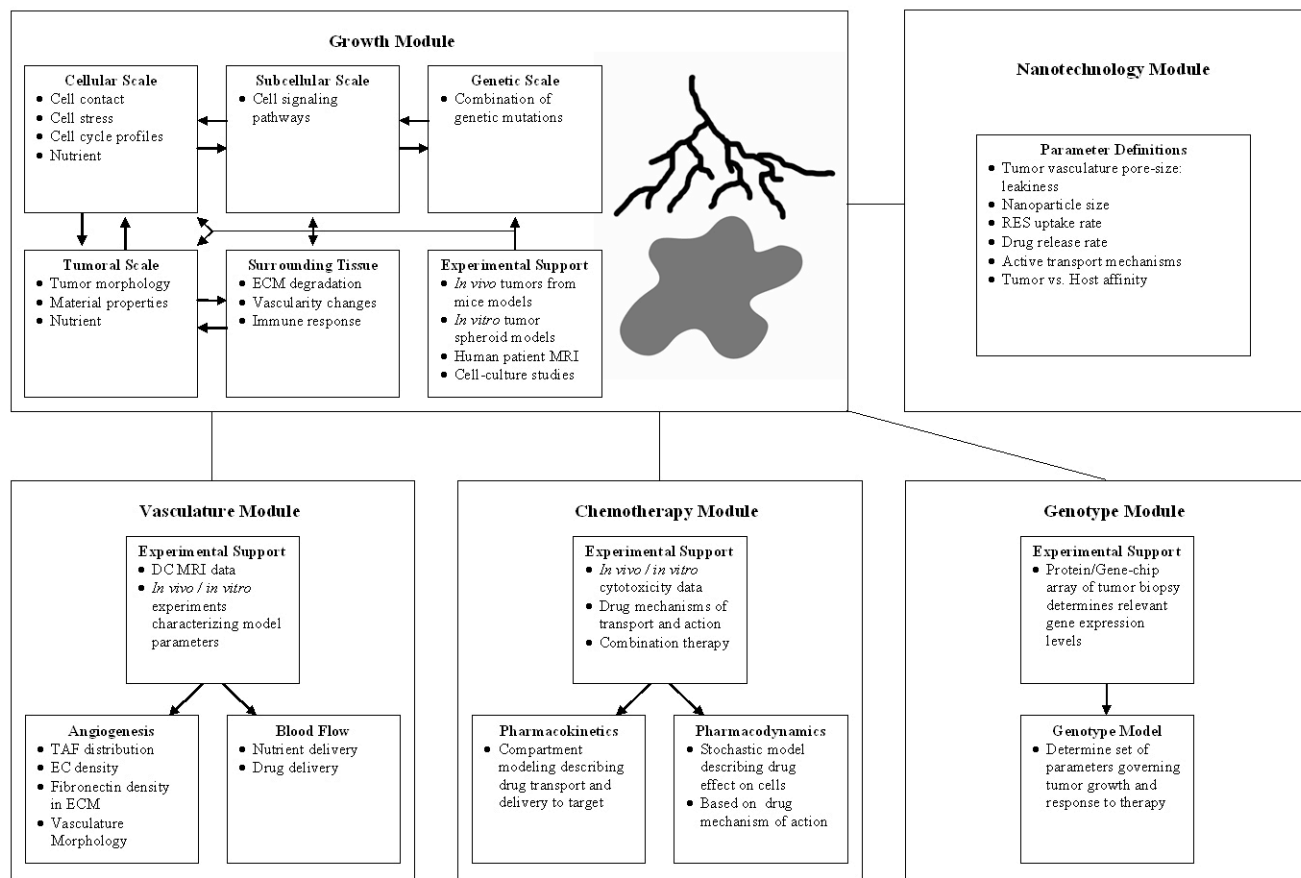


Figure 3.1: Outlook for multi-scale cancer simulation under development by Cristini and coworkers.

Multi-scale cancer simulation is founded on the integration of experimental data and mathematical models. It should provide valuable insights into the cancer phenomenon and establish a platform information technology for analyzing the effectiveness of chemotherapeutic drugs and rationally design new therapies. Modules are developed and coupled *via* sharing of information.

Tumor Growth

Cristini *et al.* (Cristini, Lowengrub, & Nie, 2003) were among the first to advance tumor modeling beyond the limited capabilities of mathematical linear analyses and into the realm of nonlinear computer simulation of complex tumor morphologies. The multifaceted nature of cancer requires sophisticated, nonlinear mathematical models to

capture more realistic growth dynamics and morphologies. Boundary-integral simulations (Cristini, *et al.*, 2003) of classic continuum-based tumor models (Burton, 1966; H. M. Byrne & Chaplain, 1995; H. M. Byrne & Chaplin, 1996; H.P. Greenspan, 1972; Maggelakis & Adam, 1990; Thomlinson & Gray, 1955) determined that a reduced set of two non-dimensional parameters (related to mitosis rate, apoptosis rate, cell mobility, and cell adhesion) regulate morphology and growth (invasiveness) of avascular and vascularized tumors. In this model there is no morphological representation of vasculature, rather the effect of vascularity is quantified by a parameter relating the concentration of nutrient in the blood, nutrient transfer rate from blood to tissue, and nutrient consumption by the cells. Critical conditions were predicted that separate compact, noninvasive mass growth from unstable, fingering, infiltrative progression (Cristini, *et al.*, 2003). However, further analysis demonstrated that highly vascularized tumors tend to grow in compact, nearly spherical shapes showing little or no signs of invasiveness. This unexpected prediction suggests that tumors could maintain stable morphology under normoxic microenvironmental conditions. This result is supported by experimental observations indicating that hypoxia stimulates invasiveness and tumor fragmentation (Kunkel, *et al.*, 2001a; Pennacchietti, *et al.*, 2003).

Tumor-Induced Angiogenesis

Angiogenesis is the process by which cancers recruit enhanced blood supply to provide the oxygen and nutrients that are commonly considered necessary to support growth into larger, more invasive tumor masses. As is the case with tumor growth, tumor-induced angiogenesis is a topic receiving considerable attention from the biological modeling community and has been extensively reviewed (Mantzaris, *et al.*, 2004).

Mathematical Models for Tumor-Induced Angiogenesis

Angiogenesis is believed to be initiated by pro-angiogenic proteins (PAP), such as vascular endothelial growth factor (VEGF), that have been induced by a lack of oxygen and nutrient to be released from the necrotic tissue of a tumor lesion into surrounding tissue (Folkman & Klagsbrun, 1987). These proteins create a chemical gradient that triggers endothelial cells (EC) from parent vessels in the pre-existing vasculature to migrate towards the tumor (chemotaxis). Eventually through a number of complex mechanisms, the accumulation of EC form finger-like capillary sprouts extending from a parent vessel. Analogous to plant growth, these sprouts extend and grow towards the tumor along the chemical gradient guided by the migration of the sprout-tip. The interaction between EC and the extracellular matrix (ECM), itself, is also significant in directing the sprout-tips. Fibronectin is generated and adheres to the matrix, serving to guide the direction of endothelial cell progression *via* a process called haptotaxis, similar and complementary to chemotaxis. Once capillary sprouts from the parent vessel extend far enough towards the tumor, they tend to lean towards each other and form tip-to-tip and tip-to-sprout fusions called anastomoses (thought to be caused by haptotaxis, or directed cell motility in response to substrate gradients) (A. R. Anderson & Chaplain, 1998; McDougall, Anderson, Chaplain, & Sherratt, 2002). Through this process of anastomosis, an initial network of poorly perfused, interconnected immature vessels is formed. The previously described process of angiogenesis and subsequent anastomoses occurs in a repetitive fashion using the initial network as parent vessels, therefore producing an extended capillary bed concentrated in the tumor. However, the neovasculature is irregular and poorly perfused in comparison to normal tissue vasculature and will be portrayed as a biological barrier to anticancer treatment efficacy in Chapter 5.

In their seminal study, Anderson and Chaplain developed a tumor-induced angiogenesis model (A. R. Anderson & Chaplain, 1998) with the ability to follow the motion of EC at the capillary tips and control important processes such as proliferation, branching, and anastomosis. Their model utilizes a hybrid approach (*i.e.*, both continuum and discrete modeling) and focuses on three significant variables related to angiogenesis: EC density, and PAP and fibronectin concentrations. While their continuum component tracks the profiles of PAP and fibronectin in the tumor microenvironment, their discrete component tracks the migration of individual endothelial cells sprouting from local capillary beds in response to PAP (chemotaxis) and fibronectin (haptotaxis). This influential study by Anderson and Chaplain enabled simulation of qualitatively realistic structure of tumor vasculature laying the groundwork for improved tumor simulation (Cristini, *et al.*, 2005; Cristini, Li, Lowengrub, & Wise, 2009; H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Bearer, *et al.*, 2007; X. Zheng, Wise, & Cristini, 2005), and serving as inspiration for other angiogenesis models (H. A. Levine, Pamuk, Sleeman, & Nilsen-Hamilton, 2002; Plank & Sleeman, 2003, 2004; Sun, Wheeler, Obeyesekere, & Patrick, 2005).

The abnormal nature of tumor vasculature compared to healthy tissue vasculature has been addressed (Baish, *et al.*, 1996; Jain, 2001b; Pries, *et al.*, 2009). Irregular tumor vasculature leads to restricted and inhomogeneous drug and nutrient extravasation to tumor tissue, which may exacerbate the situation by selecting for highly resistant clones. Anderson and Chaplain's angiogenesis model appears to capture the irregularity of tumor vasculature through appropriate adjustment of the governing mathematical parameters. However, their model and others (H. A. Levine, *et al.*, 2002; Plank & Sleeman, 2003, 2004; Sun, *et al.*, 2005) only describes the physical structure of the capillary network. Nutrient, oxygen, and drug distributions in a tumor can be modeled in a simplified

fashion by using the vasculature produced by these models as a source boundary condition in a diffusion-reaction system, with the transfer rate between the vasculature and tissue as a function of local pressure differences. In reality, nutrient, oxygen, and drug delivery depend on blood flow through the vasculature; using the entire vasculature as a uniform source with blood-tissue transfer proportional to local pressures is a satisfactory interpretation, which can be improved upon by a more direct model of blood flow.

McDougall *et al.* (McDougall, *et al.*, 2002) developed a direct extension of the Anderson and Chaplain angiogenesis model (A. R. Anderson & Chaplain, 1998) by describing the generated vascular networks as a series of straight, rigid cylindrical capillaries that join adjacent nodes. Blood flow is modeled through the cylindrical vascular network by modeling the elemental flow-rate in each segment with Poiseuille's Law, which describes flow-rate as a function of capillary lumen radius, fluid viscosity, capillary length, and pressure drop. Using this simple flow model, McDougall *et al.* (McDougall, *et al.*, 2002) identified tumor neovasculature as a biobarrier to chemotherapy. Results of their simulations indicated that the highly interconnected nature of irregular vasculature produced by tumor-induced angiogenesis could cause low rates of drug delivery to the tumor with the potential for drug to actually completely bypass the entire mass depending on the tumor shape and consequent pro-angiogenic protein distribution. Additionally, the simulation results suggest that drug delivered by bolus injection suffers from severe dilution, therefore reducing drug efficacy.

Stephanou *et al.* (Stephanou, McDougall, Anderson, & Chaplain, 2005) extended the work of McDougall *et al.* (McDougall, *et al.*, 2002) by developing an algorithm that normalizes vasculature produced by Anderson and Chaplain's angiogenesis model (A. R. Anderson & Chaplain, 1998). They examined how pruning vessels by anti-angiogenic

drugs might affect blood flow distribution, and consequently drug delivery to tumors. Their work included blood flow simulations in fully 3D vasculature. Stephanou and coworkers later included vascular adaptation effects (Stephanou, McDougall, Anderson, & Chaplain, 2006), due to shear stress generated by flowing blood (Pries, Reglin, & Secomb, 2001; Pries & Secomb, 2005), to the angiogenesis model to investigate how adaptive remodeling affects oxygen and drug supply to tumors. Alarcon *et al.* (Alarcon, *et al.*, 2003) also modeled the vascular adaptation effects in an effort to study inhomogeneity of oxygen distribution in tumors, and consequential role of hypoxic cells in tumor invasion. More recently, McDougall *et al.* (McDougall, Anderson, & Chaplain, 2006) modified their angiogenesis model to simultaneously couple vessel growth with blood flow to dynamically include the effects of vascular adaptation, rather than adapt the vasculature *a posteriori* like Stephanou *et al.* (Stephanou, *et al.*, 2006) and Alarcon *et al.* (Alarcon, *et al.*, 2003).

Pharmacology and Drug Efficacy

In the event that a cancer has metastasized, systemic treatment is generally necessary in the form of chemotherapy delivered to the primary and secondary tumors through the bloodstream. Drugs must overcome various resistance mechanisms and barriers that affect their efficacy *en route* to their respective targets, thus producing the overall MDR phenomenon. Individual mechanisms and barriers occur at different scales:

1) At the subcellular and cellular scale, there exists a range of drug influx/efflux pumps, changes in the expression of topoisomerases, and alterations in metabolic pathways (*e.g.*, influencing drug metabolism, DNA repair, and/or apoptosis); 2) At the tumor and body scale, resistance can be due to normal clearance mechanisms (*e.g.*, urinary system, reticuloendothelial system, blood-brain barrier), abnormal tumor vasculature, tumor

microenvironment, and tumor three-dimensional structure (Jain, 2001a). Consequently, these biobarriers impede the delivery of chemotherapeutic drugs at effective concentrations to all cancer cells.

Pharmacokinetics

In the study of drug delivery, it is common to conceptualize the organism as a system of interconnected pools called compartments. The investigation of the properties of these compartments and the material fluxes between them is termed “compartment modeling” (Holz & Fahr, 2001). Conventional pharmacokinetic (PK) models use compartment modeling to investigate cellular drug-uptake and intracellular drug interactions as well as provide insight into modeling cellular-scale mechanisms contributing to drug resistance. For example, a standard 3-compartment model was used by Dordal *et al.* (Dordal, *et al.*, 1995) to investigate cellular drug uptake. Their objective was to quantify increased efflux, decreased intracellular sequestration, and decreased membrane permeability as they relate to a reduction in drug effectiveness. Using flow cytometry, they assessed the cellular uptake of doxorubicin and fluorescent rhodamine-123 in drug-resistant and drug-sensitive cancer cells. By fitting the experimental data to the compartmental model, kinetic parameters for both inward and outward transport were obtained and used to quantify the relative importance of the previously mentioned cellular mechanisms. Specifically, their results indicate that of the three cellular mechanisms modeled, decreased intracellular sequestration in a non-exchangeable compartment is quantitatively the most significant contributor towards drug resistance (Dordal, *et al.*, 1995). Similarly, compartment modeling can be applied to investigate additional components affecting drug delivery such as extracellular drug binding and target repair mechanisms.

Pharmacodynamics

While PK describes drug penetration, pharmacodynamics (PD) describes drug cytotoxicity. Although the mechanisms contributing to drug effect are incompletely understood, several phenomenological models adequately yield fractional cell survival S as a function of concentration-time exposure history. The Hill-type model $S = (1 + Ax^m)^{-1}$

is often used, where x is a measure of “cellular damage,” such as extracellular or intracellular area under the curve (AUC). AUC, in turn, is given as the integral of concentration with respect to time $\equiv Cdt$. Another possibility is the exponential kill model $S = \exp(-kx)$, where x again is a suitable measure of damage and k is a constant. While these are perhaps the simplest PD models in use, other equations can be employed.

A study by El-Kareh and Secomb investigated several measures of cellular damage in conjunction with the Hill-type model given above to determine which provided the best fit to experimental data (El-Kareh & Secomb, 2003, 2005). Their investigation was prompted by the observation that models employing extracellular AUC consistently overestimated cytotoxicity in cases of extended exposure to the drugs, cisplatin and doxorubicin. Experimentally, toxicity would achieve a plateau, above which continued exposure, even to continually refreshed drug, would have no effect. To explain this they hypothesized that it was not the time of exposure *per se* that correlated with cytotoxicity, but rather the *peak* level of DNA-bound drug (El-Kareh & Secomb, 2003). Accordingly, they used this measure and showed that for short exposure times, the delay in achieving DNA-bound drug equilibrium could explain increasing cytotoxicity in time. Various experimental cell survival data were fit to determine appropriate values for the constants A and m . The new model was compared to previous models describing the relationship between cytotoxicity and exposure time. El-Kareh and Secomb’s model consistently proved to be the best fit even for long exposure in *in vitro* datasets (Troger, Fischel, Formento, Gioanni, & Milano, 1992) establishing that peak DNA-bound cisplatin is a stronger indicator of cytotoxicity than extracellular or intracellular concentrations. Later, they extended the model to doxorubicin (El-Kareh & Secomb, 2005). Experimental evidence suggests that doxorubicin has two cytotoxic mechanisms, one involving topoisomerase II inhibition by intracellular drug, and the other involving

apoptosis induction *via* extracellular drug. El-Kareh and Secomb proposed a model that combines the effects of both mechanisms into the “cellular damage” by summing peak intracellular and extracellular drug concentrations. Like the cisplatin model, their doxorubicin pharmacodynamics model provides better fits to *in vitro* cytotoxicity datasets than previous models (El-Kareh & Secomb, 2005).

Core Tumor Model Used in Chapter Five

Zheng *et al.* (X. Zheng, *et al.*, 2005) produced a first-generation multidimensional tumor simulator employing a sharp-interface (level-set) finite-element numerical method for tracking the tumor boundary. This model is capable of simulating two-dimensional tumor evolution through the major phases of growth, including avascular dormancy, neo-vascularization and subsequent rapid expansion, and infiltration of host tissue. Wise *et al.* (H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Bearer, *et al.*, 2007; Wise, Lowengrub, Frieboes, & Cristini, 2008) recently produced a second-generation three-dimensional tumor simulator employing a more physically accurate diffuse-interface formulation on a finite-difference framework, which can more realistically represent tissue interfaces and clonogenic heterogeneity. Both simulators not only serve to model tumor progression, but also provide test-beds for therapeutic strategies and hypotheses (Cristini, *et al.*, 2005; J. Sinek, Frieboes, Zheng, & Cristini, 2004).

These models build on the continuum-based approach used by Cristini *et al.* (Cristini, *et al.*, 2003), which considers tumor mass as an incompressible and viscous material that locally expands and contracts in correspondence to variable rates of cell mitosis and apoptosis. Tumor cells, themselves, are not individually represented. In Zheng *et al.*'s formulation (X. Zheng, *et al.*, 2005) local lesion environment is modeled as three sharply demarcated non-intersecting domains: 1) viable tumor, 2) necrotic tissue, and 3) host tissue (see Figure 3.2). Although Wise *et al.*'s diffuse interface formulation

(Wise, *et al.*, 2008) uses a similar partitioning of the lesion environment into tissue domains, boundaries are not as strictly defined. Instead, at a given location, intermixing of several cancerous clones along with necrotic and host tissue can be represented by specifying their relative mass fractions. This cannot be done in the sharp interface model, and is a critical improvement towards realistically simulating mutation-driven heterogeneity.

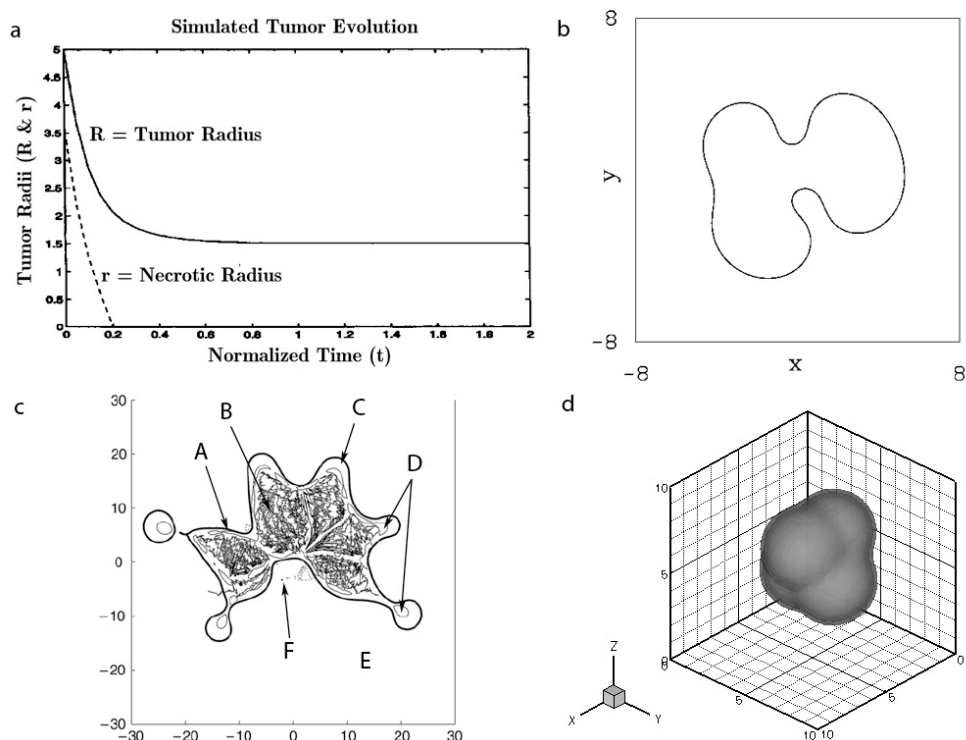


Figure 3.2: Examples of simulated tumor evolution.

(A) Byrne and Chaplain’s necrotic tumor model with a specific apoptosis rate (H. M. Byrne & Chaplin, 1996). This model describes the shrinkage of tumor and necrotic rim. (B) Cristini *et al.*’s nonlinear continuum-based boundary-integral model, which was among the first to capture complex tumor morphologies (Cristini, *et al.*, 2003). (C) Zheng *et al.*’s sharp-interface level-set model of tumoral lesion (X. Zheng, *et al.*, 2005), which captures complex tumor progression including tumor-induced angiogenesis; A \Rightarrow Tumor Boundary, B \Rightarrow Neovasculature, C \Rightarrow Viable Tumor Tissue, D \Rightarrow Necrotic Tumor Tissue, E \Rightarrow Host Tissue, and F \Rightarrow “Free” EC migrating from Parent Vessel (not shown). (D) Wise *et al.*’s diffuse-interface model (Wise, *et al.*, 2008), which captures complex tumor morphology and clonogenic heterogeneity in three-dimensions (angiogenesis not shown).

While it is the growth and regression of lesion tissue that is of primary interest, other processes support and interact with this growth, necessitating a modular design in which simulator components are dedicated to process management. The major components essential to basic lesion simulation are specific to growth and regression, nutrient delivery, and angiogenesis. Beyond these, modules pertaining to molecular-scale events (*e.g.*, gene regulatory networks, protein-protein signaling), cellular-scale behavior (*e.g.*, motility, adhesion, proliferation, apoptosis), and therapy (*e.g.*, surgery, radiation

therapy, chemotherapy) need be added to provide pertinent information for expanding the temporal and spatial range of simulation, and applicability of the technology to patient-specific medicine. The growth and regression component postulates that cell velocity is proportional to pressure gradient (Darcy's law), which is commonly used to model motion through porous substrates (such as a continuum of cells flowing through the extracellular matrix). Morphology, especially as it pertains to invasiveness, is affected by parameters that model cell adhesion, for instance through the definition of an equivalent surface tension at the tumor boundary (H.P. Greenspan, 1972). Based upon inputs from other components, the growth module produces a cell displacement (velocity) field and advects the tumor boundary in the case of Zheng *et al.*'s method (X. Zheng, *et al.*, 2005), and the species mass fractions in the case of Wise *et al.*'s method (Wise, *et al.*, 2008).

Vasculature is incorporated into these simulators as an angiogenesis module inspired by the work of Anderson, Chaplain, Plank, and Sleeman (A. R. Anderson & Chaplain, 1998; Plank & Sleeman, 2004), linked to tumor growth through the release of tumor angiogenic regulators by necrosing tumor tissue. The transition between avascular and vascular tumor growth is marked by the recruitment of microvasculature from local blood vessels (*i.e.*, angiogenesis). The governing processes of angiogenesis are still very much in debate, but one proposed mechanism (A. R. Anderson & Chaplain, 1998; Plank & Sleeman, 2003, 2004) grows new vessels from parent vessels due to chemotaxis of endothelial cells along an angiogenic regulator gradient towards the tumor; endothelial cells also interact with the extracellular matrix in a process known as haptotaxis. In Chapter 7, we develop a general agent-based model for directed cell motility that has applications towards modeling the migration of endothelial cells in response to VEGF gradients.

Underlying these components are sophisticated numerical algorithms, including adaptive computational meshes (A. Anderson, Zheng, & Cristini, 2005; Cristini,

Blawdziewicz, & Lowenberg, 2001) that enable high-resolution rendering of complex tumor morphologies, including fingering, invasion, and reconnection, across multiple length scales at the minimum computational expense.

EMPIRICAL CANCER MODELING APPROACHES: MICROARRAY GENE EXPRESSION ANALYSIS

In order to design sophisticated, targeted treatments for cancer, it is critical to consider the mechanisms underlying the pathophysiology that enables cancer cells to subvert normal physiology and express abnormal phenotypes. Microarray-based gene expression profiling has emerged as a powerful tool for simultaneously monitoring thousands of gene transcripts, and is being used to study a variety of biomedical problems such as cancer at the molecular scale. There are a number of commercially available microarray platforms that measure gene expression using similar, but different methods. Regardless of the platform chosen by the experimenter, the evaluation of gene expression experiments usually involves statistical analysis of the data to reveal co-expressing genes and assess differential expression between phenotypes (often the case in comparative studies such as disease vs. normal or drug-resistant vs. drug-sensitive tissue). These analyses benefit from having a larger number of samples. Thus, there is a strong motivation in the community to be able to adequately combine the increasing number of large scale gene expression studies in order to increase the statistical power of analyses and produce better empirically-driven models.

Normalizing microarray studies performed by different groups of researchers and across different platforms poses a number of statistical difficulties due to the inherent systematic errors. These systematic errors are attributed to differences in microarray platforms, in the amount of RNA used, processing facilities, technicians, protocol, outdated molecular knowledge used to define probes on the microarray chips, etc. Further details regarding statistical issues, variations, and systematic errors inherent of

microarray experiments have been comprehensively reviewed (Nadon & Shoemaker, 2002). Identifying and removing these systematic effects while keeping the biological information intact is a research challenge in cross-study micorarray analysis. Various research groups have proposed different normalization methods (Barnes, Freudenberg, Thompson, Aronow, & Pavlidis, 2005; Bolstad, Irizarry, Astrand, & Speed, 2003; Liang & Wang, 2008; Lim, Wang, Lefebvre, & Califano, 2007; Shabalín, Tjelmeland, Fan, Perou, & Nobel, 2008; Yang & Chang, 2002) in combination with updated probe definitions to achieve this goal (Carter, Eklund, Mecham, Kohane, & Szallasi, 2005; Dai, *et al.*, 2005; H. Liu, *et al.*, 2007). Once the normalization process minimizes systematic bias in the combined data, a meta-analysis (an analysis conducted using combined sets of data) usually proceeds to apply various statistical and data mining techniques to study the data. These unsupervised and supervised methods are expected to produce results with higher statistical power compared to studies performed with less samples (Figure 3.3).

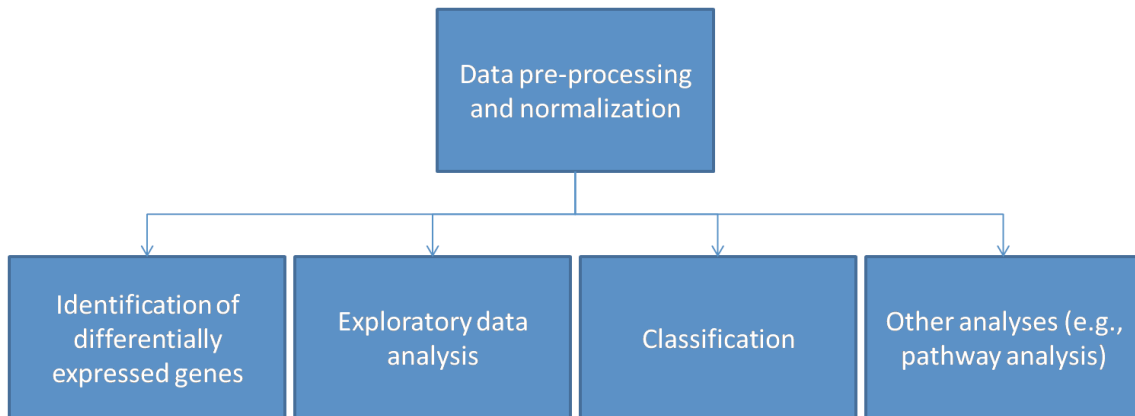


Figure 3.3: Flow of a typical microarray experiment or meta-analysis of microarray studies following acquisition of raw gene expression data, much of which is made publicly available.

Unsupervised Analysis – Exploring Gene Expression Levels to Identify Co-Expressing Genes Within and Across Samples

Unsupervised data analysis methods do not take into account prior knowledge,

and are constructed in a manner that they work with data at hand. Typically, a relative measure of distance between objects is introduced, which enables the researcher to group samples together according to their proximity. The objective of unsupervised gene expression microarray analyses is to find genes and samples that express together similarly; identifying samples sharing a similar gene expression profile can help to classify heterogeneity within cancers having similar organ origins (Belani, 2002; Covell, Wallqvist, Rabow, & Thanki, 2003; Mischel, Cloughesy, & Nelson, 2004; Ogino & Goel, 2008; Perou, *et al.*, 2000; Y. F. Wong, *et al.*, 2003). Moreover, identifying co-expressing genes can help elucidate signaling relationships underlying the tissue in question, especially if the co-expression is common to a number of samples within a group. Some commonly used methods for unsupervised gene expression analysis included hierarchical clustering (Eisen, Spellman, Brown, & Botstein, 1998), K-means clustering (Lu, Lu, Fotouhi, Deng, & Brown, 2004; F. X. Wu, 2008), principal components analysis (PCA) (Atkinson, *et al.*, 2007; Barra, 2004; Peterson, 2003), and self organizing maps (Fernandez & Balzarini, 2007; Herrero & Dopazo, 2002; Nikkila, *et al.*, 2002; Toronen, Kolehmainen, Wong, & Castren, 1999). We direct the reader to reviews for more details regarding these methods (Leung & Cavalieri, 2003; Quackenbush, 2001; Sherlock, 2001; Valafar, 2002).

Supervised Analysis

Supervised data analysis methods take into account prior knowledge, where the samples are labeled according to a differentiating feature (class A vs. class B). These methods take advantage of these labels to build empirical models for prediction or classification of samples not used to train the models, such as prediction of survival in breast cancer patients (van't Veer & Bernards, 2008; van de Vijver, *et al.*, 2002) or response to a specific chemotherapy regimen in bladder cancer patients (Takata, *et al.*,

2005) based on gene expression.

Identifying Differentially Expressed Genes (Comparative Analysis): Discovery of Predictive Gene Signatures and Biomarkers

Comparative analysis of the gene expression between different sample classes (cancer vs. normal, drug-sensitive vs. drug-resistant, good survival vs. poor survival) is a typical approach used for discovering molecular signatures and biomarkers of predictive and prognostic value. A conventional method for identifying differentially expressed genes between sample classes is to define a threshold cut-off for differences in expression levels, such as two-fold increases or decreases. While this method is simple to implement, it is not based on statistical significance of changes. It is sensitive to noise in the data as well as limits analyses by passing on gene expressions with subtle differences.

The standard for comparative microarray analysis is based on statistical calculations using replicate array data. Genes are ranked according to their possibility of expressing differently between multiple sample classes. Replication is achieved by performing microarray experiments on several samples of the same class, for example one tissue type obtained from multiple patients compared to another tissue type also obtained from multiple patients. This is particularly important for expression profiling of disease tissues in order to account for gene expression heterogeneity of the disease type within the patient population, thus any one particular sample may not be representative of the population. Statistical hypothesis tests such as the Student's T-test, ANOVA, Bayesian analysis, or the Mann-Whitney test are used to rank the genes according to p-values (Leung & Cavalieri, 2003), and thus elucidate which genes are differentially expressed between sample classes with statistical significance. Setting a standard for statistical significance must take into account the possibility of identifying false positives and false negatives, therefore setting a conventional p-value cut off of 0.05 may not be appropriate for microarray experiments capable of generating gene expression data for

the full human genome (~25,000 genes). Instead, a multiple hypothesis correction, such as the Bonferroni correction, can help determine an appropriate cut-off p-value based on the number of hypotheses (the number of genes being compared in the case of microarray gene expression analysis). An alternative method, and has become a standard in differential gene expression analysis, is implemented in the software package called Statistical Analysis of Microarrays (SAM) (Tusher, Tibshirani, & Chu, 2001); a modified Student's T-test is used to rank the statistical significance of differential gene expression between sample classes, and also performs permutations on the data to derive a false discovery rate (FDR), or the expected frequency of identifying false positives. This feature allows researchers to use FDR as a cut-off for determining differentially-expressed genes, which can then be validated experimentally using quantitative real-time polymerase chain reaction (Q-RT-PCR) or northern blots (Leung & Cavalieri, 2003).

Classifier Prediction Methods

Another form of supervised gene expression analysis is to train a classifier algorithm on a pre-defined sample groups, and use the classifier to assign new samples. Gene signatures identified by differential gene expression analysis can also serve this purpose. This type of analysis is taking advantage of data mining and machine learning methods that are receiving attention in many other disciplines including the use of decision tree methods (Edgerton, *et al.*, 2007; Pirooznia, Yang, Yang, & Deng, 2008; A. C. Tan & Gilbert, 2003), artificial neural networks (Bicciato, Pandin, Didone, & Di Bello, 2003; J. Huang, Shimizu, & Shioya, 2003; Petalidis, *et al.*, 2008), and support vector machines (Y. Liu, 2004; Statnikov, Aliferis, Tsamardinos, Hardin, & Levy, 2005). The challenge is to avoid “over-fitting” that model, which is where the model is accurate in predicting the data it was trained upon, but not in predicting new samples. Developing these types of predictive models usually follows a protocol of defining training and testing sets, where the testing sets are used to assess the prediction error of the trained

model. When the number of samples is limited, as is usually the case with gene expression microarray data (thus the motivation to combine microarray studies), cross-validation methods are used to permute the data to create multiple sets of training and test sets (Goutte, 1997; Zhu & Rohwer, 1996). These empirically-driven models are expected to make an impact in the future of clinical diagnostics and can also help reveal the molecular mechanisms underlying diseased phenotypes.

MODELING INTRACELLULAR PATHWAYS DRIVING TUMOR PHENOTYPES

The behavior of cells is governed by the state of their internal signals, much like the behavior of an airplane or motor vehicle is dependent upon the state of its parts and how they are interacting with each other. We are entering a new age of cancer medicine, where the sophisticated therapeutic approaches must consider the molecular signaling networks that cause cancer cells to behave abnormally. For this reason, reconstruction and analysis of the molecular determinants driving normal and aberrant cellular phenotypes is an active field of biological research, and has implications for drug target discovery. In the framework of a multi-scale cancer model, a description of molecular scale interactions and how they affect cellular behaviors (*e.g.*, cell motility, cell-cell interactions, cell-matrix interactions, drug sensitivity) is a critical link to develop.

Cell signaling networks are complex in their organization because individual elements are highly interconnected and tend to regulate a multitude of cell functions. Because of the size and complexity of the number of interacting elements in pathways regulating cell behavior, it has become necessary to develop mathematical models to help elucidate the system behavior of signaling networks in order to predict emergent properties that can be validated through experimentation. Modeling signaling networks is expected to augment our understanding of pathways by 1) yielding insight into how individual pathways function within the context of other pathways, 2) enabling the

estimation of biochemical parameters that are not yet possible to probe experimentally, and 3) simulating system dynamics to help reveal how and why cells behave the way that they do, thus exposing potential strategies to treat diseases such as cancer (Bhalla & Iyengar, 1999; Eungdamrong & Iyengar, 2004b; Neves & Iyengar, 2002; Ruths, Nakhleh, Iyengar, Reddy, & Ram, 2006).

In developing models of molecular signaling pathways, the first step is to define the system architecture, or the “connection map”, that describe the relevant molecular determinants and how they interact with one another. Because cellular regulation is achieved through the dynamic interaction between the many components identified in the “connection map,” the subsequent step is to apply a modeling framework, typically using mass-action kinetics (W. W. Chen, *et al.*, 2009; A. E. DeWitt, Dong, Wiley, & Lauffenburger, 2001; Kholodenko, Demin, Moehren, & Hoek, 1999; Wiley, Shvartsman, & Lauffenburger, 2003), compartmental modeling (Holz & Fahr, 2001), or diffusion-reaction equations (Rangamani & Iyengar, 2007), to simulate pathway dynamics, where some (if not all) of the system parameters can be constrained through experiments or rational estimation based on evidence in the literature. In Chapter 6, we present a research study where we use bioinformatics and data mining techniques to postulate signaling relationships underlying a breast cancer subset, and in Chapter 7, we develop a prototype agent-based model for predicting cellular scale phenotypes from underlying signaling, and to serve as a link between the molecular scale and cellular scale in our research laboratory’s multi-scale cancer modeling framework.

Reconstruction of Signaling Maps

Reconstruction of intracellular signaling networks enables a molecular systems-level understanding of network function that is critical for studying complex diseases like cancer (Eungdamrong & Iyengar, 2004b). A reconstructed network is inferred from experimental data and gives an accurate representation of the biochemical events occurring within a signaling network. This signaling connectivity map provides the architecture for applying mathematical methods to quantitatively analyze the signaling network properties (Papin, Hunter, Palsson, & Subramaniam, 2005), and is already helping link the molecular and cellular scales in multi-scale agent-based cancer models (C. Athale, Mansury, & Deisboeck, 2005; Deisboeck, Zhang, Yoon, & Costa, 2009; Kharait, *et al.*, 2007; Z. Wang, Zhang, Sagotsky, & Deisboeck, 2007; Zhang, Athale, & Deisboeck, 2007; Zhang, *et al.*, 2009).

The advance of gene expression microarray technology, generation of large-scale sets of data, and improvements in normalization methods and bioinformatics analysis algorithms has lead to gene expression profile data serving as a major data source for reverse engineering cellular networks because signaling events such as protein-protein interactions and protein-DNA interactions can be inferred from the statistical dependencies between gene expression profiles (Basso, *et al.*, 2005; Lim, *et al.*, 2007). The primary drawback for network reconstruction based on gene expression profiles is the lack of enough samples to provide and inherent noise in the data. This serves as further motivation to combine microarray studies and improve normalization protocols that minimize persisting systematic noise in the data (see above for discussion of these topics). Various techniques applied to gene expression data in attempts to extract underlying signaling knowledge, such as Boolean networks (Akutsu, Miyano, & Kuhara, 1999; Yu, Watterson, Marshall, & Ghazal, 2008), differential equations (Finkenstadt, *et*

al., 2008), Bayesian networks (Broom, Rinsurongkawong, Puztai, & Do, in preparation; Markowitz & Spang, 2007), and supervised machine learning approaches (Edgerton, *et al.*, 2007; Fisher D, Edgerton M, Tang L, Frey L, & Chen Z, 2005; Fisher DH, Edgerton ME, Chen Z, Tang L, & Frey LJ, 2006; Frey, Edgerton, Fisher, Tang, & Chen, 2005; Soinov, 2003; Soinov, Krestyaninova, & Brazma, 2003). In experiments where gene expression is profiled in time-series, these methods can be used to infer input-output relationships, or causality. Static data, on the other hand, like much of the gene expression publicly available on human cancers can be used to postulate interactions that can later be tested experimentally. Resolving gene network interactions from static gene expression data is well-suited for machine learning techniques and Bayesian network methods, thus we use them as part of the approach we take in Chapter 6 to infer signaling relationships underlying a breast cancer subset.

Chapter 4: Predictive Oncology: A Review of Multi-disciplinary, Multi-scale *In Silico* Modeling Linking Phenotype, Morphology and Growth

PREFACE

The objective of *predictive oncology* is to promote cancer prognosis and treatment by advancing our understanding of the molecular biology underlying cancer phenotypes, prognostic evaluation of tumor development and progression, and predicting tumor response to therapy on a patient-specific basis. This chapter is based on a published article, which was a collaborative effort between Sandeep Sanga, Dr. Hermann B. Frieboes, Dr. Xiaoming Zheng, Dr. Robert Gatenby, Dr. Elaine L. Bearer, and Dr. Vittorio Cristini. Sanga took the lead role in preparing the manuscript. Frieboes and Zheng assisted in the manuscript's preparation. The multi-scale modeling framework presented in this manuscript is the core philosophy of our research group under the supervision of Cristini with collaborators Gatenby and Bearer, as well as others, and is a unifying theme of the research presented in this dissertation.

Note: This chapter is based on an article originally published as (Sanga, Frieboes, Zheng, <i>et al.</i> , 2007) in <i>NeuroImage</i> , which can be accessed at http://dx.doi.org/10.1016/j.neuroimage.2007.05.043 . This article has been included in this dissertation with permission from Elsevier.

ABSTRACT

Empirical evidence and theoretical studies suggest that the phenotype, *i.e.*, cellular- and molecular-scale dynamics, including proliferation rate and adhesiveness due to microenvironmental factors and gene expression that govern tumor growth and invasiveness, also determine gross tumor-scale morphology. It has been difficult to quantify the relative effect of these links on disease progression and prognosis using conventional clinical and experimental methods and observables. As a result, successful individualized treatment of highly malignant and invasive cancers, such as glioblastoma,

via surgical resection and chemotherapy cannot be offered and outcomes are generally poor. What is needed is a deterministic, quantifiable method to enable understanding of the connections between phenotype and tumor morphology. Here in this chapter, we critically assess advantages and disadvantages of recent computational modeling efforts (*e.g.*, continuum, discrete, and cellular automata models) that have pursued this understanding. Based on this assessment, we review a multi-scale, *i.e.*, from the molecular to the gross tumor scale, mathematical and computational “first-principle” approach based on mass conservation and other physical laws, such as employed in reaction-diffusion systems. Model variables describe known characteristics of tumor behavior, and parameters and functional relationships across scales are informed from *in vitro*, *in vivo* and *ex vivo* biology. We review the feasibility of this methodology that, once coupled to tumor imaging and tumor biopsy or cell culture data, should enable prediction of tumor growth and therapy outcome through quantification of the relation between the underlying dynamics and morphological characteristics. In particular, morphologic stability analysis of this mathematical model reveals that tumor cell patterning at the tumor-host interface is regulated by cell proliferation, adhesion and other phenotypic characteristics: histopathology information of tumor boundary can be inputted to the mathematical model and used as a phenotype-diagnostic tool to predict collective and individual tumor cell invasion of surrounding tissue. This approach further provides a means to deterministically test effects of novel and hypothetical therapy strategies on tumor behavior.

THE ROLE OF PREDICTIVE SCIENTIFIC COMPUTATION AS “*IN SILICO*” CANCER MODELING

Cancer Progression and Invasion: Current Understanding

A wealth of empirical evidence links disease progression with tumor morphology, invasion, and associated molecular phenomena. However, not only is there a lack of quantitative understanding of the underlying physiological processes driving tumor-scale behavior, in particular, morphology at the tumor-host interface, but the qualitative explanations themselves may be indecisive or inconsistent. For example, a positive correlation of cell adhesion molecules (integrins) and cancer cell migration was observed in glioma cells (Tysnes, *et al.*, 1996), yet integrins can also serve as negative effectors that impede invasion and progression (Zutter, Santoro, Staatz, & Tsung, 1995). Similarly, conflicting data on the function of proteases in tumor invasion and metastasis (Friedl & Wolf, 2003) is illustrated by variable results from clinical trials of potential anti-invasive therapies (Lah, Alonso, & Van Noorden, 2006). While the primary role of angiogenesis in promoting tumor growth and invasion has been well demonstrated, the results of clinical trials using various drugs to suppress neovascularization have yielded mixed results. Despite encouraging signs of tumor regression following anti-angiogenic therapy, in some cases length of survival remains the same (Bernsen & Van der Kogel, 1999; Bloemendal, Logtenberg, & Voest, 1999; Kuiper, Schellens, Blijham, Beijnen, & Voest, 1998). In addition, experimental observations indicate that anti-angiogenic treatments may exacerbate hypoxia (Steeg, 2003a), and paradoxically promote tumor fragmentation, cancer cell migration, and host tissue invasion (Bello, *et al.*, 2004; Kunkel, *et al.*, 2001b; Lamszus, Kunkel, & Westphal, 2003; Page, Anderson, & Sakamoto, 1987; Seftor, *et al.*, 2002).

Links Between Cellular- and Tumor-scale

In spite of abundant experimental and clinical data surrounding molecular and cellular phenomena, it is difficult to quantify their aggregate effect on gross tumor-scale

behavior using conventional methods that, for the most part, investigate isolated mechanisms. Prognosis of cancer patients suffering from highly invasive tumors, such as glioblastoma, is grim despite advances in surgical and chemotherapeutic treatment, since not all tumor cells can be removed or treated because of limited delineation between healthy and tumor tissue at the tumor border, which may lead to fatal recurrences (A. R. Kansal, S. Torquato, I. G. Harsh, E. A. Chiocca, & T. S. Deisboeck, 2000b). In particular, mechanisms governing glioma invasion likely include intrinsic properties of cell proliferation, migration, and adhesion. Glioma cells have been experimentally shown to infiltrate and scatter throughout the entire central nervous system after a period of only seven days post-implantation (Chicoine & Silbergeld, 1995; Silbergeld & Chicoine, 1997; Swanson, Alvord, & Murray, 2000). This might be one reason why current treatments that focus on surgery, radiation, and chemotherapy, while perhaps having an effect on primary bulk mass characteristics, may fail to extend survival time.

A Novel, *In Silico* Approach to Cancer Modeling

In this chapter we describe a multidisciplinary method integrating mathematical models with experimental (*in vitro* and *in vivo*) and clinical data. This methodology reflects an “engineering” approach that views tumor lesions as complex micro-structured materials, where three-dimensional tissue architecture (“morphology”) and dynamics are coupled in intricate, complex ways to cell phenotype, which in turn is influenced by factors in the microenvironment. Cellular and microenvironmental factors act both as tumor morphology regulators and as determinants of invasion potential by controlling the mechanisms of cancer cell proliferation and migration (Friedl & Wolf, 2003; Sierra, 2005; van Kempen, Ruiters, van Muijen, & Coussens, 2003). In particular, recent experimental results demonstrate that interactions between cellular proliferation, adhesion, and other phenotypic properties are reflected in both tumor-host interface

morphology and invasive characteristics of tumors (Bello, *et al.*, 2004; Bernsen & Van der Kogel, 1999; Bloemendal, *et al.*, 1999; Friedl & Wolf, 2003; Kuiper, *et al.*, 1998; Kunkel, *et al.*, 2001b; Lah, *et al.*, 2006; Lamszus, *et al.*, 2003; Page, *et al.*, 1987; Seftor, *et al.*, 2002; Steeg, 2003a; Tysnes, *et al.*, 1996; Zutter, *et al.*, 1995). The goal is then to create computational (*in silico*) multi-scale tools capable of predicting the complexity of cancer at multiple temporal and spatial resolutions, with the aim of supplementing diagnosis and treatment by helping plan more focused and effective therapy *via* surgical resection, standard chemotherapy, novel treatments (*e.g.*, angiogenic, anti-invasive), or some combination of them. The tools would quantitatively examine the effect of tumor morphology regulators, which include tissue rigidity, density, adhesiveness, microenvironment gradients (*e.g.*, oxygen, nutrient, growth factors), and the combinatorial effects of oncogenes (controlling cell proliferation, motility, and nutrient consumption) and tumor suppressor genes (controlling cell apoptosis and motility) on gross morphology. They would also define the degree of diffuse invasion of tumor cells peripheral to the tumor mass that may be beyond the detection of current non-invasive medical imaging techniques (Swanson, *et al.*, 2000), or extrapolate tumor invasiveness and metastatic potential from its morphology in fixed tissue. *In silico* model development is built upon continuum, discrete, and in particular cellular automata models (Adam, 1996; Araujo & McElwain, 2004; Ayati, Webb, & Anderson, 2006; Bellomo & Preziosi, 2000; Boushaba, Levine, & Nilsen-Hamilton, 2006; Bru, Albertos, Subiza, Garcia-Asenjo, & Bru, 2003; H. M. Byrne, Alarcon, Owen, Webb, & Maini, 2006; Castro, Molina-Paris, & Deisboeck, 2005; M. A. Chaplain & Anderson, 2003; M. A. J. Chaplain & Lolas, 2005; H.B. Frieboes, Sinek, Nalcioglu, Fruehauf, & Cristini, 2006; A. Friedman, 2004; Garner, *et al.*, 2005; Hatzikirou, Deutsch, Schaller, Simon, & Swanson, 2005; Hoge, Murray, & Sethian, 2006; Jackson, 2004; Khain & Sander, 2006; Khain, Sander, & Stein, 2005; Macklin & Lowengrub, 2007a, 2007b; Nagy, 2005; Painter,

Maini, & Othmer, 2000; Quaranta, Weaver, Cummings, & Anderson, 2005; Sander & Deisboeck, 2002; Sanga, Frieboes, Sinek, & Cristini, 2007; Sanga, *et al.*, 2006; Stein, Demuth, Mobley, Berens, & Sander, 2007; Swanson, Bridge, Murray, & Alvord, 2003).

Incorporation of Patient Data: Predictive Modeling

Multi-scale modeling quantifies the time- and space-dependent physics and chemistry (*e.g.*, diffusion of substrates, mechanical forces exchanged among cells and with the matrix, molecular transport, receptor-ligand interactions, pharmacokinetics determinants) underlying tumor biological behavior. We envision that future simulators, building on the developments described herein, will operate as described in Figure 4.1. Initial conditions regarding tumor physical location, structure, and vasculature are obtained from patient data, such as contrast-enhanced MRI (magnetic resonance imaging) (a full discussion of these technologies is beyond the scope of this paper; please see (Clatz, *et al.*, 2004; Clatz, *et al.*, 2005)), possibly coupled with CT (computed tomography) (Roberts, Roberts, Lee, & Dillon, 2002; Xie, *et al.*, 2004). Image information would be translated using a computer program to the coordinate system of the multi-scale model, *e.g.*, a finite-element computational mesh discretizing the space occupied by a tumor and surrounding host tissue (A. R. A. Anderson, 2005; Cristini, *et al.*, 2001; Y. C. Tan, Fisher, Lee, Cristini, & Lee, 2004; X. Zheng, *et al.*, 2005; X. M. Zheng, Lowengrub, Anderson, & Cristini, 2005). Cell-scale model input parameters are obtained from histopathology, including information about viable regions and cell density therein, necrosis, microvasculature, and other details (Bearer, *et al.*, 2009; H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Elaine, *et al.*, 2007). The computer model then calculates local tumor growth, angiogenesis, and response to treatment under various conditions (H. Frieboes, Wise, Lowengrub, & Cristini, in press; H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Elaine, *et al.*, 2007; Wise, *et al.*, 2008; X. Zheng, *et*

al., 2005) by solving in time and space conservation (*e.g.*, diffusion equation) and other laws at the tissue scale. These laws are linked to the cell molecular biology by functional relationships and parameters informed by the biopsy data. Computational solutions are obtained using finite elements and other numerical techniques (*e.g.*, (H. Frieboes, *et al.*, in press; H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Elaine, *et al.*, 2007; X. Zheng, *et al.*, 2005)). Additional patient data obtained from tissue culture, gene arrays, proteomic profiling, and other means would sharpen these parameter estimations (H. B. Frieboes, *et al.*, 2006) in order to enable accurate prediction of behavior. Further details on the parameter estimation procedure are described below and in references cited.

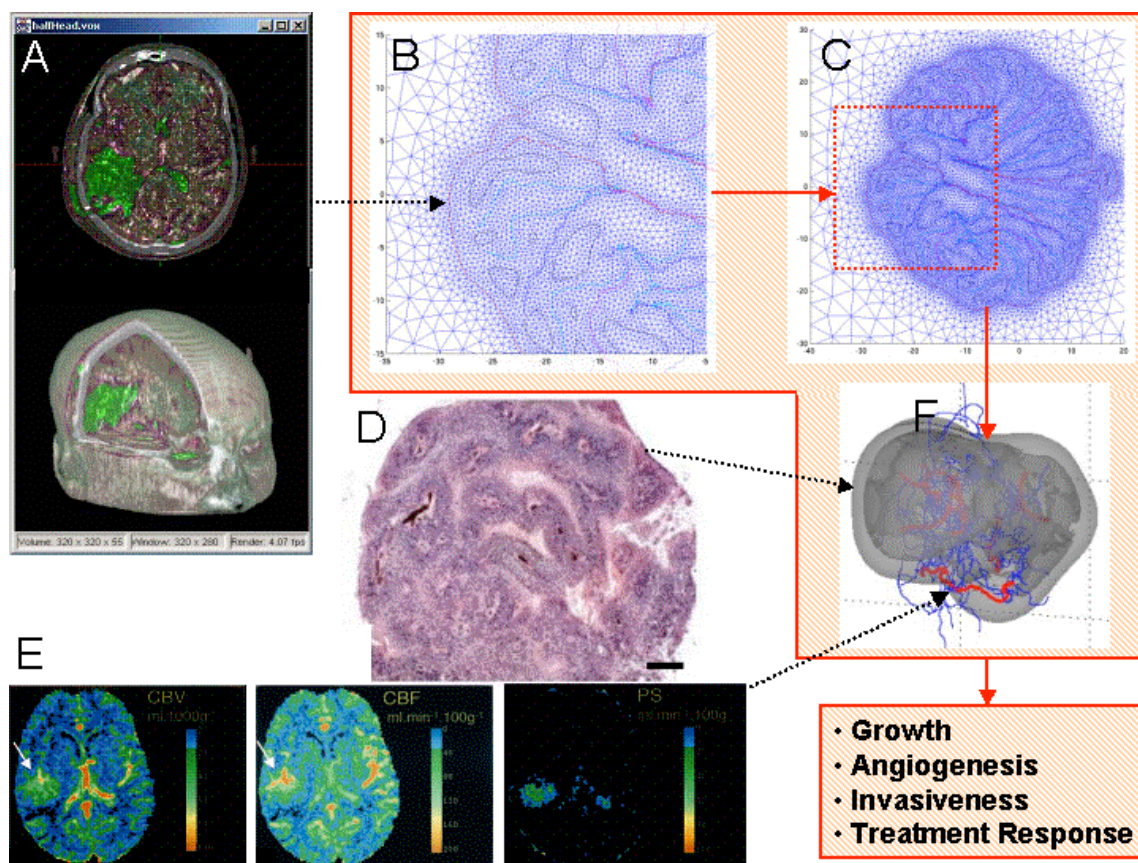


Figure 4.1: Patient data can be inputted into an in silico model of cancer.

The red line encloses sample *in silico* representations of tumor morphology. Tumor morphology at a given time is inputted from 3-D voxel fusion of CT and MR data (A) (reprinted with permission from Xie *et al.* (c) 2004 IEEE) that is translated (using a computer program) into the *in silico* coordinate system (B) (in this example, an unstructured adaptive mesh by Cristini *et al.*, 2001) to build a tumor representation in virtual computational space (C). Spatial information of viable cell regions and microvasculature structure is obtained from histopathology (D) (reprinted from *NeuroImage*, Frieboes *et al.* (2007), Copyright 2007, with permission from Elsevier; bar, 200 μ m). Vasculature specific information is defined from dynamic contrast enhanced CT (E), yielding parameters such as blood volume (left), blood flow (middle), and microvascular permeability (right) (reprinted with permission from Roberts *et al.*, 2002). Other input data to the model include cell-scale parameters such as proliferation rates obtained, for example, from *in vitro* cultures. The model then predicts (F) tumor growth (viable cells: light gray; necrosis: darker gray), angiogenesis (red: mature; blue: immature capillaries), invasiveness, and response to treatment for this patient by solving time- and space-dependent conservation of mass and momentum and other physical laws (reprinted from *NeuroImage*, Frieboes *et al.* (2007), Copyright 2007, with permission from Elsevier). These laws contain parameters that characterize the phenotypes and are linked to the underlying molecular biology by functional mathematical relationships founded on these and other experimental data. Incorporation of models of this biology (*e.g.*, evolution of genotype, cell signaling pathways) guarantees that the *in silico* predictions of tumor behavior are realistic and accurate. Computational solutions of this multi-scale model are obtained using finite elements and other numerical techniques.

MULTI-SCALE MODELING AND SIMULATION OF TUMOR MORPHOLOGY AND INVASION

Model Development Goals and Choices

The progression of tumor lesions cannot be completely characterized by studying effects in isolated cells since it is known that the forces and mechanisms regulating the motion of individual cells converge and synchronize into the collective, organized, structural motion of a whole body or cluster of cells (“Functional Collective Cell-Migration Units,” FCCMU) that often precedes the onset of epithelial-mesenchymal and other phenotypic transitions leading to individual cell shedding from a tumor and eventually to metastasis (Friedl & Wolf, 2003). Both individual and collective migration modes are regulated (and complicated) by multifaceted interactions among tumor cells, stroma and tumor microenvironment (Sierra, 2005; van Kempen, *et al.*, 2003).

While “discrete” *in silico* models (A. R. A. Anderson, 2005; Dickinson & Tranquillo, 1993; Dimilla, Barbee, & Lauffenburger, 1991; dos Reis, Mombach, Walter, & de Avila, 2003; Ferreira, Martins, & Vilela, 2002; A. R. Kansal, S. Torquato, G. I. Harsh, E. A. Chiocca, & T. S. Deisboeck, 2000a; Kansal, *et al.*, 2000b; Leyrat, Duperray, & Verdier, 2003; Patel, Gawlinski, Lemieux, & Gatenby, 2001; S. Turner & Sherratt, 2002) are able to capture individual cell migration and easily incorporate biological rules, such as cell-cell & cell-medium interactions and motion due to chemotaxis and haptotaxis, they are limited to relatively small numbers of cells due to computational cost, among the other deficiencies and over-simplifications introduced by the discrete approach. In contrast, “continuum” models (Bellomo & Preziosi, 2000; H. Byrne & Preziosi, 2003; H. M. Byrne, *et al.*, 2006; H. M. Byrne & Chaplain, 1995; H. M. Byrne & Chaplain, 1996, 1997; H. M. Byrne & Chaplin, 1996; Cristini, *et al.*, 2005; Cristini, *et al.*, 2003; H. B. Frieboes, *et al.*, 2006; X. R. Li, Cristini, Nie, & Lowengrub, 2007; Macklin

& Lowengrub, 2005, 2007b), describing tissue matter as a continuum medium rather than discrete individual cells, capture the collective motion of FCCMUs with less computational expense.

The fact that collective migration is often associated with relatively higher degrees of cell differentiation (Friedl & Wolf, 2003) than for the case of single-cell migration suggests that molecular mechanisms are relatively more robust across a tumor cell population. Thus, the multitude of cells can be averaged out and re-described as a single multi-cellular FCCMU unit obeying deterministic dynamics laws, while still employing mathematical models of single-cell migration when needed, *e.g.*, to describe epithelial-mesenchymal transitions (Friedl & Wolf, 2003). Moreover, the domain size of realistic discrete simulations is limited to sub-millimeter-size *in vitro* tumor spheroids or *in vivo* patches of tumor tissue. We propose that discrete models of cell proliferation and migration should be coupled to continuum models of FCCMU to extend the computational capability to realistic, cm-size three-dimensional tumor lesions as defined and described in the following. A *hybrid*, multi-scale modeling methodology (Bearer, *et al.*, 2009) that links continuum (*i.e.*, tissue-scale) with discrete (*i.e.*, cellular-scale) formulations with appropriate functional relationships of cell adhesion and migration due to environmental conditions should provide, over the next decade, a more comprehensive understanding of the molecular basis of diversity and adaptation of cell migration, thus more efficiently and accurately predicting invasion potential from real-time tumor morphology.

This approach has the advantage that well-established engineering methods and analyses of morphology can be applied (*e.g.*, based on continuum methods when possible). Experimental measurements, computer simulations and morphologic stability analyses can be used to study, in detail, microenvironment transport processes (*e.g.*, of

oxygen, nutrients, chemokines, growth factors), cell motion and proliferation, signaling pathways and molecular phenomena regulating cell cycling, cell-cell communications and expression of cell adhesion molecules and matrix degrading enzymes. For example, the link between hypoxic gradients and invasion, and between normoxic conditions and compact non-infiltrative tumor morphologies, can thus be explained “by exploiting the ability of mathematics to model physical and biological systems in ways that enable prediction and control” (quoting John Lowengrub, Chair, Mathematics, UC Irvine).

Significance of *In Silico* Modeling: A Novel Hypothesis-Generation Tool

The computational models reviewed in this paper represent important steps in generating hypotheses that postulate functional relationships linking the effect of molecular/cellular changes to tumor-scale morphology and invasiveness. By directly solving the mathematical equations describing underlying physical and chemical processes occurring within tumors, the complex biology of tumor behavior and the often hidden mechanisms of growth and invasion automatically are unveiled and can be accurately quantified in virtual, *in silico* simulation space. Examples of novel hypotheses generated from simulations studies and tested in experiments will be provided in the following. Although these types of models are not multi-scale *per se*, parameters characterizing cell response to substrate concentration can be interpreted as representing underlying biochemistry and molecular biology driving tumor-scale dynamics, specifically an invasive phenotype. However, modeling of tumor behavior and cell microenvironment remains a challenge. Existing mathematical models are only capable, in general, of recapitulating *a posteriori* the highly variable empirical observations of morphology, once appropriate *phenomenological* parameters that do not incorporate direct molecular-scale description have been “fitted” to the experiments.

Here we propose that the next decade of investigation should focus on the task of developing *predictive* multi-scale models (e.g., see Figure 4.2) that incorporate new, functional relationships among macro-scale parameters characterizing differences in, and transitions among, cellular patterns, and variations in the molecular repertoire used by tumor cells to regulate proliferation, adhesion and other phenotypic properties.

Vasculature		
Flow & Diffusion <ul style="list-style-type: none"> • Cell substrate transport • Drug delivery • Nanovector passive transport 	Angiogenesis <ul style="list-style-type: none"> • Distribution of angiogenic regulators • Endothelial cell density • Fibronectin density • Vasculature morphology 	Experimental Validation <ul style="list-style-type: none"> • Tumor imaging • <i>In vivo/in vitro</i> experiments

Conservation of mass: Cell substrates

$$0 = D_n \nabla^2 \frac{n}{n_v} + R_i$$

$$R_i = v \left(1 - \frac{n}{n_v} \right) \left(1 - \frac{p}{p_v} \right)^+ \delta - \eta_i \rho_i \frac{n}{n_v}$$

Conservation of mass: Angiogenesis

$$\frac{\partial e}{\partial t} = D_e \nabla^2 e - \nabla \cdot (\alpha_c e \nabla c) - \nabla \cdot (\alpha_f e \nabla f)$$

$$\frac{\partial f}{\partial t} = v_f e - \eta_f e f$$

$$\frac{\partial c}{\partial t} = -\eta_c e c$$

D_n : diffusion coefficient for cell substrates
 n : local concentration of cell substrates
 n_v : concentration inside the vasculature
 R_i : net local rate of cell substrate delivery from vasculature and uptake by tumor cell species i
 v : transfer coefficient from vasculature
 p : pressure in tumor
 p_v : pressure inside vasculature
 δ : indicator function of vasculature
 η_i : rate of cell substrate uptake by species i
 ρ_i : local mass fraction of species i
 e : local density of endothelial cells
 D_e : diffusivity of endothelial cells
 α_c : strength of chemotaxis
 c : local concentration of tumor angiogenic factors
 α_f : strength of haptotaxis
 f : local concentration of fibronectin
 v_f : rate of production of fibronectin
 η_f : rate of uptake of fibronectin
 η_c : rate of uptake of tumor angiogenic factors

Tumor			
Surrounding Tissue	Tumor Scale	Cellular Scale	Experimental Validation
<ul style="list-style-type: none"> • ECM degradation • Vasculature changes • Immune response 	<ul style="list-style-type: none"> • Morphology • Material properties • Oxygen/Nutrient 	<ul style="list-style-type: none"> • Cell contact • Cell adhesion • Cell cycle • Cell substrates 	<ul style="list-style-type: none"> • <i>In vivo</i> tumor models • <i>In vitro</i> tumor models • Clinical data • <i>In vitro</i> cell culture

Conservation of mass: Tumor cells

$$\frac{\partial \rho_i}{\partial t} + \nabla \cdot (\mathbf{u} \rho_i) = \nabla \cdot (M_{i,j} \nabla \frac{\delta E}{\delta \rho_i}) + S_i$$

$$S_1 = \lambda_{M,1} \frac{n}{n_v} \rho_1 - \lambda_{A,1} \rho_1 - \lambda_{N,1} \mathcal{H} \left(\frac{n_{N,1}}{n_v} - \frac{n}{n_v} \right) \rho_1 - \lambda_{TR} \rho_1$$

$$S_2 = \lambda_{M,2} \frac{n}{n_v} \rho_2 - \lambda_{A,2} \rho_2 - \lambda_{N,2} \mathcal{H} \left(\frac{n_{N,2}}{n_v} - \frac{n}{n_v} \right) \rho_2 + \lambda_{TR} \rho_1$$

Conservation of momentum: Velocity

$$\mathbf{u} = -\mu \nabla p + \chi \nabla n / n_v$$

Pressure:

$$[p]_{\Sigma} = \gamma_i \kappa$$

- ρ_i : local mass fraction of tumor cell species i
- \mathbf{u} : tumor cells' velocity
- $M_{i,j}$: tumor cell mobility function
- E : cell adhesion energy
- S_i : mass source term for species i
- $\lambda_{M,i}$: proliferation rate of species i
- $\lambda_{A,i}$: apoptosis rate of species i
- $\lambda_{N,i}$: proliferation rate of species i
- \mathcal{H} : Heaviside function
- $n_{N,i}$: nutrient limit for cell viability of species i
- λ_{TR} : mutation rate (function of position and time)
- μ : hydraulic conductivity of tissue
- χ : velocity of chemotaxis
- Σ : tumor-host interface
- γ_i : tumor surface tension due to species i
- κ : local curvature of tumor-host interface

Genotype		
Genetic Information	Molecular Scale	Experimental Validation
<ul style="list-style-type: none"> • Arrays of genes favoring tumorigenesis: proto-oncogenes, tumor suppressor genes, and mutator genes 	<ul style="list-style-type: none"> • Cell signaling pathways 	<ul style="list-style-type: none"> • Protein/Gene-chip array of tumor biopsy
<ul style="list-style-type: none"> • Multiple cell species • Transition between species regulated by gene mutations according to a mutation rate 	Genetic Scale	
	<ul style="list-style-type: none"> • Combination of genetic mutations 	

Genetic Information:

$$\mathcal{M}[\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3]$$

- \mathcal{M}_1 : Set of tumor suppressor genes
- \mathcal{M}_2 : Set of oncogenes
- \mathcal{M}_3 : Set of mutator genes

Figure 4.2: Main component modules of a sample predictive model based on first principles (e.g., conservation laws of mass and momentum) linking cellular/molecular scale processes to tumor growth and morphology, assuming a tumor with two clones for simplicity.

Each component (Vasculature, Tumor, Genotype) lists biological processes that are implemented through a set of equations (e.g., the diffusion-reaction equation determining the local concentration n of a cell substrate within the tumor), as well as suggested experiments for validation of these functional relationships. Additional biological processes, clones, and properties of the stroma can be incorporated by augmenting the number of variables and equations. Functional relationships to be determined experimentally describe mathematically the dependence of the listed phenotypic parameters (e.g., cell proliferation rates λ_M) on the array M that contains genetic information. Note that several of these parameters are phenomenological, i.e., they do not correspond to direct measurables (e.g., the “strengths” α of haptotaxis and chemotaxis, which are related to the expression of integrins and the mechanical forces exchanged with molecules in the ECM). Data obtained from in vitro experiments, in vivo / ex vivo models,

and clinically (*e.g.*, tumor size, morphology) is thus input to the various modules. This data will also help refine the model's functional relationships through an iterative exercise of multidisciplinary research that will progressively lead, over the next decade, to less phenomenological and more "exact" models, in which the parameters are directly measurable in experiments. For further mathematical details and definitions of variables and parameters, and for demonstrations of a "prototype" *in silico* model see Frieboes *et al.* (2009) and references therein.

This methodology is expected to improve current modeling efforts because a multi-scale approach connects previous work focused on specific scales (*e.g.*, molecular) and processes (*e.g.*, gene transformation), affording the possibility to go beyond the current reductionist picture of tumor invasion and migration (J. Condeelis, *et al.*, 2005; Elvin & Garner, 2005; Friedl, 2004; Friedl, Hegerfeldt, & Tilisch, 2004; Friedl & Wolf, 2003; Jones, Byrne, Gibson, & Dold, 2000; Keller, Pampaloni, & Stelzer, 2006; Kopfstein & Christofori, 2006; Ridley, *et al.*, 2003; Sahai, 2005; Sierra, 2005; van Kempen, *et al.*, 2003; K. Wolf & Friedl, 2006; Yamaguchi, Wyckoff, & Condeelis, 2005) by providing a platform to study cancer as a *system*. Next, we describe biologically founded, *in silico* modeling efforts of tumor progression relying on known characteristics of tumor behavior to predict the combination of variables most likely driving progression towards invasiveness.

This effort builds on an approach (*e.g.*, (Cristini, *et al.*, 2005; Cristini, *et al.*, 2003; H. B. Frieboes, *et al.*, 2006; X. R. Li, *et al.*, 2007; X. Zheng, *et al.*, 2005)) that includes reformulations and generalizations of mathematical models (Adam, 1996; Ambrosi & Preziosi, 2002; D. M. Anderson, McFadden, & Wheeler, 1998; Bellomo & Preziosi, 2000; H. Byrne & Preziosi, 2003; H. M. Byrne & Chaplain, 1996; H. M. Byrne & Chaplain, 1996; M. A. Chaplain & Anderson, 2003; M. A. J. Chaplain, Graziano, & Preziosi, 2006; Garcke, Nestler, & Stinner, 2004; H. P. Greenspan, 1976; Jackson & Byrne, 2002; Jacqmin, 1999; Lee, Lowengrub, & Goodman, 2002; Leo, Lowengrub, & Jou, 1998; Lowengrub & Truskinovsky, 1998; Macklin & Lowengrub, 2005, 2007b), solved numerically using state-of-the-art algorithms and techniques (Berger & Colella,

1989; Brandt, 1977; Cristini, *et al.*, 2001; J. Kim, Kang, & Lowengrub, 2004; Wise, *et al.*, 2008; Wise, Lowengrub, Kim, & Johnson, 2004; Wise, *et al.*, 2005; X. M. Zheng, *et al.*, 2005). Figure 4.2 describes the main component modules (Vasculature, Tumor, and Genotype) of this model along with equations that represent mathematically the relevant biological parameters.

Determination of Functional Relationships and Parameter Values

The specific process of multi-scale model “training” relies on conducting experiments (Figure 4.2) in which molecular factors are measured in the cell and the environment, and outcome of tumor growth (*e.g.*, morphology, shape, extent of vascularization and invasion) is correlated with expression of these factors. This data allows estimation of the mathematical model parameters and functional relationships by perturbing these parameters and comparing the resulting simulation predictions of morphology against direct measurements, thus leading, through an iterative process that reveals deficiencies in modeling choices and triggers refinements in the relationships introduced, to a validated mathematical model with calibrated constitutive parameters. By virtue of its predictive power, this approach (Bearer, *et al.*, 2009) can help plan new experiments by identifying parameter regimes of noteworthy behavior—regimes that might otherwise be time-consuming and costly to discover by systematic experimentation.

Theoretical (*e.g.*, (H. B. Frieboes, *et al.*, 2006)) and experimental work (*e.g.*, (Chomyak & Sidorenko, 2001; H. B. Frieboes, *et al.*, 2006; J. B. Kim, 2005; Mueller-Klieser, 2000) and references therein) can be used to develop and test functional relationships, and to estimate the microphysical parameter values of a multi-scale *in silico* model. Examples (Bearer, *et al.*, 2009) of these functional relationships include

those between expression of membrane transport proteins (*e.g.*, glucose transporter-1 and Na/H exchanger) and hypoxia/proliferation; between extracellular matrix macromolecules (*e.g.* tubulin, actin), haptotaxis and chemotaxis; and between cell-cell adhesion parameters as an increasing function of oxygenation, *e.g.*, from recent measurements by Robert Gatenby (personal communication) showing a gradient of cell-adhesion molecules (E-cadherins) opposed to hypoxia.

In vivo animal models (*e.g.*, dorsal wound chamber by (H. B. Frieboes, *et al.*, 2006)) can supply detailed measurements of angiogenesis and blood flow, which provide additional constraints to the *in silico* model to determine parameter values associated with a developing neovasculature. Computational models of angiogenesis (H. A. Levine, *et al.*, 2002; McDougall, *et al.*, 2006; Plank & Sleeman, 2003, 2004; Stephanou, *et al.*, 2005; Sun, *et al.*, 2005) (Figure 4.2) can account for endothelial cell chemotactic and haptotactic movement, proliferation, development and remodeling of capillaries and the flow of blood through the local pressure and other constraints. Under *in vivo* conditions, additional measurements can be performed to determine pH and pO₂ gradients that provide further functional constraints on the parameters relating to proliferation and cellular adaptation to hypoxia (Bearer, *et al.*, 2009). Finally, *in vivo* measurements of matrix degradation at the tumor-host boundary due to acidosis and proteases can provide parameter values for the invasion component of the *in silico* model (Bearer, *et al.*, 2009).

COMPUTATIONAL MODELING: A FRAMEWORK FOR LINKING PHENOTYPE, MORPHOLOGY, AND CANCER INVASION

Extensive mathematical modeling has produced preliminary quantifications of the links between invasive, malignant cell phenotypes and tumor-scale morphologies. These involve cell-cell and cell-extracellular matrix (ECM) interactions, cell motility, micro

vessel density and acidosis, and local concentration of cell substrates (Mareel & Leroy, 2003). We illustrate representative discrete models in each of these areas. We then review recent mathematical and computational studies of a continuum FCMU model developed in our group. Both discrete and continuum models are based on conservation formulations such as described in Figure 4.2.

Effects of Cell-Cell and Cell-Matrix Interactions

The effects of tumor cell and environment heterogeneity on the overall tumor morphology were recently studied (A. R. A. Anderson, 2005) by capturing spatial distribution of oxygen, matrix-degrading enzymes, and matrix molecules in the tumor microenvironment as continuous variables, while tumor cell properties, *e.g.*, their migration, were represented by a discrete automata-like model. Results support the notion that tumor cell-matrix interactions ultimately control tumor shape by driving tumor cell migration *via* haptotaxis and chemotaxis towards fingering, invasive tumor morphologies (Figure 4.3 **A** and **B**).

However, degradation of ECM, specifically surrounding the tumor boundary, may have a stronger influence on invasion than cell-cell adhesion. Using a derivation of a Potts model (F. Y. Wu, 1982) that incorporates homotypic and heterotypic adhesion as well as secretion of proteolytic enzymes that drive haptotaxis along their gradient, a more quantitative perspective into the role of cell adhesion and proliferation in promoting an invasive phenotype was obtained (S. Turner & Sherratt, 2002). The model assumes genetic mutations affect cellular adhesiveness, secretion of matrix degrading enzymes, the ability to undergo taxis along gradients, and proliferation rate (Stetlerstevenson, Aznavoorian, & Liotta, 1993; S. Turner & Sherratt, 2002). Using the maximum host tissue penetration as an index of invasiveness, simulation results predict that increases in

the secretion of matrix degrading enzymes in synergy with increases in cell proliferation and haptotaxis can produce fingering morphologies at the tumor-host interface as cells adhere to the ECM and spread into host tissue. The model hypothesizes a functional relationship between proliferation rates and changes in adhesiveness based on experimental evidence (S. Huang & Ingber, 1999).

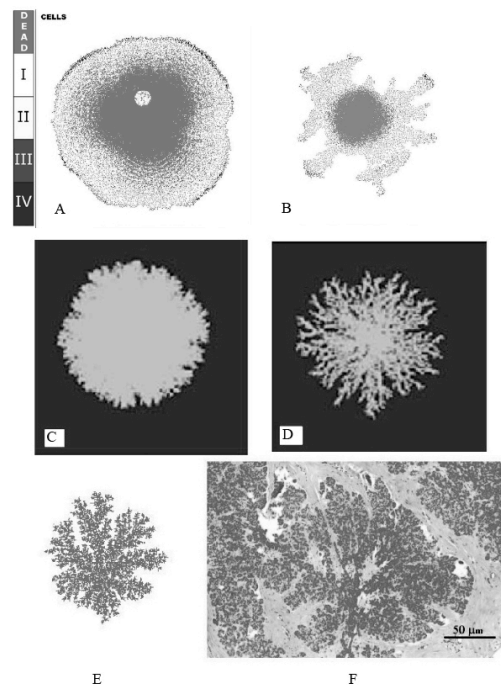


Figure 4.3: Effects of tumor cell and environment heterogeneity on overall tumor morphology predicted by a number of “discrete” (each tumor cell’s position is “tracked” during a simulation) computational models recently introduced.

Simulation results from Anderson (2005) show the importance of tumor cell-matrix interactions in aiding or hindering migration of individual cells thus defining the overall tumor-scale geometry (A and B). Tumor types I-IV correspond to cell phenotypes displaying increasing levels of aggressiveness, *i.e.*, combinations of cell-cell adhesiveness, proliferation, degradation, and migration rates. Both simulations use the same parameter values with the exception of differing initial ECM conditions (*i.e.*, different distributions of matrix molecules). In (A), the matrix environment is initially described as homogeneous, whereas it is heterogeneous in (B). Consequently, (B) depicts invasive, fingering morphology. Cells of phenotype IV are on the tumor-host boundary, promoting invasion into host tissue, emphasizing that more aggressive cells drive fingering morphologies. These aggressive cells have minimal cell-cell adhesion, thus they do not tend to form compact structures. Simulations are carried out on a 400 x 400 grid representing 1 cm² of a virtual, 2-D tumor. Adapted from Anderson, A.R.A, A hybrid mathematical model of solid tumor invasion: the

importance of cell adhesion, *Mathematical Medicine and Biology*, 2005, vol. 22, issue 2, pages 175-176, by permission of Oxford University Press. Simulation results from dos Reis *et al.* (2003) showing how tumors growing in host tissue environments of low (C) and high (D) “rigidity” can influence compact, non-invasive (C) and fractal, fingering, invasive (D) morphologies. Simulations are carried out to approximately 5000 cells, where cells are represented as interacting particles in a 2-D continuous space with periodic boundary conditions. Reprinted from *Physica A*, vol. 322, dos Reis *et al.*, The interplay between cell adhesion and environment rigidity in the morphology of tumors, page 550, Copyright 2003, with permission from Elsevier. Cell patterns simulated using the model of Ferreira *et al.* (2002) (E) suggest how parameters characterizing cancer cells’ response to nutrient concentrations and embodying complex genetic and metabolic processes can influence the formation of fractal, fingering tumor morphologies. For comparison, a real fractal pattern observed in trichoblastoma (F). Reprinted figure with permission from Ferreira *et al.*, *Physical Review E*, 65, page 021907-6, 2002. Copyright 2002 by the American Physical Society.

The notion that formation of fingering protuberances at the tumor-host boundary is primarily due to an intrinsic physical property termed rigidity of the host environment to resist tumor growth has also been computationally examined (dos Reis, *et al.*, 2003). Low rigidity allows a tumor to expand through the host environment resulting in a well-defined tumor-parenchyma interface, whereas higher rigidity forces a tumor to grow by invading the host tissue resulting in a fingering morphology (Figure 4.3, C and D), as predicted by simulation results. In addition, cell adhesion changes growth patterns from fractal morphologies at the tumor-host interface to compact shapes.

Effects of Cell Motility

Computational investigations of the invasiveness of high, medium, and low-grade gliomas illustrate that the ratio of tumor growth rate and cell motility can quantify the invasive nature of a tumor (Swanson, *et al.*, 2000). Specifically, this ratio might be useful in predicting a tumor’s invasive and metastatic potential; high proliferation rates and low motility correspond to lower grade tumors with less invasive potential whereas low proliferation rates and high motility correspond to higher-grade tumors with more invasive potential.

In contrast, a 3D cellular automaton model of glioblastoma capable of predicting tumor growth according to four microscopic parameters (probability of division, necrotic

thickness, proliferative thickness, and maximum tumor extent) successfully predicted tumor-scale dynamics of a test case for untreated glioblastoma progression compiled from medical literature; simulations reproduced data such as lesion radius, cell number, growth fraction, necrotic fraction, and volume doubling time at particular time points (Kansal, *et al.*, 2000a; Kansal, *et al.*, 2000b). Human glioblastoma patients have a median survival time of 8 months from diagnosis, which these models accurately predict using primary tumor volume as an indicator for survival.

Effects of Micro Vessel Density and Acidosis

The potential importance of micro vessel (MV) density and acidosis in promoting tumor growth and invasion has been demonstrated through recent computational models (H. B. Frieboes, *et al.*, 2006; R. A. Gatenby & Gawlinski, 1996, 2003a; Patel, *et al.*, 2001). Simulations show that the production rate of H^+ ions by cancer cells, due to their increased dependence on anaerobic glucose metabolism, is linked to an optimal micro vessel (MV) density such that the microenvironment favors tumor cells over normal cells, hence promoting growth and invasion (Patel, *et al.*, 2001). MV density below this optimal value produces an environment too acidic even for cancer cells, while MV density above the optimum reduces or even completely negates the advantage enjoyed by cancer cells over normal cells in an acidic environment, thus inhibiting overall tumor growth and invasion by promoting nutrient competition.

Depending on the metabolic phenotype, various tumor morphologies can be predicted including invasive, fingering protrusions seen in experiments and with other *in silico* models. This and other modeling and experimental work further supports the acid-mediated tumor invasion hypothesis (H. B. Frieboes, *et al.*, 2006; R. A. Gatenby & Gawlinski, 1996, 2003a; Patel, *et al.*, 2001), illustrating the potential importance of MV

density in driving pH gradients in the microenvironment and associated tumor-scale behavior. Such microenvironmental factors, in addition to cellular dynamics, are thus quantitatively linked to tumor-scale morphology.

Effects of Cell Substrate Concentration

Competition for nutrient and oxygen amongst normal and cancer cells, in addition to cell proliferation, motility, death, and secretion of matrix degrading enzymes, may be an important factor driving tumor invasion. Using a cellular-automaton model (Ferreira, *et al.*, 2002), cell dynamics were described where at each time step, a cell (of type normal, cancer, or necrotic tumor) has equal probability of dividing, migrating, or undergoing necrosis; each action is governed by the local substrate concentration. Cells are modeled to release a series of enzymes that progressively degrade the ECM, thus providing more space for tumor cells to invade. In Figure 4.3, **E** and **F**, fingering morphologies are predicted by the model (and observed) as a result of high proliferation rates demanding large amounts of substrates. Predicted tumor morphology remains compact in situations of high nutrient supply and low cell consumption, while cell clusters expressing a phenotype that increases nutrient consumption exhibit thinner “fingers.”

Continuum-Based Parameter-Sensitivity Studies of FCCMU

Morphologic Instability as a Mechanism of Tumor Invasion

The current conceptual framework of continuum FCCMU models is based on reaction–diffusion formulations (Figure 4.2). Accordingly, tumor morphologic “stability” is regulated by the competition of pro- and anti- migratory/proliferative factors. When the

former prevail, complex, unstable FCCMU patterns can develop (Cristini, *et al.*, 2003; X. R. Li, *et al.*, 2007). The power of this approach is that it is based on a physical mechanism that can account for the various invasive morphologies observed, and is thus potentially predictive of tumor growth. This mathematical analysis of morphologic stability has suggested that tumor tissue dynamics is regulated by two dimensionless parameters: the parameter G quantifies the competition between local tumor mass growth due to proliferation, and cell adhesion that tends to minimize tumor surface area and thus maintain compact nearly spherical tumors; the parameter A quantifies tumor mass shrinkage due to cell death (these parameters are obtained from some of those listed in Figure 4.2 using dimensional analysis; for the sake of simplicity, the associated cumbersome formulation is not reported here: see (Cristini, *et al.*, 2003; X. R. Li, *et al.*, 2007)). During glioblastoma tumor growth *in vitro*, cell proliferation is observed in a viable region where nutrients, oxygen, and growth factor levels are adequate, and cell death and necrosis in the inner regions where diffusion limitations prevent these substances from being present in adequate levels (H. B. Frieboes, *et al.*, 2006). *In the presence of these substrate gradients*, morphology can be “unstable”, *i.e.*, invasive, when cell adhesion is weak (large G). In contrast for small G , spheroid morphology is “stabilized” (*i.e.*, spherical or nearly spherical) by cell adhesion (Cristini, *et al.*, 2003).

This is illustrated in a morphologic stability diagram (Cristini, *et al.*, 2003; H. B. Frieboes, *et al.*, 2006), Figure 4.4, A. The G -curves divide the parameter space into stable (left) and unstable (right) regions. The G -curves are more shifted to the left as cell adhesion decreases (higher G), thus reducing the range of sizes of tumors that will be morphologically stable. The *in silico* model parameters A and G were calibrated using data from “stable” spheroids (shaded area) until agreement was obtained between the model calculations of morphology and growth and *in vitro* measures of tumor growth

curves and thickness of the viable rim of cells. The model was then tested by predicting morphology stability conditions for an independent set of experiments (filled symbols) where the cell medium levels of growth factors and glucose were changed over a wider range to manipulate glioblastoma cell proliferation and adhesion (Frieboes *et al.*, 2006a).

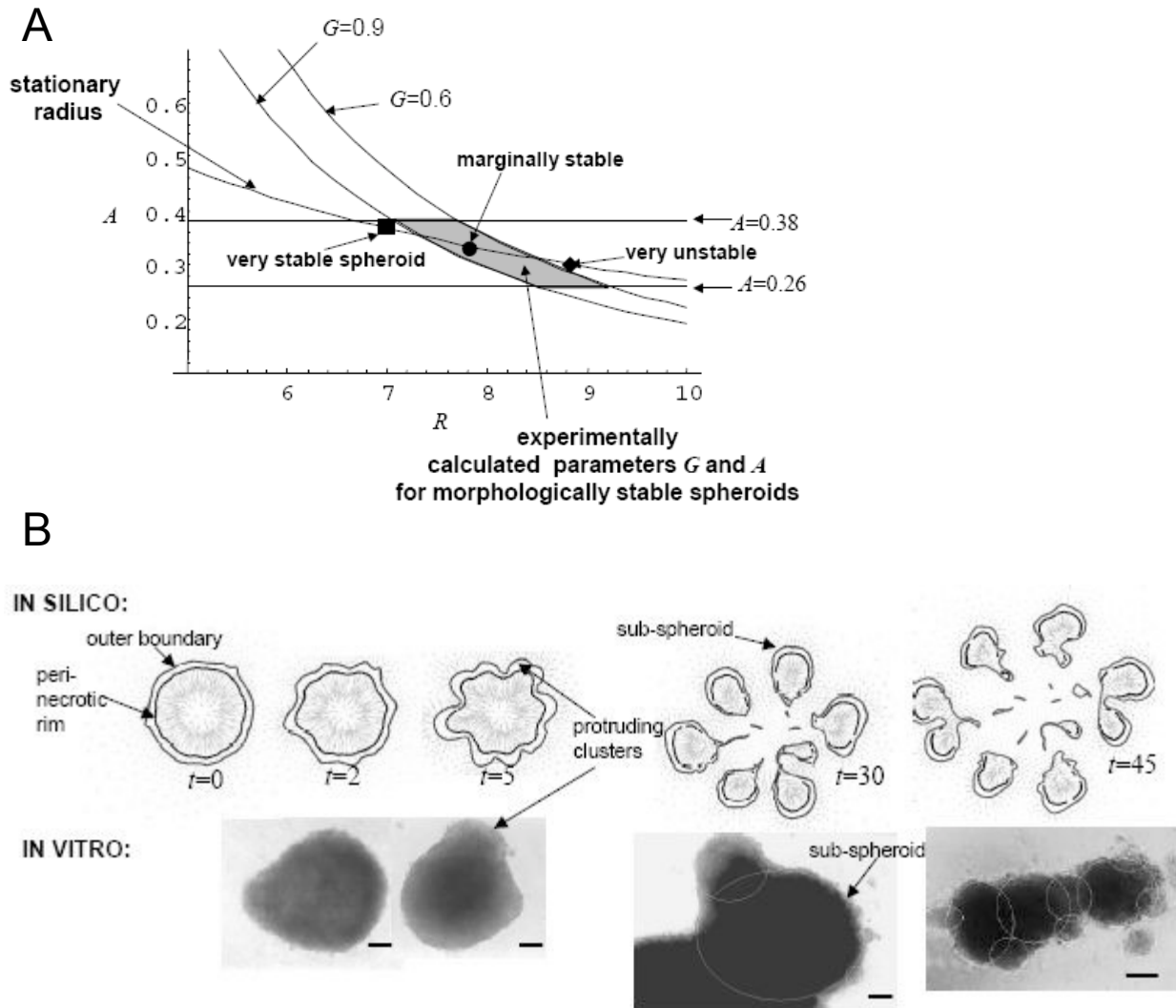


Figure 4.4: In silico model predictions and in vitro measurements of locally invasive cell clusters in collective migration, using a “continuum model” (see text for a definition).

Adapted from Frieboes *et al.* (2006) with permission from the American Association of Cancer Research. In the morphologic stability diagram (A), obtained from a mathematical analysis of the model, “stationary radius” describes tumor dimension R (unit length = 100 μ m), monotonically decreasing as death (described by the parameter A) increases. The G -curves calculated by the mathematical model divide, for each value of G (a parameter related to cell adhesion), the parameter space into morphologically stable (left) and unstable (right) tumors. Stable tumors remain roughly spherical during growth; unstable tumors are invasive and form wavy protrusions at the tumor-host boundary that further develop into sub-tumors that break-up from the parent tumor (B). The shaded region was determined from calibration of the parameter values under “stable” *in vitro* conditions. Representative “stable” and “unstable” spheroids (filled symbols) from different sets of experiments are shown to agree with the model predictions. This means that the mathematical model was capable of predicting invasive behavior of tumors under

conditions of growth factors and substrate concentrations different from the “control” that was used to calibrate the model parameters. These results indicate that wavy patterns of cell arrangements at the tumor-host boundary could be inputted to a mathematical model to predict future invasive potential. **(B)** Time progression (arbitrary units) of avascular glioma predicted by simulations of the *in silico* model (top) compared to observations *in vitro* (bottom). Tumor morphology and invasiveness are predicted to be heavily influenced by substrate gradients (*e.g.*, nutrient) in the cellular microenvironment, driving detachment of bulbs or clusters of cells. Bar, $130 \mu\text{m}$.

A remarkable result of this study was that the *in silico* model was capable of predicting growth and invasion of tumors from experiments that were not used for model training, thus successfully testing, under relatively simple highly controllable *in vitro* conditions, the hypothesized phenomenological relationships of adhesion and proliferation with substrate levels and their effects on tissue-scale growth and morphology (*e.g.*, see Figure 4.2). Computer simulations and experimental observations of “unstable” spheroids that develop protrusions and detachment of cell clusters are shown in Figure 4.4, **B**. As described below, these infiltrative morphologies are also universally observed in tumors *in vivo* and in data from patients.

Clinical Relevance

Morphologic instability during tumor growth is predicted to result from genomic changes that produce variations in sub-tumor clonal expansion, rates of mitosis and apoptosis, oxygen consumption, and diffusion gradients. This physical hypothesis is corroborated by *in vitro* and *in vivo* observations (Bello, *et al.*, 2004; H. B. Frieboes, *et al.*, 2006; Kunkel, *et al.*, 2001b; Lamszus, *et al.*, 2003; Rubenstein, *et al.*, 2000) and by

patient data. In a study of several clinical samples of glioblastoma multiforme from multiple patients (Figure 4.5), histology reveals protruding fronts of cells in collective motion away from a necrotic area into an area of the host brain where neo-vascularization is present, thus following substrate gradients. This data strongly resembles the morphology of the tumor boundary predicted by computer simulation (Bearer, *et al.*, 2009; H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Elaine, *et al.*, 2007) and by the *in vitro* experiments described above (H. B. Frieboes, *et al.*, 2006). These infiltrative shapes were consistently observed in high-grade gliomas, although their size may vary.

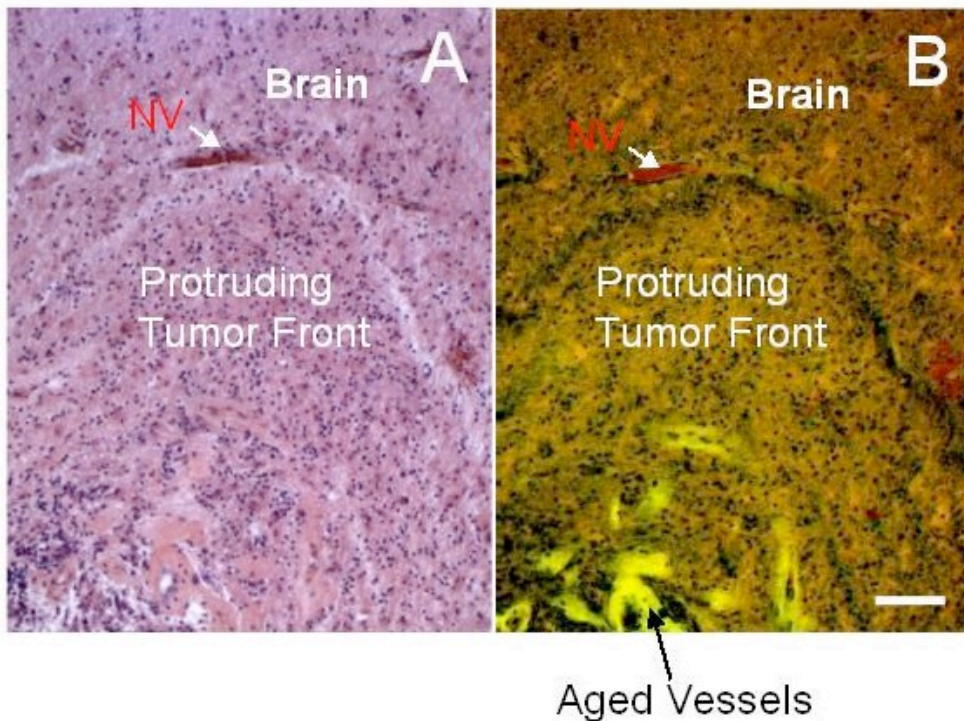


Figure 4.5: Glioblastoma histopathological sections from one patient

Stained for H&E and viewed by bright field (A) and fluorescence (B) microscopy (Frieboes *et al.* 2007, reproduced with permission from Elsevier) showing tumor (bottom) pushing into more normal brain (top). Note demarcated margin between tumor and brain parenchyma to the middle top of the image and green fluorescent outlines of large vascular channels deeper in the tumor. Neovascularization (NV) at the tumor-brain interface can be detected by red fluorescence from the erythrocytes inside the vessels. Altogether, these data support the “morphologic instability” hypothesis. Bar, 100 μ m.

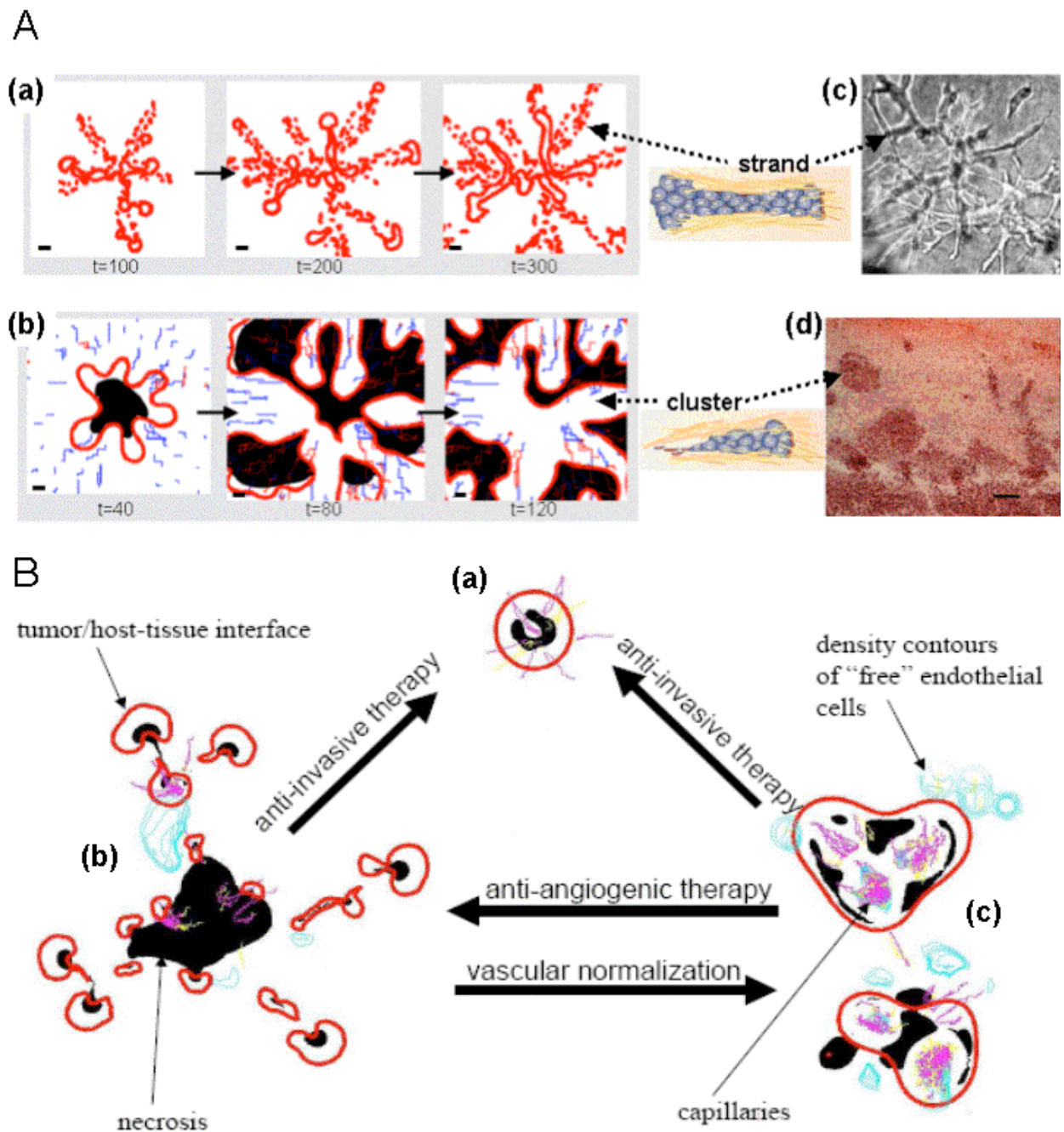


Figure 4.6: In silico predictions of tumor morphology based on varying cellular and micro-environmental conditions in a parameter-sensitivity simulation study of the continuum model by Zheng *et al.* 2005

(Note: figures are not to scale). These results extend the findings illustrated in Figure 4.4 and demonstrate that the *in silico* model can account for the variety of invasive morphologies observed in tumors, and that

the *in silico* model is thus potentially predictive of tumor growth. **(A)** When cell taxis but not proliferation is present, cells are predicted to align in chains or strands **(a)**. When cell proliferation is significant, more bulb- or cluster-like protrusions form and detach into the host **(b)**. Red: tumor boundary; Black: hypoxia; Blue and Pink: neovascularization (immature and mature, respectively); time units are arbitrary. Drawings of cell strand and cluster adapted by permission from Macmillan Publishers Ltd: Nature Rev. Cancer, Vol. 3, p. 366, Friedl & Wolf, Copyright 2003. Corresponding structures observed after inducing hypoxia *in vitro* (proliferation was also inhibited) **(c)** (Pennacchietti *et al.*, 2003) and *in vivo* (Rubenstein *et al.*, 2000) through anti-angiogenic therapy **(d)** are shown for comparison. Bar, 80 μ m. Reprinted from Cancer Cell, Vol. 3, Pennacchietti *et al.*, page 354, Copyright (2003), with permission from Elsevier. Reprinted from Neoplasia, vol. 2, Rubenstein *et al.*, page 311, Copyright 2000, with permission from Neoplasia Press. **(B)** Snapshots from three simulated tumor morphologies corresponding to different values of cell adhesion and vascularization parameters: high cell adhesion **(a)**; low cell adhesion **(b)**; low cell adhesion and with higher microvascular density or more efficient vascularization **(c)**. Arrows indicate morphology transitions following different therapy strategies. Adapted from Cristini *et al.* (2005) with permission from the American Association of Cancer Research.

Effects of Phenotype on Morphology and Growth

In Figure 4.6, **A**, different morphologies predicted by a continuum FCCMU model (Cristini, *et al.*, 2005; X. Zheng, *et al.*, 2005) are shown, starting from the same initial condition of a spherical tumor. The model predicts that when cell taxis, but no proliferation is present **(a)**, cells tend to align in chains or strands. When cell proliferation is significant **(b)**, more “bulb like” protrusions form and detach into the host. These are also predicted to be more hypoxic. In all cases, these complex morphologic patterns developed because cell adhesion parameters were set very low (“morphologic instability,” Cristini *et al.*, 2003). Corresponding structures observed after inducing hypoxia **(c)** *in vitro* (proliferation was inhibited) (Pennacchietti, *et al.*, 2003) and **(d)** *in vivo* (Rubenstein, *et al.*, 2000) are reported for comparison. The underlying molecular phenomena (Friedl & Wolf, 2003) responsible for the prevalence of one over another morphology, and for the spatial frequency of finger-like protrusions originating from a primary tumor, are captured by (phenomenological) model parameters describing proliferation and taxis and the associated convective cell fluxes on one side, and cell-adhesion forces on the other (Cristini, *et al.*, 2003).

This model was further used to predict changes in the system dynamics following

a range of possible perturbations of parameters related to cell-adhesion forces and to oxygen distribution in the environment, with the goal of providing suggestions for novel treatment protocols aimed at restoring normoxia and thus preventing “unstable” morphologies (Cristini, *et al.*, 2005; H. B. Frieboes, *et al.*, 2006). Since these phenotypic and environmental parameters are also associated with invasion, this perturbation study provided preliminary quantification of the effects of anti-invasive and “vasculature-normalizing” anti-angiogenic therapeutic strategies that alter the balance of morphology-stabilizing and -destabilizing micro-environmental and molecular processes. In particular, this study helped explain the undesirable effects on morphology following current anti-angiogenic therapy due to exacerbation of micro-environmental hypoxic gradients and enhancement of cell migration (as reviewed above).

In Figure 4.6, **B**, case **(a)** corresponds to sufficiently high cell adhesion so that the simulated tumor growth is morphologically “stable”. Due to hypoxic gradients, necrotic regions have formed where concentrations are inadequate, leading to a diffusion-limited tumor size. Angiogenic factors (not shown) emanate from the peri-necrotic regions and diffuse outward, reaching pre-existing vessels and triggering neo-vascularization of the lesion. Even after angiogenesis, the model predicts that lesion **(a)** will maintain a compact shape because of high cell adhesion.

Case **(b)** corresponds to low cell adhesion, in which the tumor experiences morphological instability driven by hypoxic gradients as it progresses (H. M. Byrne & Chaplain, 1997; Cristini, *et al.*, 2003; Macklin & Lowengrub, 2007b; X. Zheng, *et al.*, 2005). Cell adhesion is insufficient to maintain proliferating cells together. The lesion breaks up into fragments, or detached cell clusters (Friedl & Wolf, 2003), moving away following outward gradients of nutrient and oxygen concentration (Macklin & Lowengrub, 2007b; X. Zheng, *et al.*, 2005). The model predicts that anti-invasive therapy enforces a morphology transition from **(b)** to **(a)** by increasing cell adhesion.

Case **(c)** corresponds to low cell adhesion (as in **(b)**), but with a more spatially uniform distribution of vessels. The simulation predicts that this “vascular normalization” would lead to reduced oxygen gradients, and hence to suppression of instability and to clustering of cells into a more compact tumor morphology. This result could be achieved by pruning immature and inefficient blood vessels, leading to a more normal vasculature of vessels reduced in diameter, density, and permeability (Jain, 1990, 2001a, 2001b). In contrast, after anti-angiogenic therapy (**(c)** to **(b)**), increased scattering of tumor cell clusters in response to hypoxia is predicted, as documented *in vivo* by (Bello, *et al.*, 2004) and by others. Remarkably, the simulations also predict that in this case some tumor cell clusters tend to co-opt the vasculature to maximize nutrient uptake, as documented previously *in vivo* (Kunkel, *et al.*, 2001b; Lamszus, *et al.*, 2003; Rubenstein, *et al.*, 2000).

Figure 4.7 shows a summary of some of the biology revealed by the predictive model reviewed here (Cristini, *et al.*, 2005; H. B. Frieboes, Lowengrub, Wise, Zheng, Macklin, Elaine, *et al.*, 2007; H. B. Frieboes, *et al.*, 2006; Sanga, *et al.*, 2006; J. Sinek, *et al.*, 2004; X. Zheng, *et al.*, 2005) under the categories of Tumor, Microenvironment, Treatment Response, and Vasculature, including gross tumor morphology in 3-D **(A)**, gradients of cell substrates **(B)**, tissue fragmentation in response to chemotherapy involving large nanoparticles and adjuvant anti-angiogenic therapy **(C)**, and tumor vasculature with both conducting and non-conducting vessels **(D)**.

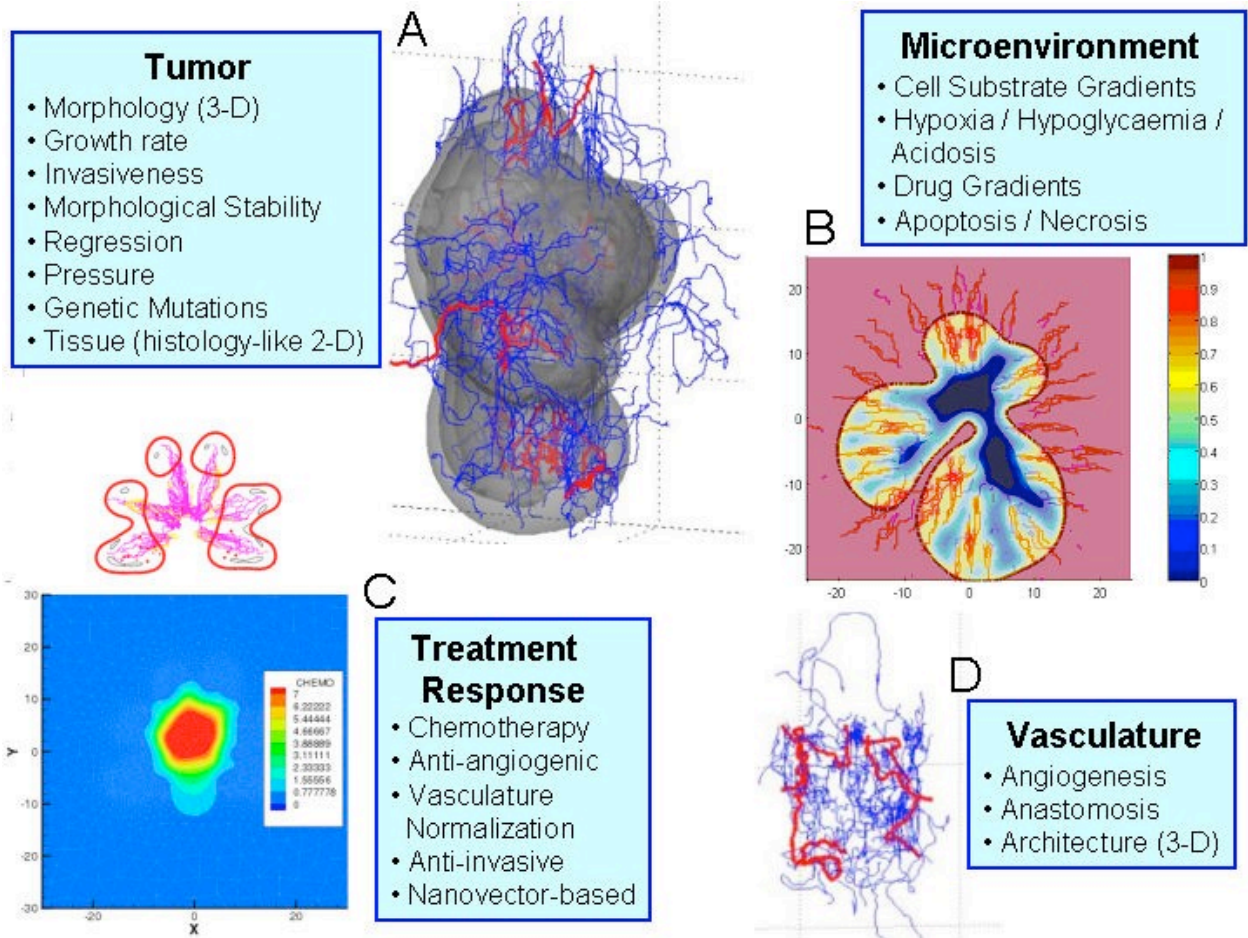


Figure 4.7: Tumor biology revealed by parameter-sensitivity studies of a continuum FCCMU computer model is listed under the categories of Tumor, Microenvironment, and Vasculature, along with response to various treatments.

(A) Gross tumor growth and morphology in 3-D are predicted based on cell-scale parameters (e.g., proliferation, cell adhesion) set from experimental values. Viable (light grey) and necrotic (dark grey) tissue as well as extensive vascularization (conducting vessels in red, non-conducting in blue) are shown. Reprinted from *NeuroImage*, Frieboes *et al.* (2007), Copyright 2007, with permission from Elsevier. (B) Gradients of cell substrates (from highest (red) to lowest concentration (blue)) are predicted from the vasculature topology (dark red lines). Thin dashed line denotes tumor boundary. Reprinted from *Journal of Mathematical Biology*, Sinek *et al.* (2009), Copyright 2007 Springer. With kind permission of Springer Science and Business Media. (C) Local tumor fragmentation (top) is predicted in response to chemotherapy involving large nanoparticles and adjuvant anti-angiogenesis (bottom). Boundary of tumor fragments is in red, vessels are pink (conducting) and light blue (non-conducting). Gradient of drug (red, highest, blue, lowest) is centered in middle area of tumor tissue. Adapted from *Biomedical Microdevices*, Vol 6, 2004, p. 307, Sinek *et al.*, Figure 4.5. Copyright 2004, Kluwer Academic Publishers. With kind permission of Springer Science and Business Media. (D) Abnormal tumor vasculature architecture with conducting (red)

and non-conducting vessels (blue) is predicted based on angiogenic regulators produced by tumor and host tissue. Reprinted from *NeuroImage*, Frieboes *et al.* (2007), Copyright 2007, with permission from Elsevier.

CONCLUSIONS AND FUTURE WORK

The research direction we envision focuses on the development and application to tumor biology of quantitative methods traditionally confined to engineering and the physical sciences. Indeed, it is clear that such complex biological systems dominated by large numbers of processes and highly complex dynamics are very difficult to approach by experimental methods alone. They can typically be understood only by using appropriate mathematical models and sophisticated computer simulations complementary to experimental investigations. In the innovative and powerful multidisciplinary approach reviewed here, mathematical and computational modeling completes the circle of discovery: laboratory experiments provide data that, in turn, informs the construction of a mathematical model that can then predict behavior and guide the design of future experiments to test these predictions.

This multi-scale approach captures tumor progression by taking into account ongoing molecular and cellular scale events (H. B. Frieboes, *et al.*, 2006; MartinezZaguilan, *et al.*, 1996; Rofstad & Danielsen, 1999; Schlappack, Zimmermann, & Hill, 1991). One of the key links established in a more quantitative manner is that mutations drive increased cellular uptake, which introduces perturbations in spatial gradients of oxygen and nutrient; these gradients enhance hypoxia and cause heterogeneous cell proliferation and migration leading towards diffusional shape instabilities. This supports the hypothesis that cellular and extra-cellular properties driving tumor growth and invasiveness also determine tumor morphology (Cristini, *et al.*, 2005; H. B. Frieboes, *et al.*, 2006) and suggests that morphological characteristics

including neo-vasculature and harmonic content of the tumor edge should serve as predictors of tumor growth (Cristini *et al.*, 2006).

Predictive modeling assumes that criteria and critical microphysical conditions for tumor invasion can be formulated in terms of physical laws linking tissue architecture and morphology to cell phenotype. Future multidisciplinary investigations should exploit the power of predictive modeling that allows observable properties of the tumor, such as its morphology in general and specifically the cell spatial arrangements at the tumor boundary, to be used to both understand the underlying cellular physiology and predict subsequent invasive behavior. We envision this research taking steps towards further establishing the dependence of tumor cell motion into surrounding host tissue on the balance between cell proliferation and adhesion, as well as perturbations caused by microenvironmental factors such as oxygen, nutrient, and H^+ diffusion gradients. This will include the continuing application of mathematical and empirical methods to quantify the competition between gradient-related pro-invasion phenomena and molecular forces that govern proliferation and taxis, and forces opposing invasion through cell adhesion. In addition, a more detailed description of the complex *in vivo* environment, which better recapitulates the conditions of tumors in patients, would be valuable.

Currently, pathologic analysis is often limited to a set of morphological features that are rarely quantitatively assessed (the main quantitative factors are mitotic rates and size of invasive tumor “fingers”), and these measures differ depending on the types of tumor. “Degree of pleiomorphism” (variable phenotypes) is also used as a prognosticator, although this has no absolute quantitative definition and is subjective. Multi-scale modeling of cancer would allow predictions of cellular and molecular perturbations that alter invasiveness and are measured through changes in tumor morphology. This opens

the possibility of designing novel individualized therapeutic strategies in which the microenvironment and cellular factors are manipulated with the aim of imposing compact tumor morphology by both decreasing invasiveness and promoting defined tumor margins—an outcome that would benefit cancer therapy by improving local tumor control through surgery or chemotherapy.

ACKNOWLEDGEMENTS

We gratefully acknowledge John Lowengrub and Steven M. Wise (Mathematics, U.C. Irvine), and the reviewers for helpful comments and suggestions. We thank Ed Stopa at Rhode Island Hospital and the Columbia University Alzheimer's Brain Bank for archived human glioma specimens. Funding from NIH-NIGMS RO1 GM47368 and RO1- NS046810 (E.L.B.), from the National Science Foundation (V.C.) and National Cancer Institute (V.C.; R.G.).

Chapter 5: Predicting Drug Pharmacokinetics and Effect in Vascularized Tumors Using Computer Simulation

PREFACE

The field of predictive oncology conventionally applies methodologies such as microarrays and immunohistochemical staining for comprehensively profiling gene expression and protein activities of genes in cancer tissue, and identifying biomarkers and signatures that are either prognostic and/or predictive of chemotherapy response. However, the complex nature of cancer has made it difficult to identify unique molecular and pathophysiological signatures for each disease variant, consequently hindering the ability to predict the performance of therapies in individual patients solely using these approaches (Sanga, *et al.*, 2006). In the next chapter, we take an alternate approach; we investigate the drug delivery (pharmacokinetics) and effect (pharmacodynamics) of doxorubicin and cisplatin in vascularized tumors and show that microenvironmental considerations such as lesion-scale drug and nutrient distributions may significantly hamper therapeutic efficacy and should be considered as carefully as genetic and proteomic determinants. Our model takes into account tumor vascularity and morphology as well as cellular and lesion-scale pharmacokinetics determinants such as p-glycoprotein efflux and cell density. Drug transport is encapsulated using a multi-compartment model calibrated from published experimental data; this model tracks drug as it extravasates from the blood stream into the tumor interstitial space, diffuses through the lesion, enters cells and eventually reaches its intended target: DNA. Cell inhibition is modeled as a function of this DNA-bound drug. Unlike a truly *in vivo* situation, our *in silico* model provides the means to quantify expected *in vivo* IC50 under varying drug, oxygen, nutrient, and drug transporter conditions. The nonlinear interaction among various

determinants representing cell and lesion phenotype as well as therapeutic strategies towards predicting tumor response to therapy is a unifying theme of our results. Our results suggest that macroscopic environmental conditions, notably drug and nutrient distributions, give rise to considerable variation in tumor response to chemotherapy, hence clinical resistance. Moreover, the synergy or antagonism of combined therapeutic strategies depends heavily upon this environment.

This research was a collaborative effort between Dr. John P. Sinek, Sandeep Sanga, and Dr. Xiaoming Zheng under the supervision of Dr. Vittorio Cristini. Sinek and Sanga strategized the model development, model calibration, coordinating model simulations, and analyzing results. Sinek took the lead role in preparing the manuscript for publication with the assistance of Sanga. Sanga took a lead role in running model simulations presented in the research study. Analysis of simulation results was conducted by Sinek, Sanga, and Cristini. The core tumor modeling framework was originally developed by Zheng, Wise, and Cristini (X. Zheng, *et al.*, 2005), and was adapted by Zheng to model drug delivery using the pharmacokinetics model. Cristini directed the progress of the research study.

Note: This chapter is based on an article originally published as (J. P. Sinek, *et al.*, 2009) in the *Journal of Mathematical Biology*, which can be accessed at <http://dx.doi.org/10.1007/s00285-008-0214-y>. This article has been included in this dissertation with permission from Springer.

ABSTRACT

In silico modeling is an increasingly powerful tool in probing processes driving cancerous behavior and, in particular, tumor response to therapeutics. It has the capability of integrating individually well-understood phenomena across a wide range of scales, such as the vascular extravasation of oxygen and drug, interstitial diffusion, and cells' response to local concentrations in a spatially multi-dimensional setting, thereby giving

rise to behavior that is more than the sum of its inputs. As such, the model can be used as an *in silico* laboratory, investigating tumor therapeutic response under almost any conceivable set of conditions. The results can suggest hypotheses to be corroborated by further *in vitro* and *in vivo* experimentation, the convergence of results shedding light on cancer phenomena and treatment.

In this chapter we investigate the pharmacokinetics and effect of doxorubicin and cisplatin in three simulated two-dimensional vascularized tumors. We take into account especially vascular and morphological heterogeneity as well as cellular and lesion-level pharmacokinetic determinants like Pgp efflux and cell density. To do this we construct a multi-compartment PKPD model calibrated from published experimental data and simulate two-hour bolus administrations followed by 18-hour drug washout. Our results show that lesion-scale drug and nutrient distribution may significantly impact therapeutic efficacy and should be considered as carefully as genetic determinants modulating, for example, the production of multidrug-resistance protein or topoisomerase II. We visualize and rigorously quantify distributions of nutrient, drug, and resulting cell inhibition. A main result is the existence of significant heterogeneity in all three, yielding poor inhibition in a large fraction of the lesion, and commensurately increased serum drug concentration necessary for an average 50% inhibition throughout the lesion (IC_{50}). For doxorubicin the effect of hypoxia and hypoglycemia (“nutrient effect”) is isolated and shown to further increase cell inhibition heterogeneity and double the IC_{50} , both undesirable. We also show how the therapeutic effectiveness of doxorubicin penetration therapy depends upon other determinants affecting drug distribution, such as cellular efflux and density, offering some insight into the conditions under which otherwise promising therapies may fail and, more importantly, when they will succeed. Cisplatin is used as a contrast to doxorubicin since both published experimental data and our simulations indicate its lesion distribution is more uniform than that of doxorubicin.

Because of this some of the complexity in predicting its therapeutic efficacy is mitigated. Using this advantage, we show results suggesting that *in vitro* monolayer assays using this drug may more accurately predict *in vivo* performance than for drugs like doxorubicin.

The nonlinear interaction among various determinants representing cell and lesion phenotype as well as therapeutic strategies is a unifying theme of our results. Throughout it can be appreciated that macroscopic environmental conditions, notably drug and nutrient distributions, give rise to considerable variation in lesion response, hence clinical resistance. Moreover, the synergy or antagonism of combined therapeutic strategies depends heavily upon this environment.

INTRODUCTION

While drug resistance to solid tumors is often considered to be a consequence of genetic factors, such as over-expression of anti-apoptotic proteins or production of drug efflux pumps, factors residing at coarser physiological scales may have profound influence on tumor therapeutic response. Lesion mass is a heterogeneous three-dimensional composite of fibrous and connective tissues, stromal components, vasculature, and multiple clones of cancer cells. Atop this intrinsic heterogeneity is layered the anatomical and functional irregularity of tumoral vasculature, characterized by erratic flow, collapsed vessels, diminished oxygen tension, and a large mean tissue-to-vessel distance (Baish, *et al.*, 1996; Gullledge & Dewhirst, 1996; Haroon, Peters, Greenberg, & Dewhirst, 1999; Jain, 2001a; Padera, *et al.*, 2004). As a consequence, the tumor microenvironment is highly variable, marked by gradients of nutrient (oxygen and glucose), regions of hypoxia, acidity, and necrosis, and heterogeneous proliferation. In order for an anticancer agent to work, it must extravasate, diffuse through lesion, and then be transported into cells, where it must bind to its target and effect cell apoptosis or

mitotic inhibition. Clearly, the tumor environment is not conducive to these processes. The vessel bed's blood flow and spatial distribution hinder uniform extravasation, calling into question the capability of drug molecules to adequately distribute throughout tissue. *In vitro* experiments by Tannock and others underscore this concern, demonstrating limited penetration of drugs through lesion tissue, especially highly protein-bound molecules like doxorubicin and paclitaxel (Kohno, Ohnuma, & Truog, 1994; Tannock, Lee, Tunggal, Cowan, & Egorin, 2002; J. H. Zheng, Chen, Au, & Wientjes, 2001). Once a drug molecule has traversed lesion from its point of extravasation and is presented to a cell, the path from extracellular space to intracellular target is fraught with difficulties ranging from protonation due to the acidic environment, which renders anthracyclines incapable of traversing membrane, to intracellular removal by drug efflux pumps, to nuclear processes that effect DNA repair and drug clearance (Arancia, Calcabrini, Meschini, & Molinari, 1998; Demant & Friche, 1998; Hurwitz, Terashima, Mizunuma, & Slapak, 1997; Simon & Schindler, 1994; Takemura, *et al.*, 1991). In addition to pharmacokinetics, drug pharmacodynamics is equally impaired. Significant hypoxia and hypoglycemia throughout lesion tend to induce cell quiescence, and there is evidence that under these conditions, the efficacy of chemotherapeutic agents can be reduced (Durand, 1986, 1990). Quiescence may shield cells directly from the action of cycle phase-specific drugs like doxorubicin and cisplatin, and hypoglycemia can induce changes in topoisomerase II detrimental to the functioning of doxorubicin (Mellor, Ferguson, & Callaghan, 2005; Shen, Subjeck, Lock, & Ross, 1989; Siu, Arooz, & Poon, 1999).

The heterogeneity and three-dimensionality of the tumoral environment presents a challenge to drug assessment, both during development and in the clinic. Whereas a particular drug may show marked activity against a particular cancer line *in vitro*, its potency may vanish or become far less reliable *in vivo*. This is evidenced by the differential between positive predictive accuracy of *in vitro*-assisted therapy selection

(around 70%) and negative predictive accuracy (around 90%), a situation not remarkably changed over the years (J. P. Fruehauf, 2002; J.P. Fruehauf & Bosanquet, 1993). Supraoptimal delivery of drug to cultured cells eliminates the gauntlet of biobarriers *in vivo* described above, precluding the variability they induce. A drug that consistently works *in vitro* can therefore be expected to only sometimes work *in vivo*. Unraveling the myriad interactions of therapeutic determinants within the complex three-dimensional tumoral environment is evidently difficult, resulting in high costs of drug development and patient suffering.

Perhaps the crystal ball we are attempting to build is incomplete when made only of glass typically found in the wet-lab; the *in silico* lab, *i.e.*, computer simulation, might fulfill a key aspect of the lens. A significant capability of *in silico* experimentation (including simulated assays) is the complete control over and monitoring of the simulated *in vivo* tumor environment. Moreover, computer modeling can create hypothetical environments and conditions *impossible* to achieve otherwise, the study of which is nonetheless instrumental in unraveling disease and drug mechanisms. This expansive control, founded upon an adequately mechanistic mathematical basis, could facilitate the discovery of hypotheses as to why certain drugs or therapeutic strategies would or would not be effective, potentially on a patient-by-patient basis. The relative ease and cost-efficiency of performing simulation could furthermore enable a thorough investigation of strategies, revealing the optimal among them. The judicious combination of this burgeoning technology with the capabilities of the wet-lab is an attractive development in both drug discovery and the clinical management of cancer leading to the easing of patients' burdens.

The past two decades have witnessed explosive growth in the mathematical and computational modeling of vascular and avascular tumors (Alarcon, *et al.*, 2003; Araujo

& McElwain, 2004; Bellomo & Preziosi, 2000; Breward, Byrne, & Lewis, 2003; M. A. J. Chaplain, 1996; Cristini, *et al.*, 2003; Jiang, *et al.*, 2005; Please, Pettet, & McElwain, 2005). It was not long before similar models were employed to investigate lesion response to anticancer drugs (Jackson, 2003; Lankelma, Fernandez Luque, Dekker, Schinkel, & Pinedo, 2000; Norris, King, & Byrne, 2006; Ward & King, 2003). Notable limitations of these models, however, are their one-dimensionality (employing cylindrical or spherical symmetry) and lack of realistic vasculature. Without true multi-dimensionality and discrete vasculature, it is difficult to simulate the heterogeneities of nutrient and drug that profoundly affect therapeutic efficacy. In the present paper we use the two-dimensional (non-symmetric) tumoral growth engine of Zheng *et al.* (X. Zheng, *et al.*, 2005), which employs the discrete vasculature algorithm of Anderson and Chaplain (A. R. Anderson & Chaplain, 1998), to perform chemotherapy simulations using cisplatin and doxorubicin. In particular, we focus on the effects of drug and nutrient distribution heterogeneity and their impact on *in vivo* therapeutic efficacy. Sinek *et al.* (J. Sinek, *et al.*, 2004) had earlier performed a similar investigation; however, the pharmacokinetics and pharmacodynamics (PKPD) component was rudimentary, assuming one homogenous lesion compartment and not based upon externally acquired parameter values. In the present work we implement an extensive multi-compartment PKPD component whose parameter values are calibrated *via* published experimental data. This enables a comparison of the tissue- and cell-level drug dynamics of the two drugs, and facilitates the generation of hypotheses to explain their *in vivo* characteristics. We ask that the reader consider that if doxorubicin and cisplatin were discovered only today, the simulations herein could be seen as providing great insight into their *in vivo* performance, potentially streamlining and reducing costs of development and clinical trials, and of assisting in clinical therapeutic strategy to improve patient comfort and survival.

MATERIALS AND METHODS

Model Description

The multi-scale tumor growth and angiogenesis simulator developed by Zheng *et al.*, (X. Zheng, *et al.*, 2005) is used to grow the lesions upon which we simulate chemotherapy. The simulation field incorporates three phases: viable cancerous tissue, normal host tissue, and necrotic debris. The lesion/host interface is demarked by thick black contours, while the microvasculature appears as thin red curves. Dark regions are necrotic debris. Briefly, nutrient (oxygen and glucose) is provided through the discrete microvasculature, which is generated in response to angiogenic regulators produced from perinecrotic cells. This results in proliferation and tumor growth. The simple steady-state diffusion equation

$$0 = k_v(1 - n)^{TM} + D_n \nabla^2 n - k_n n \quad (1)$$

is used to model nutrient delivery and uptake, where n is the local nutrient normalized by the intravascular level, k_v is a measure of vascular porosity (0 is impermeable, ∞ is completely porous), δ is the Dirac delta function located along the vasculature, D_n is nutrient diffusivity, and k_n is the local rate of consumption by cells (X. Zheng, *et al.*, 2005). The characteristically high porosity of tumor vasculature implies a very high setting of k_v so that, essentially, vasculature provides a constant boundary condition of 1. Experiments given in (Mueller-Klieser, 1984) demonstrate that oxygen penetrates approximately 150 μm into *in vitro* spheroids before falling to about 10 percent of serum level. At this point necrosis ensues. Combining this with a diffusivity D_n of around

60,000 $\mu\text{m}^2\text{min}^{-1}$ (Nugent & Jain, 1984a, 1984b, 1984c; Swabb, Wei, & Gullino, 1974),

the nutrient uptake rate is calculated to be $k_n = 24 \text{ min}^{-1}$. Waste resulting from necrotic cell degradation is assumed to be removed *via* convection towards and through the tumor-host interface as well as *via* scavenger cell phagocytosis. In regions where nutrient is sufficient to maintain viability, mitosis is assumed to be directly proportional to its concentration, with the proportionality constant dependent upon the average cell cycle time of the malignant population.

Once the tumors are grown, drug administration *via* the vasculature is simulated by our multi-compartment pharmacokinetics model, based upon earlier work of (Demant & Friche, 1998; Dordal, *et al.*, 1995; El-Kareh & Secomb, 2003, 2005). For cisplatin, there are three compartments corresponding to (1) extracellular, (2) cytosolic, and (3) DNA-bound drug. For doxorubicin, there is a fourth compartment corresponding to intracellular organelles, *e.g.*, lysosomes. The system of equations governing transport for both drugs is

$$\begin{aligned}
\dot{s}_1 &= k_v (s_v - s_1)^{TM} + D_s \nabla^2 s_1 - k'_1 s_1 + k'_2 (s_2 / V_C) \\
\dot{s}_2 &= k_{12} V_C s_1 - k_{21} s_2 + k_{32} s_3 - k_{23} s_2 (1 - s_3 / s_M) + k_{42} s_4 - k_{24} s_2 \\
\dot{s}_3 &= k_{23} s_2 (1 - s_3 / s_M) - k_{32} s_3 - k_3 s_3 \\
\dot{s}_4 &= k_{24} s_2 - k_{42} s_4
\end{aligned} \tag{2}$$

where s_i represents drug concentration in compartment i , k_{ij} represents a transfer rate from compartment i to j , and k_i represents a rate of permanent removal from compartment i . s_v is intravascular drug concentration during bolus, and s_M is a DNA saturation parameter relevant to doxorubicin. V_C is the volume of a cell (assumed spherical with diameter 10 μm , yielding $V_C = 520$ fL) and appears in the first equation to reconcile the dimensions of s_1 (μM) with the dimensions of all other compartments (fmoles/cell). k_v and TM are the same as in Eq. 1. The primed rates appearing in the first equation are related to their unprimed counterparts *via* a constant due to the fact that extracellular volume is only a fraction F of total tumor volume. Taking a baseline tumor density of $1.0\text{E}9$ cells/ml, a well-known representative value, in combination with the cell volume previously quoted results in an extracellular fraction of 0.48, also a reasonable value. Finally, D_s is the diffusivity of the drug through interstitial space.

Both cisplatin and doxorubicin pass through cell membrane according to k_{12} (which includes possible pump and transporter activity, as do all other rates). From there, the drugs may efflux according to k_{21} or may bind to DNA according to k_{23} . The kinetics differ from here for the two drugs. Cisplatin may be removed according to the rate k_3 ,

which destroys the functioning of the drug and repairs the DNA (D. Wang & Lippard, 2005). Doxorubicin, however, has an off rate given by k_{32} , and moreover may be sequestered and released by lysosomes according to k_{24} and k_{42} (Arancia, *et al.*, 1998; Hurwitz, *et al.*, 1997; Rizzo, Sacchi, & Menozzi, 1989). Although lysosomal flow to membrane and exocytosis of sequestered drug plays a role in some drug resistant cell lines, we are not necessarily modeling drug resistance *via* this function, and so assume this process to be negligible, a valid assumption (Demant & Friche, 1998). On the other hand, we *are* concerned with the quantity of drug lysosomes can sequester, as this contributes to the cellular uptake of drug, and hence, its penetration characteristics.

The pharmacodynamics model consists of the Hill-type equation along the lines of those employed in (El-Kareh & Secomb, 2003, 2005)

$$E = \frac{N(n)}{1 + A^{-1}x^{-m}} \quad (3)$$

where E is cell inhibition (1 minus surviving fraction), x is DNA-bound drug-time product (area under the curve, or AUC), and A and m are phenomenologically fit parameters. $N(n)$ is a function of nutrient n ranging from 0 to 1 used to mimic the effect of hypoxia and hypoglycemia. Results with doxorubicin show that cells in deeper layers of spheroids do not respond as well to drug as do cells on the surface, even when intracellular drug levels are taken into account (Durand, 1986, 1990). Other experiments demonstrate reduced response in monolayer when cells are forced into quiescence due to reduced oxygen (Siu, *et al.*, 1999). Still others show that hypoglycemia can deplete topoisomerase II, thus reducing the effect of some anthracyclines (Shen, *et al.*, 1989). These results imply that the response of cells to doxorubicin *in vivo* might correlate to the local nutrient, which we herein refer to as the “nutrient effect.” For our purposes, the

exact form of N is not important. For simplicity, we choose $N = n^p$, where p is a phenomenological parameter derived from the data of (Durand, 1990), and equals 0.4. Since in our model n is normalized with respect to the intravascular level, it runs from 0 to 1, and thus so does N . Furthermore, at full nutrient levels, $N = 1$, and so cell inhibition (Padera, *et al.*, 2004) is maximal. In our simulations, drug pharmacokinetics (Eq. 2) is allowed to proceed from bolus initiation to washout 20 hours later. During this time the locally varying DNA-bound AUC is calculated and used to find cell inhibition (Eq. 3).

Parameter	Description	Baseline Value	
		Dox	Cis
V_c	Cell volume (fL cell ⁻¹)	520	520
ρ	Cell density (cells mL ⁻¹)	1.0E9	1.0E9
F	Interstitial fraction	0.48	0.48
D_n	Nutrient/ECM diffusivity ($\mu\text{m}^2 \text{min}^{-1}$)	60E3	60E3
D_s	Drug/ECM diffusivity ($\mu\text{m}^2 \text{min}^{-1}$)	1.0E3	30E3
k_n	Nutrient metabolism (min^{-1})	24	24
k_{12}	Drug uptake (min^{-1})	5.40	0.054
k_{21}	Drug efflux (min^{-1})	5.40	1.56E-3
k_{23}	Drug-DNA binding (min^{-1})	8.02E5	3.82E-4
k_{32}	Drug-DNA release (min^{-1})	1.80E3	0.0
k_3	Drug-DNA repair (min^{-1})	0.0	0.015
k_{24}	Lysosomal sequestration (min^{-1})	10.0	0.0
k_{42}	Lysosomal release (min^{-1})	0.07	0.0
s_m	Drug-DNA capacity (fmole)	1.00	∞
A	Phenomenological PD parameter	0.188	7.75
m	Phenomenological PD parameter	1.14	1.58
p	Nutrient effect parameter	0.4	0.0

Table 5.1: A complete summary of baseline pharmacokinetics and pharmacodynamics parameters along with references. Tumor growth and angiogenesis parameters can be found in (X. Zheng, *et al.*, 2005)

Pharmacokinetics Model Parameters

It will be useful to bear in mind that a generally acceptable theoretical setup for performing experiments to measure compartmental concentrations (and therefore to derive the rate constants we are after) is either a suspension or monolayer in an inexhaustible drug-laden medium corresponding to s_1 . Under these conditions, the relevant model consists of the last three equations in Eq. 2, with s_1 held constant. We will refer to this model as the modified version of Eq. 2. All model parameters and values are summarized in Table 5.1 along with references. These will be referred to as the baseline values, some of which will be adjusted later to simulate different tumor characteristics and therapeutic treatments. We emphasize that parameter values, having been derived from a variety of published experimental data spanning many years and cell types, correspond to a prototypical tumor and cancer cells suitable for the simulations herein, but not necessarily representative of any particular clinical specimen.

We begin with cisplatin, setting k_{24} and k_{42} to 0 since we assume only three compartments, and k_{32} to 0 since we assume the repair rate k_3 is the dominant removal rate of DNA-bound drug. k_3 is next obtained as follows. In experiments performed by Sadowitz *et al.* (Sadowitz, *et al.*, 2002), adducts per million nucleotides on isolated peripheral blood mononuclear cell DNA fell from 75 to 5 and 185 to 40 in two hours in two different experiments. Thus, assuming the exponential repair model $\lambda_3 = k_3 s_3$ we calculate the repair rate to be about 0.015 min^{-1} . An initial estimate of k_{23} is then made as follows. Sadowitz *et al.* (Sadowitz, *et al.*, 2002) shows that for $7 \text{ } \mu\text{M}$ cis, in two hours peripheral blood mononuclear cells accumulate from about 25 (non-thiol-blocked cells) to 175 (thiol-blocked cells) adducts per million nucleotides. Assuming that DNA consists of about 1.25 E6 kbp , this converts to from 1.04 E-19 to 7.27 E-19 moles of Pt docked on

the DNA (1 atom/adduct). Neglecting the cell membrane and supposing the DNA to be exposed directly to the drug, we have the ODE $\dot{s}_3 = \gamma_{23} - k_3 s_3$, where γ_{23} is a clearance parameter ($\text{fL}\cdot\text{min}^{-1}$). The solution of this is $s_3 = \gamma_{23} / k_3 (1 - e^{-k_3 t})$. Substituting values of $k_3 = 0.015 \text{ min}^{-1}$, $t = 120 \text{ min}$, and $1.04 \text{ E-4 } s_3 = 7.27 \text{ E-4 fmole}$ yields $0.27 \text{ } \gamma_{23} = 1.9 \text{ fL}\cdot\text{min}^{-1}$. To convert this to a rate we use the relation $k_{23} = \gamma_{23} / V_C$, arriving at $3.82 \text{ E-4 min}^{-1}$. The assumption that DNA was exposed directly to the cis solution means that this rate is only a bootstrap approximation and must be refined. At this point we note that the extremely low ratio of adducts per kbp implies that the saturation capacity of DNA with respect to cisplatin is never approached, and so set s_M to ∞ .

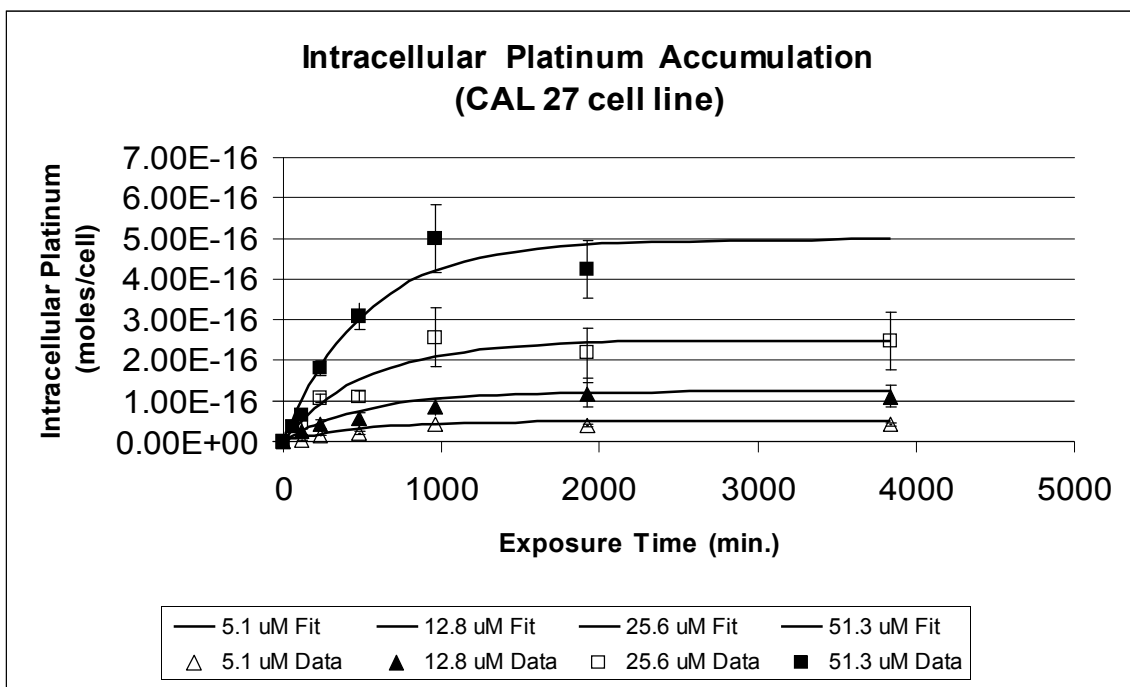


Figure 5.1: Data from (Troger, *et al.*, 1992) used to calibrate k_{12} , k_{21} , and k_{23} for the cisplatin model.

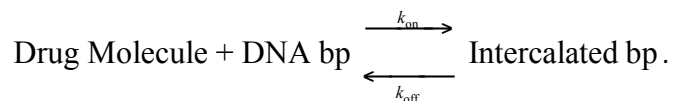
Parameters are fit simultaneously to all four curves treated as one set of data; they are not different for each curve.

Next, we estimate k_{12} and k_{21} . While doing this we will refine our initial estimate of k_{23} . The whole procedure involves fitting the best curves to data from (Troger, *et al.*, 1992) (Figure 5.1). Troger exposed CAL-27 cells in monolayer to four different concentrations of cisplatin and then measured the total intracellular amount of Pt at selected times. This corresponds to s_2+s_3 in our model. Beginning with the previous estimate of k_{23} and setting s_1 to the concentrations used by Troger, we adjust k_{12} and k_{21} in the modified version of Eq. 2 until a good fit of Troger's data is obtained. Simultaneously, we adjust k_{23} to keep the DNA-bound drug true to the results of Sadowitz *et al.* (Sadowitz, *et al.*, 2002) previously discussed. We remark that the disparity between the inward and outward rates derived for cisplatin may be due in part to carrier-mediated transport, *e.g.*, the CTR1 influx transporter.

Proceeding to doxorubicin we first obtain an acceptable range for k_{12} and k_{21} from the literature. For a variety of anthracyclines, including doxorubicin, initial estimates of cell membrane permeability P are taken from experiments with SU-4 and SU-4R wildtype and resistant human lymphoma cells (Dordal, *et al.*, 1995), from experiments with EHR2 and EHR2/DNR+ wildtype and resistant Ehrlich ascites tumor cells (Demant & Friche, 1998), and from experiments with MDA-468 breast cancer cells (Lankelma, *et al.*, 2000). The range reported is $2.4 \times 10^{-4} \text{ m} \cdot \text{min}^{-1}$. The relation $k_{12} = PA_C/V_C$, where A_C represents the cell membrane area, can then be used to arrive at an initial range of $1.4 \times 10^{-4} \text{ min}^{-1}$, which will be refined later. In the case of passive diffusion, $k_{21} = k_{12}$. We note that these values are far larger than those obtained for cisplatin previously. More generally, it has been remarked that cell membrane permeability for cis is much lower than for doxorubicin, etoposide, and vinblastine, although all four drugs are thought to enter cells by passive diffusion (Jekunen, *et al.*, 1993).

We next turn our attention to DNA-binding affinity. Given the great DNA affinity of the anthracyclines, saturability of the DNA must be taken into account, requiring an estimate of s_M . A difficulty arises since there is evidence a typical anthracycline molecule intercalation occludes from 3 to 10 binding sites in a manner that cannot be corrected exactly by a factor (McGhee & von Hippel, 1974; Rizzo, *et al.*, 1989; Tarasiuk, Frezard, Garnier-Suillerot, & Gattegno, 1989). To a first approximation, we assume that such a correction can be applied. Demant & Friche (Demant & Friche, 1998) report a DNA binding site concentration of about 5 mM within a cell volume of 1000 fL, yielding 5 fmoles of sites. A low value of 0.7 fmoles is obtained by using our assumed value of 1.25 E6 kbp and the reported site exclusion parameter of about 3 from Rizzo *et al.*, (Rizzo, *et al.*, 1989). Tarasiuk *et al.*, (Tarasiuk, *et al.*, 1989) find that the DNA of human lymphocytes is comprised of about 6.0 E6 kbp and that one intercalating molecule of doxorubicin requires 10 base pairs. Thus, Tarasiuk's data implies a factor-corrected quantity of 1 fmole of binding sites, which we take as a representative value of s_M .

DNA binding kinetics of the anthracyclines is nontrivial, perhaps requiring multiple steps and demonstrating sequence specificity (Qu, Wan, Becker, Zhong, & Zewail, 2001; Rizzo, *et al.*, 1989). Bearing this in mind, as an approximation it will suffice to assume non-specific, one-step binding and unbinding according to the chemical reaction



A representative value for the binding coefficient in the above equation for doxorubicin is reported as $k_{\text{on}} = 4.2 \text{ E8 M}^{-1} \text{ min}^{-1}$ and a value of the unbinding coefficient as $k_{\text{off}} = 1800 \text{ min}^{-1}$ (Rizzo, *et al.*, 1989). k_{off} is identical with k_{32} . From k_{on} we calculate a clearance

parameter (as with cisplatin) given as $\lfloor_{23} = k_{\text{on}S_M} = 4.2 \text{ E8 fL}\cdot\text{min}^{-1}$ (being cautious with the scales of our dimensions). k_{23} can then be calculated as \lfloor_{23}/V_C , given in Table 5.1.

We next turn our attention to the rates k_{24} and k_{42} governing lysosomal sequestration. Experiments of Hurwitz *et al.* (Hurwitz, *et al.*, 1997) using U-937 myeloid leukemia cells and their dox-resistant variant U-A10 show that the ratio of DNA-bound to lysosomally-sequestered drug is about 3 (Hurwitz uses daunorubicin, an anthracycline related to doxorubicin) In our modified model equations with all other parameters set as described above, the amount of sequestered drug at equilibrium is dependent only upon the ratio k_{24}/k_{42} . This ratio furthermore does not affect the equilibrium quantity of DNA-bound drug. Arbitrarily selecting $k_{24} = 1 \text{ min}^{-1}$, we find that the appropriate DNA-bound to lysosomally-sequestered ratio is obtained by setting k_{42} to 0.007. Considering that lysosomal membrane permeability is quite high (Demant & Friche, 1998), the lysosomally-bound drug must achieve equilibrium quickly, which can be modified by changing k_{24} while keeping the ratio k_{24}/k_{42} constant. We find that increasing k_{24} by a factor of 10 reduces the time required to achieve 95% of equilibrium value (max_{95}) to about 300 minutes, below which further increases in k_{24} only reduce this time negligibly. Thus, we conservatively set $k_{24} = 10$ and $k_{42} = 0.07$.

To refine our initial range of k_{12} and k_{21} , we use the modified PK model to compare our simulated monolayer uptake profiles of total intracellular drug with those of DeGregorio *et al.* (DeGregorio, Lui, Macher, & Wilbur, 1984) using human Ewing's sarcoma and rhabdomyosarcoma cells. At 5.40 min^{-1} both equilibrium values and uptake rates compare favorably at three test concentrations.

The last pharmacokinetics parameter values needed are the diffusivities D_s of cisplatin and doxorubicin through tumor interstitium. For molecules of their size (dox M.W. = 544, cis M.W. = 300), diffusivity should be about $30,000 \text{ }\mu\text{m}^2\cdot\text{min}^{-1}$ (Nugent

&

Jain, 1984c; Swabb, *et al.*, 1974). However, doxorubicin faces particularly severe barriers due to its binding to extracellular constituents such as hyaluronic acid (Kohno, Ohnuma, Kaneko, & Holland, 1988; Kohno, *et al.*, 1994), and its diffusivity in some tissues has been estimated to be as low as $1000 \text{ } \mu\text{m}^2\text{-min}^{-1}$, which we take as our baseline value

(Lankelma, *et al.*, 2000).

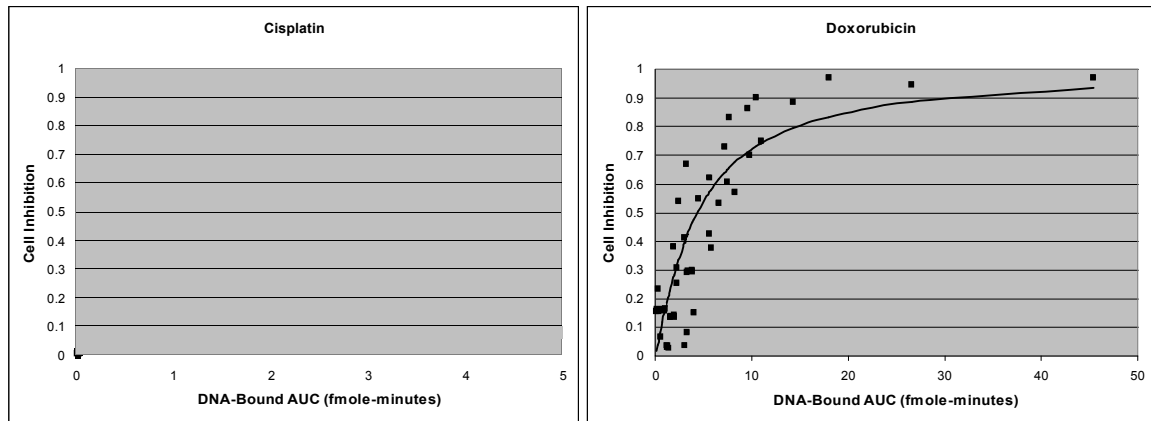


Figure 5.2: Cell inhibition fits for Equation 3 using data from (Levasseur, *et al.*, 1998) on A2780 ovarian cancer cells exposed in monolayer.

Pharmacodynamics Model Parameters

In order to calibrate the pharmacodynamics model (Eq. 3), we use the *in vitro* data of Levasseur *et al.*, (Levasseur, Slocum, Rustum, & Greco, 1998) with A2780 ovarian cancer cells exposed in monolayer to both doxorubicin and cisplatin over a range of times and concentrations. We assume the previously discussed modified pharmacokinetics model along with the values derived, and simulate Levasseur's exposures followed by approximately 24 hours of drug washout in drug-free medium (s_1 is set to 0). During this time, DNA-bound AUC is calculated. These data are then used in conjunction with Levasseur's surviving fraction data to fit the parameters A and m in Eq. 3 using Excel

(Figure 5.2). During this process, nutrient is assumed plentiful ($n = 1$) thus bypassing the nutrient effect.

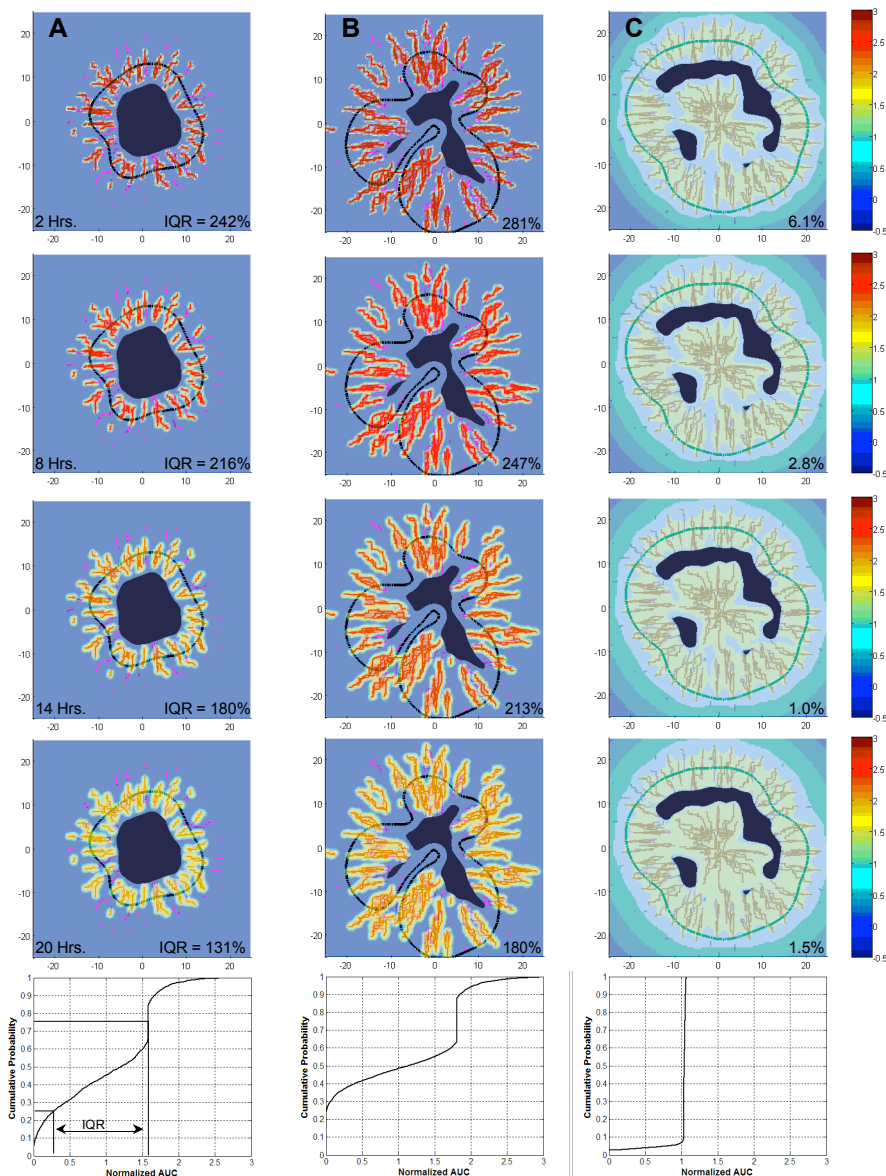


Figure 5.3: DNA-bound AUC at four times (rows: 2, 8, 14, and 20 hours) post bolus initiation for three two-dimensional simulated baseline tumor lesions (columns).

Results are normalized to average lesion AUC at the time taken to enable comparison of distribution heterogeneities. Thick black contours are tumor boundaries. Thin red curves are vasculature. Dark regions

are necrotic areas. Each unit represent $200 \mu\text{m}$. Bottom probability distributions show final AUC distributions at 20 hours. A concise measure of heterogeneity is given by the inter-quartile range (IQR), depicted in the lower left graph and explained in the text. Although AUC in host tissue is also shown in plots, analysis considers only DNA-bound drug in viable lesion.

***In Silico* Experiments**

We first grow three vascularized two-dimensional *in silico* lesions, shown in Figure 5.3. Each lesion is produced based upon the same set of growth and vasculature control parameters (See Zheng *et al.* (X. Zheng, *et al.*, 2005) for a complete description of these), but randomness in the vasculogenesis algorithm and slightly different initial shapes produce different morphologies. We then perform simulations to demonstrate and analyze distributions of DNA-bound drug AUC, nutrient, and cell inhibition resulting from intravascular bolus administrations of cisplatin and doxorubicin. Each experiment is replicated in each of the three lesions. In each case we hold the intravascular concentration of drug (s_v in Eq. 2) constant for two hours, then set it to zero for eighteen more hours to allow washout. Although this sharp “square wave” is perhaps a caricature of clinical bolus administration, it allows for consistent analysis and comparison of results. Intravascular concentrations are calibrated in each case to produce a total cellular growth inhibition of 50 percent (IC_{50} concentration). It is assumed that a true *in vivo* tumor does not grow or regress appreciably during the 20 hour course of the therapy we are attempting to simulate, hence we freeze tumor and vascular growth during our *in silico* therapies.

Our first set of experiments compares DNA-bound drug AUC distributions of doxorubicin and cisplatin under baseline conditions (see Table 5.1). We furthermore show the homogenizing effect of doxorubicin retention on final DNA-bound AUC (Durand, 1990). We next investigate the impact of inhibition heterogeneity on dosing requirements, paying particular attention to the nutrient effect for doxorubicin under

baseline conditions and improved penetration by, for example, removing hyaluronic acid (Kohno, *et al.*, 1988; Kohno, *et al.*, 1994). In our third set of simulations we more deeply investigate the effect of doxorubicin penetration therapies under three circumstances: baseline tumor density, high tumor density, and baseline tumor density with Pgp efflux activity. These are chosen because they demonstrate a spectrum of possibilities due to their effect on cellular drug uptake. High tumor density increases uptake, while Pgp efflux decreases it. In order to simulate increased penetration, we increase D_s for doxorubicin from its baseline value to $5000 \text{ } \mu\text{m}^2\text{-min}^{-1}$ for a moderate increase, and 30,000

for the maximum increase, thus matching the performance of cisplatin. To simulate high tumor density we increase ρ by 50 percent to $1.5 \text{ E9 cells-ml}^{-1}$. This has the effect of lowering the interstitial fraction F to 0.22, which in turn increases k'_{12} and k'_{21} while leaving all other rates unchanged. Pgp efflux is simulated by increasing k_{21} by a factor of 10, which has the effect of reducing all intracellular compartment concentrations by approximately the same factor. This is consistent with results of (Chen & Simon 2000) that show Pgp activity can reduce intracellular concentrations of daunorubicin (an anthracycline related to doxorubicin) by up to a factor of 100. In our fourth and final set of experiments we investigate permeabilization therapy with respect to cisplatin, whereby a detergent, such as digitonin, or electroporation is used to increase the permeability of cell membrane (Jekunen, *et al.*, 1993; Tanaka, *et al.*, 2001). We take an extreme case, increasing the rate constants k_{12} and k_{21} from baseline both by a factor of 100. Note that this does not increase the limiting intracellular or DNA-bound levels of drug attained in simulated monolayer, only the rate at which these come to equilibrium. Thus highly permeabilized, DNA-bound max_{95} is attained at 3.4 hours of exposure; further permeabilization reduces this negligibly. For comparison, max_{95} takes longer than

27 hours for unpermeabilized cells. This therapy is simulated under both *in vivo* baseline and very high cell densities achieved by increasing the baseline density 75 percent to $1.75E9$ cells- ml^{-1} . At this density, the interstitial fraction F drops to a mere 0.08. Both of these are further compared to monolayer results to probe the conditions under which *in vitro* assays can be used to predict clinical efficacy.

RESULTS

Although all treatments described in this section are duplicated in each of the three *in silico* tumors, we display only representative plots with appropriate summaries of all data. Note that the nutrient effect is only used where noted.

We begin by examining DNA-bound AUC distributions at various times in the baseline simulated lesions (each lesion corresponding to a column, *A*, *B*, or *C*), shown in Figure 5.3. From top to bottom, the times correspond to two hours, eight hours, fourteen hours, and twenty hours post bolus initiation. Levels are normalized relative to the average AUC within viable lesion for comparison of heterogeneity. Although surrounding host tissue cells uptake and bind with drug differently than cancer cells, we make no distinction in these color plots; however, quantitative analytical results only consider DNA-bound drug within viable lesion. The two left column sequences (Lesions *A* and *B*) show doxorubicin AUC, while the rightmost column shows cisplatin. For both Lesions *A* and *B*, at 2 hours doxorubicin AUC is seen to be about 3 times the average (dark red) close the vasculature, and almost 0 (blue) elsewhere. The distribution is only slightly more homogeneous by 8 hours. By 14 hours the heterogeneity has lessened, with the peaks close to the vasculature reaching only about 2.2. Finally, at the conclusion of washout 20 hours after bolus initiation, the distribution has become much more homogeneous, with the peaks only reaching about 1.7 times the average. In contrast, the

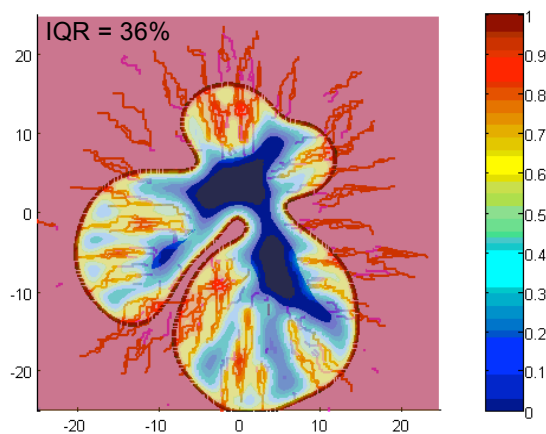
cisplatin distribution within Lesion *C* remains extremely homogeneous, right at the average, throughout the entire treatment.

The probability distributions at the bottom, corresponding to AUC at 20 hours post bolus initiation, allow for a more quantitative comparison. The two corresponding to doxorubicin show much heterogeneity relative to cisplatin on the right. Using the leftmost distribution as an example, the average DNA-bound AUC is found to be 6.04 fmole-min. 25 percent of tumor cells receive less than 1.66 fmole-min each, while 25 percent of tumor cells receive more than 9.54 fmole-min. The remaining 50 percent of the tumor cells receive between these two values, a range of 7.88 fmole-min. When normalized with respect to the average and expressed as a percent, this yields 131% (the interquartile range, or IQR), and gives a concise measure of distribution heterogeneity. IQR's are given at each of the other time points as well. All three tumors, despite varied lesion and vasculature morphologies, demonstrate similar results (not all shown). Doxorubicin AUC IQR's typically lessen from about 250 percent at 2 hours to 150 percent at 20 hours; cisplatin AUC IQR's drop from about 10 to 2 percent. Interestingly, in the run shown, the heterogeneity for cisplatin increases slightly in the last frame. This happens in some of the other cisplatin simulations as well.

We next investigate the impact of drug and nutrient heterogeneity on cell inhibition distributions and IC_{50} 's. Bolus administrations are simulated for cisplatin using the baseline lesions exactly as in Figure 5.3. The PD model (Eq. 3) is then used to calculate cell inhibition. For doxorubicin we use the baseline lesions as well as lesions in which drug penetration therapy is applied. The experiments for dox are run both with and without the nutrient effect.

A table of average IC_{50} 's and $\log(IC_{50}/IC_{50,mono})$'s for these experiments is given in Figure 5.4. Here and throughout this paper $IC_{50,mono}$ refers to *baseline* cells exposed in

monolayer and serves as a reference; IC_{50} refers to cells in lesion with parameters set to simulate particular conditions. Note that, as these are simulated monolayer exposures, $IC_{50,mono}$ is deterministic. Figure 5.4 also shows a typical nutrient profile, using Lesion *B* as an example with an IQR of 36%. This measurement is completely analogous to that used in Figure 5.3, except that here it is applied to the *nutrient* distribution and there is no normalization since nutrient levels are bounded absolutely from 0 to 100 percent, the level within the vasculature, itself. The nutrient IQR's for the other two lesions are within 2% of this value.



Drug	$IC_{50,mono} (\propto M)$	$IC_{50} (\propto M)$	$\log(IC_{50}/IC_{50,mono})$
Dox Baseline Nut. Eff. Off	0.175	0.482 ± 0.163	0.424 ± 0.138 $*p < 0.05$
Dox Baseline Nut. Eff. On	0.175	1.34 ± 0.874	0.830 ± 0.261 $*p < 0.05$
Dox w/ Penetration Nut. Eff Off	0.175	0.197 ± 0.0172	0.0511 ± 0.0371 $p > 0.05$
Dox w/ Penetration Nut. Eff. On	0.175	0.371 ± 0.0356	0.325 ± 0.0407 $*p < 0.05$
Cis Baseline	7.05	7.14 ± 0.0757	0.00529 ± 0.00462 $p > 0.05$

Figure 5.4: Means \pm SD's of the IC_{50} 's and the logs of their ratios with respect to monolayer treatments for experiments to investigate the impact of drug and nutrient heterogeneity.

$IC_{50,mono}$ is the IC_{50} of baseline cells in monolayer. At the 5% significance level using a one-tailed t -test, the average log ratio for cisplatin does not exceed 0. On the other hand, in three of the four experiments with doxorubicin, they do. Paired one-tailed t -tests show that the average log IC_{50} ratios for doxorubicin with the nutrient effect are greater than that without regardless of penetration therapy. Contour plot shows nutrient distribution in Lesion B demonstrating significant heterogeneity. Other lesions are similar.

At the 5% significance level, one-tailed t -tests show that the average log IC_{50} ratio is not greater than 0 for cisplatin, underscoring the homogeneity of its distribution. In contrast, out of the four experiments performed for doxorubicin from the combinations of nutrient effect and penetration therapy, three indicate that the average log ratios are greater than 0 at the 5% significance level. Within this group of four we can analyze the strength of the nutrient effect. For the baseline lesion, the nutrient effect increases the log IC_{50} ratio by 0.406 units (a factor of about 2.5). For the lesion with penetration therapy, the increase is 0.274 units (a factor of about 1.9). Paired t -tests show that these differences are significant at the 5% level.

The cell inhibition distributions closely mirror their AUC distributions, with that of cisplatin being virtually uniform at 50 percent inhibition throughout. Conversely, doxorubicin displays heterogeneity, increased with the addition of the nutrient effect. Using Lesion B as a representative example for doxorubicin, the upper block of frames in Figure 5.5 demonstrates the inhibition distributions for the baseline lesion with and without the nutrient effect. While the broadening of the cumulative probability plot as well as a comparison of the color distribution plots indicate that the nutrient effect increases heterogeneity, the inhibition IQR is *reduced* from 81 to 77 percent (again, not normalized). Employing penetration therapy in the lower block of frames, again we see increased heterogeneity in the plots with the nutrient effect, this time the IQR now also reflecting the increase. Lesions A and C yield similar results.

In our third set of simulations, we investigate the effect of therapies designed to improve doxorubicin penetration under several combinations of drug/interstitium

diffusivities, cell densities, and drug efflux activities (*e.g.*, Pgp). Figure 5.6 gives bar graphs of $\log(IC_{50}/IC_{50,mono})$'s for three scenarios. The leftmost triplet corresponds to baseline tumor density and no efflux, resulting in a condition of “normal” cellular uptake. The middle triplet corresponds to high density with no efflux, a condition of high uptake. The rightmost corresponds to baseline density with efflux, a condition of low uptake. In the baseline tumor case there is a change of -0.388 log units in going from no removal of hyaluronic acid to almost complete removal. When density is increased, the change becomes -0.709; however, when Pgp efflux is activated, ANOVA reveals there is no statistical difference, and in fact, the measured change is positive. Results are similar when the nutrient effect is included, with all bars essentially increased by a constant, approximately 0.37.

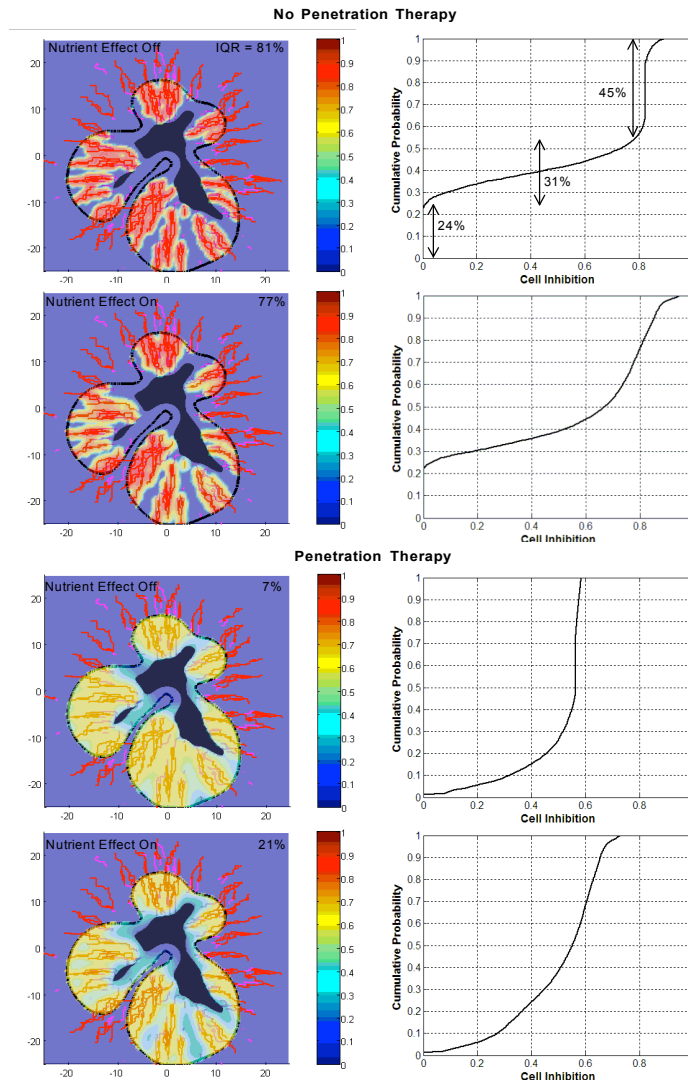


Figure 5.5: Top block shows the cell growth inhibition profile of Lesion B at baseline settings with and without the nutrient effect after bolus administration depicted in Figure 5.4.

Probability plot and IQR are now of inhibition distribution and are not normalized with respect to any average. Although the IQR indicates decreased heterogeneity with the nutrient effect, both the color plot and the probability plot indicate increased heterogeneity as is evidenced by the broadening of the curve. Bottom block repeats the same experiment, except with doxorubicin penetration increased. Now both the plots and IQR show increased heterogeneity. The appropriate IC_{50} is used in each experiment.

In our fourth and final set of simulations we investigate the effect of permeabilization therapy vis-à-vis cisplatin. Figure 5.7 shows $\log(IC_{50,perm}/IC_{50,unperm})$ for

three cases: monolayer, *in vivo* with baseline cell density, and *in vivo* with high cell density. Here, the subscripts “perm” and “unperm” denote the application or withholding of permeabilization therapy. Permeabilization results in a decrease of 0.154 log IC₅₀ units for simulated monolayers, *i.e.*, a reduction of IC₅₀ by a factor of 0.7, and is thus effective *in vitro*. An interesting question is whether this carries over *in vivo*, *i.e.*, whether a monolayer assay can be used to predict clinical efficacy. Improvements for the two *in vivo* simulations are comparable to monolayer results, with all three log-differences about -0.14, and no statistical difference for the baseline case at the 5% significance level using a two-tailed *t*-test.

DISCUSSION

According to our simulations, heterogeneities of drug and nutrient, caused in part by irregular vasculature and lesion morphology, exist and can significantly impact therapeutic results. Moreover, the sense and magnitude of their influence is not always intuitively obvious. A good example of this is that, despite its well-noted penetration difficulties, doxorubicin performs well clinically. Our simulations show that this may be somewhat explained by its retention in tissue removed from vasculature, causing homogeneity of exposure to increase long after the bolus has been terminated (Figure 5.3). This phenomenon has been experimentally verified in (Durand 1990) with spheroids. Because of this, the resulting cell inhibition distribution is more homogenous than would otherwise be expected. On the other hand, cisplatin maintains an homogenous DNA-bound distribution at all times from bolus initiation to 20 hours later, resulting in an extremely uniform cell inhibition distribution. This result, as well as the near equality of its IC₅₀ and IC_{50,mono} demonstrated in Figure 5.4, has also been experimentally verified

with spheroids (Durand, 1986; Inoue, Ohnuma, Holland, & Wasserman, 1985; Kohno, *et al.*, 1988).

While retention in tissue contributes to the performance of doxorubicin *in vivo*, Figure 5.4 demonstrates that its heterogeneity of distribution contributes to increased serum drug concentrations to match the same cell inhibition in monolayer. In one case, the average amount of drug increases by nearly one log unit. It is reasonable to expect that heterogeneity of nutrient, resulting in hypoxia and hypoglycemia, should compound this problem for doxorubicin. Indeed, this is the case as can be seen by the approximate doubling of the IC_{50} 's (0.482 μ M vs. 1.34 and 0.197 vs. 0.371) when the nutrient effect is applied. By graphically and quantitatively showing corresponding cell inhibition

distributions Figure 5.5 offers further insight into these phenomena. It is easily seen that cell inhibition distributions are as heterogeneous as their corresponding DNA-bound AUC distributions, with areas of lesion removed from vasculature experiencing reduced cell inhibition. An examination of the upper probability plot corresponding to a baseline lesion with no nutrient effect shows that a full 24% of viable lesion undergoes no inhibition at all. It is clear from these graphs that penetration therapy greatly decreases heterogeneity of cell inhibition (and commensurately, IC_{50}) as does removal of the nutrient effect. This latter is clinically feasible through, for example, carbogen breathing or recombinant human Epo (rHuEPO) administration (Shannon, Bouchier-Hayes, Condron, & Toomey, 2003; Teicher, Holden, & Jacobs, 1987).

One puzzling behavior is that while both the color and probability plots demonstrate consistently increased heterogeneity brought about by the nutrient effect (as is evidenced by the broadening of the probability curves), the IQR actually decreases in the baseline case from 81 to 77%. This occurs with Lesions *A* and *C* as well. A solution

to the mystery is obtained by noticing that a large portion of the tumor either experiences no inhibition (about 24%) or an already heterogeneous inhibition (about 31%). The remaining 45% receives a near homogeneous level of inhibition (the vertical portion of the curve), and this fraction corresponds to tissue close to the vasculature. Thus, the only significant heterogeneity that can be induced by the nutrient effect is within this fraction. Indeed, it is just this part of the curve that broadens in the second probability distribution, indicating greater heterogeneity, as expected. As the IQR is designed to measure heterogeneity somewhat more globally, it misses—in fact, misdiagnoses—the change occurring within this fraction.

In addition to dosing requirements, there is a second and subtler reason to consider heterogeneity of the inhibition distribution when treating clinical tumors. Heterogeneities in growth and regression have been linked to increased lesion fragmentation and invasiveness (Cristini, *et al.*, 2005; H. B. Frieboes, *et al.*, 2006; Kunkel, *et al.*, 2001b; Pennacchietti, *et al.*, 2003). While the mechanisms underlying this phenomenon are complex, involving myriad protein signaling events and activities at the cellular level, they may at least partly rely on gross lesion effects.

Figure 5.6 shows what might be expected from therapies that increase doxorubicin penetration by, for example, removing hyaluronic acid. As expected, for the baseline tumors, greater homogeneity and level of AUC is achieved, resulting in reductions of IC_{50} . This effect has been experimentally verified using spheroids (Kohno, *et al.*, 1988; Kohno, *et al.*, 1994). That it should be more pronounced for high-density tumors and completely absent in the presence of Pgp efflux is intriguing. A potential explanation is availed by simplifying the pharmacokinetics model (Eq. 2), reducing it to the one-dimensional, one-compartment steady state diffusion equation $0 = D\nabla^2 s - ks$ with diffusivity D and uptake rate k . In two dimensions, a segment of blood vessel acting as a source next to a section of tissue approximates the one-dimensional case. This equation

has one governing parameter, the characteristic diffusion length $L = \sqrt{D/k}$, and a (non-unique) solution $s_v \exp(-x/L)$, where x is distance from the source and s_v is the constant level of drug in the vasculature. Considering a section of tissue of thickness d next to a vessel, average AUC is proportional to the integral, $\int_0^d s_v \exp(-x/L) dx = s_v L(1 - \exp(-d/L))$. Let AUC_{50} be the fixed average AUC required for fifty percent cell inhibition. Then, ignoring the constant of proportionality, $AUC_{50} = IC_{50} L(1 - \exp(-d/L))$. Increasing L by some factor $C > 1$ to simulate penetration therapy results in a new characteristic length of CL , and hence new IC_{50} . The ratio of IC_{50} 's is therefore $\frac{AUC_{50}}{CL(1 - \exp(-d/CL))} \cdot \frac{L(1 - \exp(-d/L))}{AUC_{50}} = \frac{(1 - \exp(-d/L))}{C(1 - \exp(-d/CL))}$, which for large characteristic lengths L approaches 1, and for small lengths approaches $1/C$. Now, increasing cell density has approximately the effect of increasing k , resulting in a small L , thus manifesting the differential in IC_{50} 's. Conversely, activating Pgp efflux has the effect of decreasing k , resulting in a large L , thus nullifying the differential. This could be an important point when deciding upon appropriate therapies for tumors exhibiting different characteristics such as efflux mechanisms and relatively high or low densities. Any therapy involves risk. There may be less to gain under certain conditions, advising that the therapy not be performed or perhaps, that concomitant therapy be performed designed to optimize conditions.

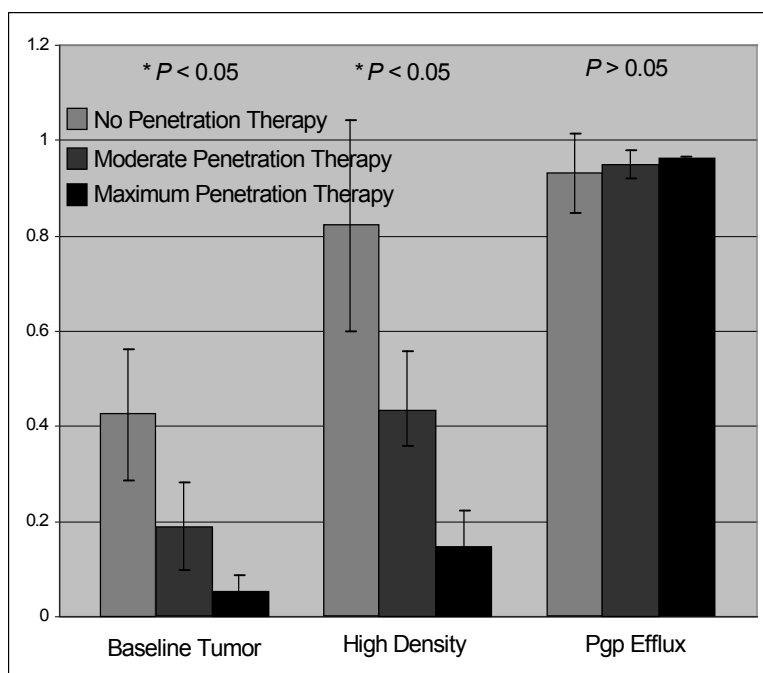


Figure 5.6: The effect of increasing doxorubicin penetration is shown in three cases: baseline tumor (excepting penetration therapy), high-density tumor, and tumor with Pgp efflux.

High density has the effect of increasing drug uptake, while Pgp efflux has the opposite effect. Bars are $\log(\text{IC}_{50}/\text{IC}_{50,\text{mono}})$, where $\text{IC}_{50,\text{mono}}$ is the IC_{50} for baseline cells exposed in monolayer. Results are similar for simulations carried out under the effects of hypoxia and hypoglycemia, with all bars approximately increased by a constant. Three replications per bar with results of ANOVA displayed.

The great homogeneity of both cisplatin AUC and cell inhibition demonstrated in Figures 5.3 and 5.4 indicate that *in vitro* assays using this drug may have relatively high positive predictive accuracy. While our simulations do not yield enough resolution to claim that the results herein answer this question, the outcome shown in Figure 5.7 is of interest. There is no statistical difference at the 0.01 significance level between improvement obtained by permeabilization therapy in the simulated tumors (even very dense ones) and monolayer. This compares favorably with the findings of (Tanaka, *et al.*, 2001) in which the improvement obtained *via* cisplatin permeabilization therapy *in vitro* is strongly reflected by the improvement in isolated lung perfusion in rats.

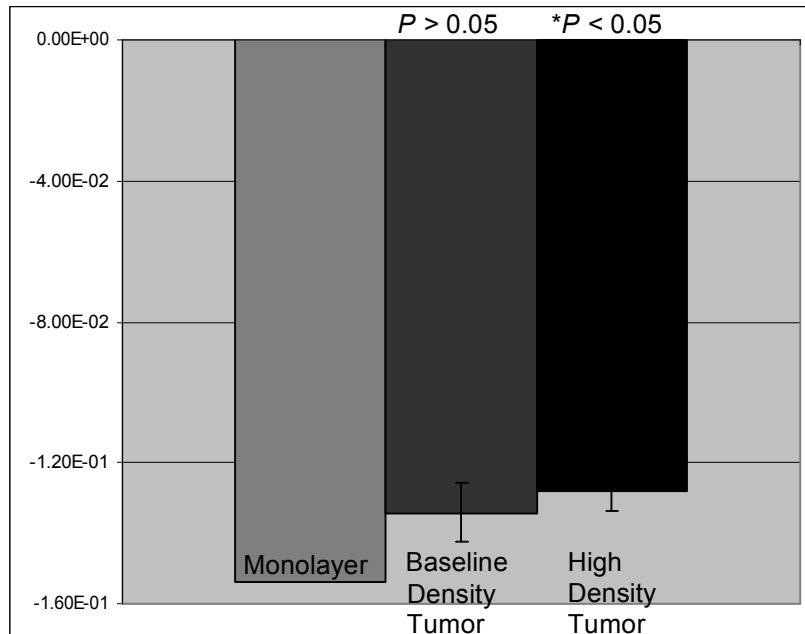


Figure 5.7: Effect of permeabilization therapy with respect to cisplatin is shown in three cases.

Bars are of $\log(\text{IC}_{50,\text{perm}}/\text{IC}_{50,\text{unperm}})$ where $\text{IC}_{50,\text{perm}}$ and $\text{IC}_{50,\text{unperm}}$ correspond to permeabilized and unpermeabilized conditions. Three replications per bar with results of two-tailed t -tests relative to monolayer displayed. While there is a statistical difference at the 0.05 significance level for the high-density tumor, this disappears at the 0.01 significance level.

While it can and should be argued that the simulations herein fail to account for some (many!) critical aspects of tumor growth and drug response (such as clonal heterogeneity, cell phase sensitivity, and signaling pathways) and that parameter settings may in some cases be inexact, it should not be concluded that these shortcomings invalidate characteristics the simulations have revealed. Indeed, we have correctly postdicted several results: that doxorubicin retention results in a more uniform AUC and cell inhibition than would otherwise be indicated by its penetration difficulties (Durand, 1990); that cisplatin achieves a highly uniform AUC, and its cellular inhibition *in vitro* can closely match that *in vivo* (Durand, 1986; Inoue, *et al.*, 1985; Kohno, *et al.*, 1988); and that improvement due to cisplatin permeabilization therapy *in vitro* can accurately

predict improvement *in vivo* (Tanaka, *et al.*, 2001). We have given compelling evidence that not only do macroscopic environmental conditions, namely, drug and nutrient distribution heterogeneity, greatly impact therapeutic efficacy, but also that the outcome of therapeutic strategies can depend upon them in nonlinear and *a priori* unpredictable ways. The results of our third set of experiments (penetration therapy) provide an example. In light of this, it is naive to solely devote attention to genetic factors at the expense of factors residing at coarser and more global scales.

One of our broader goals is to demonstrate how increasingly sophisticated *in silico* technology, driven by mathematical modeling and calibrated with experimental data, can and is being developed to provide an alternate investigative and clinical tool complementary to traditional methods (Bangs & Paterson, 2003; H.B. Frieboes, *et al.*, 2006; Sanga, Frieboes, Sinek, *et al.*, 2007; Sanga, *et al.*, 2006; J. Sinek, Frieboes, Sivaraman, Sanga, & Cristini, 2006). It can well be imagined that were doxorubicin and cisplatin discovered today, the *in vivo* simulations herein presented could be used to anticipate their lesion- and cellular-scale pharmacokinetics, helping to refine clinical trial design and lower costs. In clinical application, the results could be used to guide therapeutic strategy. For example, any risks associated with doxorubicin penetration therapy could be avoided if it were known that the patient's tumor were expressing Pgp or otherwise had lowered cellular uptake, according to the results given in Figure 5.6. The power of *in vitro* experimentation lies in its ease of implementation while remaining in the biological realm. By its very nature, *in vitro* experimentation attempts to refine and isolate. Yet, much of what happens *in vivo* is the result of a nonlinear system whose behavior is more than the sum of its parts. The power of *in silico* simulation lies in its ability to integrate components into a virtual system capable of reproducing such behavior, implicitly taking into account circuits of information flow difficult to explicitly

analyze. Accurately calibrated and rigorously validated, such an integrated model could provide a “dry-lab” to be used as a powerful complement to the traditional wet-lab in fundamental research, drug discovery, and the clinic. It could be used to probe scenarios and test hypotheses that are either difficult or impossible to instantiate in the body. Results could then suggest supportive *in vitro* and *in vivo* experimentation, the end result being new therapeutic targets or strategies. Simultaneously, weaknesses (or strengths!) of the *in silico* model could be uncovered and addressed. Computational models have the potential to facilitate an era of great discovery and progress in understanding and treating cancer, and providing new hope to its victims.

ACKNOWLEDGMENTS

The authors would like to thank Steve Wise (U.C. Irvine) for assistance with numerical methods and programming, Ardith El-Kareh (U. of Arizona) for insights into the cellular pharmacokinetics and pharmacodynamics of doxorubicin and cisplatin, Hermann Frieboes (U.C. Irvine) for helpful discussions regarding *in vitro* experimental protocols and drug response, and John Fruehauf (U.C. Irvine) for suggestions and corrections covering a broad spectrum of oncological knowledge.

Chapter 6: Gene Expression Meta-Analysis Supports Existence of Molecular Apocrine Breast Cancer with a Role for Androgen Receptor and Implies Interactions with ErbB Family

PREFACE

The advance of microarray gene expression platforms in recent decades has enabled cancer researchers to simultaneously probe the disease on a genome-wide basis. This has led to generation of massive amounts of gene expression data (much of which is publicly available), which are forming the basis for the discovery of molecular signatures and biomarkers having clinical value as prognostic indicators and predictors of patient response to therapy. While these empirically-based predictive models provide researchers with knowledge of the genes and proteins involved, they alone do not provide information about pathway relationships, which can help elucidate the mechanisms driving cancer phenotypes, facilitate therapy design, and provide structure for molecular signaling models in a multi-scale modeling framework.

Biomolecular network reconstruction is motivated by the desire to map molecular scale interactions to cellular and tissue level traits of interest (Bickel, *et al.*, 2009). The pharmaceutical and biotechnology industries have strong motivation to “reverse-engineer” molecular networks in order to rationally derive molecular targets (T. Chen, He, & Church, 1999; Hopkins & Groom, 2002; Markowitz & Spang, 2007). In the field of predictive oncology, reconstructing the maps of interacting molecules that ultimately drive cancer phenotypes is the first step in framing cell signaling models that may serve in a multi-scale cancer modeling framework. Advances in bioinformatics and computational systems biology analysis methods in combination with measurements of gene expression has provided researchers with tools to efficiently hypothesize signaling relationships that may be validated through experimentation.

Gene expression analysis has divided breast cancer into multiple molecular subclasses, with the major groups defined primarily by their estrogen receptor (ER), progesterone receptor (PR), and Her2/neu receptor (ErbB2) status (Perou, *et al.*, 2000). These subclasses have been shown to differ by clinical outcomes (Sorlie, *et al.*, 2001). The best prognosis categories are the luminal A and B groups, which are ER⁺ and ErbB2⁻. The two most aggressive subsets of breast cancer are the ErbB2⁺ subgroup, which overexpress ErbB2 (clinically referred to as Her2neu), and the ER⁻/PR⁻/ErbB2⁻ subgroup, termed the basal subtype in the original profiling studies. The ErbB2⁺ subgroup can be offered herceptin as an adjuvant molecularly targeted therapy, while the basal phenotype, clinically recognized as the triple negative phenotype by immunohistochemical assay for ER, PR and Her2neu, does not respond to either hormonal therapy or targeted therapy. Outside of clinical trials with anti-angiogenesis therapy, the primary option for patients with a triple-negative tumor is to receive chemotherapy and radiation. They have an expected three-year survival rate of 76.8% compared to 93.5% for patients with other phenotypes (Tischkowitz, *et al.*, 2007). The question arises-are there other distinguishable subsets of clinical significance present that might be distributed across these major groups or as a subset of one of them, and if so could a target be defined within their profiles that can be used for therapeutic purposes? Identifying more such subclasses at a more granular level will encourage the discovery of new molecular targets and prognostic biomarkers by more finely dividing breast cancer molecular phenotypes.

In the next chapter, we present a research study where we perform a meta-analysis of two independently-conducted breast cancer gene expression studies that claim to have identified another molecularly-defined subgroup of ER⁻ breast cancer. The first part of this study addresses the need to define criteria for comparing the molecular equivalence between independently conducted studies. In comparing the gene expression data from the two studies, we describe our experiences with data normalization, and propose a

normalization protocol for conducting gene expression meta-analyses. We continue to analyze the data using conventional bioinformatics methods, and derive a gene expression signature to predict this phenotype in the breast cancer population. Finally, we perform network reconstruction methods on the normalized data to postulate signaling relationships underlying this breast cancer phenotype, and show that we are able to derive signals that have already been identified experimentally in cell lines. Thus, we report that our data mining and computational systems biology approach is able to hypothesize signaling relationships to rationally guide experimentation. Furthermore, our pathways analysis implicates a strong role for the Androgen Receptor and related pathways in driving this phenotype, which should be considered in strategizing therapy for this patient population.

This research was a collaborative effort between Sandeep Sanga, Dr. Bradley M. Broom, Dr. Vittorio Cristini, and Dr. Mary E. Edgerton. Sanga carried out the data mining, normalization and analysis using the majority of the bioinformatics and computational systems biology methods referred to in the manuscript, with the exception of those performed by Broom. Sanga also participated in the study design, and took the lead role in drafting the manuscript. Broom carried out the Gene Shaving and Robust Bayesian Network methods. Cristini participated in the study design. Edgerton conceived of the study, directed its design and coordination, and helped to draft the manuscript.

Note: This chapter is based on a manuscript that has been accepted for publication in BMC Medical Genomics as (Sanga, Broom, Cristini, & Edgerton, In Press) and is currently in press.

ABSTRACT

Background: Pathway discovery from gene expression data can provide critical insight into the relationship between signaling networks and the biology of cancer. Oncogenic

signaling pathways are most often inferred by comparison with cell lines. We performed a meta-analysis of breast cancer data and demonstrate that we can discover signaling pathways and interactions by detecting patterns in the patient data without having to compare them to profiles from cell lines.

Results: We first demonstrate that two separately introduced ER⁻ breast cancer subsets represent the same tumor type. We combine the gene expression data and compare results of pathway analysis using two different network inference methods, LeFeminer and Backward Chaining Rule Induction. Both algorithms support a role for AR signaling in the data. Using a classifier built from this data, we identify more tumors with the same profile from published data. We apply Gene Shaving and Robust Bayesian Network Analysis and detect interactions between the AR pathway and members of the ErbB family.

Conclusions: Network inference analysis associates this ER⁻ breast cancer subset with AR signaling and demonstrates interactions between AR and both EGFR and ErbB2, implying that therapies targeting AR might be hampered if interactions with these ErbB family signals are not addressed. Data mining strategies provide an alternative method to comparison with cell lines for discovering seminal pathways and interactions between signaling networks.

INTRODUCTION

Gene expression array data can be mined to provide critical insight into our understanding of the relationship between signaling networks and the biology of cancer (Bild, *et al.*, 2006; Desmedt, *et al.*, 2008; Heiser, *et al.*, 2009). In addition to identifying individual pathways, recent attention has been given to “cross-talk” or interactions that cause aberrant signaling patterns in cancer (Citri & Yarden, 2006; Migliaccio, *et al.*, 2006; Naderi & Hughes-Davies, 2008). The conventional method of identifying

oncogenic pathways and their interactions has been through studying cell lines (Bild, *et al.*, 2006; Doane, *et al.*, 2006; Farmer, *et al.*, 2005; Heiser, *et al.*, 2009). Our goal is to be able to identify dominant pathways using data mining methods that do not require direct comparison with cell lines.

To pursue our goal we investigate a recently introduced subtype of ER⁻ breast cancer that is hypothesized to result from AR signaling. We analyze the data using several different bioinformatics approaches to pathway discovery. We are able to detect patterns that support the same conclusions reached with comparison to cell lines data by the original authors. In addition, we introduce interactions not previously discovered in the data that have important therapeutic implications. Thus, our results contribute to both bioinformatics and to breast cancer biology.

The ER⁻ breast cancer subtype that we study here has been termed the “molecular apocrine” subtype (Farmer, *et al.*, 2005; Weigelt, *et al.*, 2008) and the “ER⁻ class A” subtype (Doane, *et al.*, 2006) in two separate studies that proposed its existence. The studies were independently performed, but both groups hypothesized AR signaling as a defining feature of the transcript profile, leading us to question whether or not they represent the same tumor subset. One study identifies six of 16 ER⁻ tumors as the molecular apocrine subtype and the other study identifies ten of 41 ER⁻ as the class A tumors. Since there has not been a meta-analysis of both studies to actually confirm that the individual tumor clusters actually represent the same breast cancer subset as defined by gene expression, we start by performing a comparative study. We call this a test of “molecular equivalence,” and we propose a set of criteria for establishing molecular equivalence cancer subsets defined by gene expression data: 1) the majority of the molecular phenotype should cluster together and their combined profile should be distinct from the remaining samples in unsupervised clustering of the combined data; 2) there should be significant overlap of the gene signatures used to classify the phenotype from

each institution; and 3) a classifier trained on data from one institution should be able to predict the phenotype correctly in the other institution's data, and vice versa. In the process of establishing molecular equivalence, we test different methods of normalizing the data to remove institutional bias and we comment on their effectiveness.

Once having established the molecular equivalence of the group, we use Learner of Functional Enrichment algorithm (LeFEminer), which is based on gene set enrichment (Eichler, Reimers, Kane, & Weinstein, 2007), and Backward Chaining Rule Induction (BCRI), which is a de novo discovery method (Edgerton, *et al.*, 2007; Fisher D, *et al.*, 2005; Fisher DH, *et al.*, 2006) to identify pathways in the combined data. Both of these methods incorporate existing pathway knowledge from the literature within their methodology. Our results indicate a role for AR in this breast cancer subset. Subsequently, we use a gene expression classifier to identify more molecular apocrine data for discovery of pathway interactions. We use Gene Shaving and Robust Bayesian Network Analysis because it facilitates discovery of interactions that have variable prevalence in the patient population (Broom, *et al.*, in preparation; Hastie, *et al.*, 2000). We demonstrate that there are highly prevalent interactions between AR signaling and members of the ErbB family. We discuss the therapeutic implications of cross-talk between AR and members of the ErbB family in molecular apocrine type breast cancer. Taken together, these results demonstrate that data mining methods can be used to generate network information directly from gene expression data.

Data

The data used in this study were generated on Affymetrix U133A oligonucleotide microarrays and are publicly available (Doane, *et al.*, 2006; Farmer, *et al.*, 2005; Ivshina, *et al.*, 2006; Rouzier, *et al.*, 2005; Sotiriou, *et al.*, 2006). The cohort from Farmer *et al.* (Farmer, *et al.*, 2005) includes 22 ER⁺ breast carcinoma samples with six classified as

molecular apocrine. The cohort from Doane *et al.* (Doane, *et al.*, 2006) includes 41 ER⁻ breast carcinoma samples with ten classified as molecular apocrine. We refer to data generated by Farmer *et al.* and Doane *et al.* as the “index cohorts.” We use additional cohorts from Ivshina *et al.* (Ivshina, *et al.*, 2006), Rouzier *et al.* (Rouzier, *et al.*, 2005), and Sotiriou *et al.* (Sotiriou, *et al.*, 2006), which contain 59, 51, and 34 ER⁻ breast carcinoma samples, respectively, to confirm the existence of the molecular apocrine phenotype in larger cohorts outside the index cohorts and to explore gene network interactions.

RESULTS

Data Normalization

We combine the index cohorts into a single, homogeneous dataset with quantile normalization (QN) performed using the dChip software package (Bolstad, *et al.*, 2003; C. Li, 2008) followed by a recently published cross-study normalization scheme (XPN) that results in removal of persistent systematic bias and noise (Shabalin, *et al.*, 2008). Additionally, we use updated probeset definitions (Barnes, *et al.*, 2005; Carter, *et al.*, 2005; Dai, *et al.*, 2005; H. Liu, *et al.*, 2007). XPN brings the two gene expression datasets into better agreement as evidenced by improvements in the expected linear relationship of median probeset expression levels between the index cohorts after three sequential steps: 1) QN, 2) QN + XPN, and 3) QN + XPN with updated probeset definitions (Figure 6.1). The Pearson correlation coefficient for the three steps is 0.877, 0.923, and 0.913, respectively. We note that step 2 gives normalized data with a slightly higher coefficient than with using updated probeset definitions (step 3). However, we choose to follow recommendations in the bioinformatics literature to take advantage of the most up-to-date gene sequence information for grouping and mapping transcript-consistent probesets

(Barnes, *et al.*, 2005; Carter, *et al.*, 2005; Dai, *et al.*, 2005; H. Liu, *et al.*, 2007). As we proceed with our analysis, we compare our results using XPN with results we generate using median-centering, a conventional method for cross-study normalization of data performed on a single microarray platform.

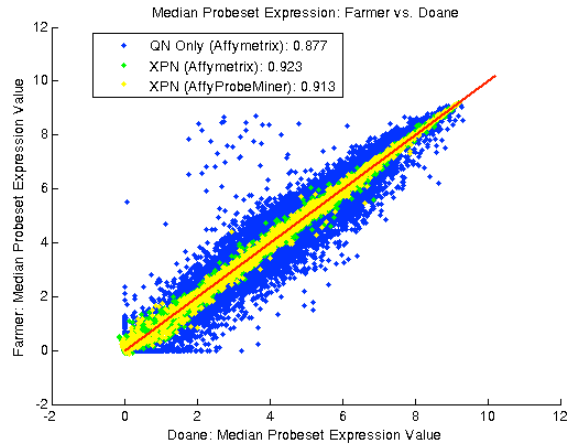


Figure 6.1: Scatter plots of Median Probeset Expression Values of Farmer *et al.* Data vs. Doane *et al.* Data.

The blue markers represent the data after quantile normalization (QN), the green markers represent the data after subsequent cross-study normalization by XPN, and the yellow markers represent the data after subsequent update of probeset definitions. The red line depicts the unity line, *i.e.* $y = x$. In the legend lists Pearson correlation coefficient for each normalization step. All data has been natural-log transformed.

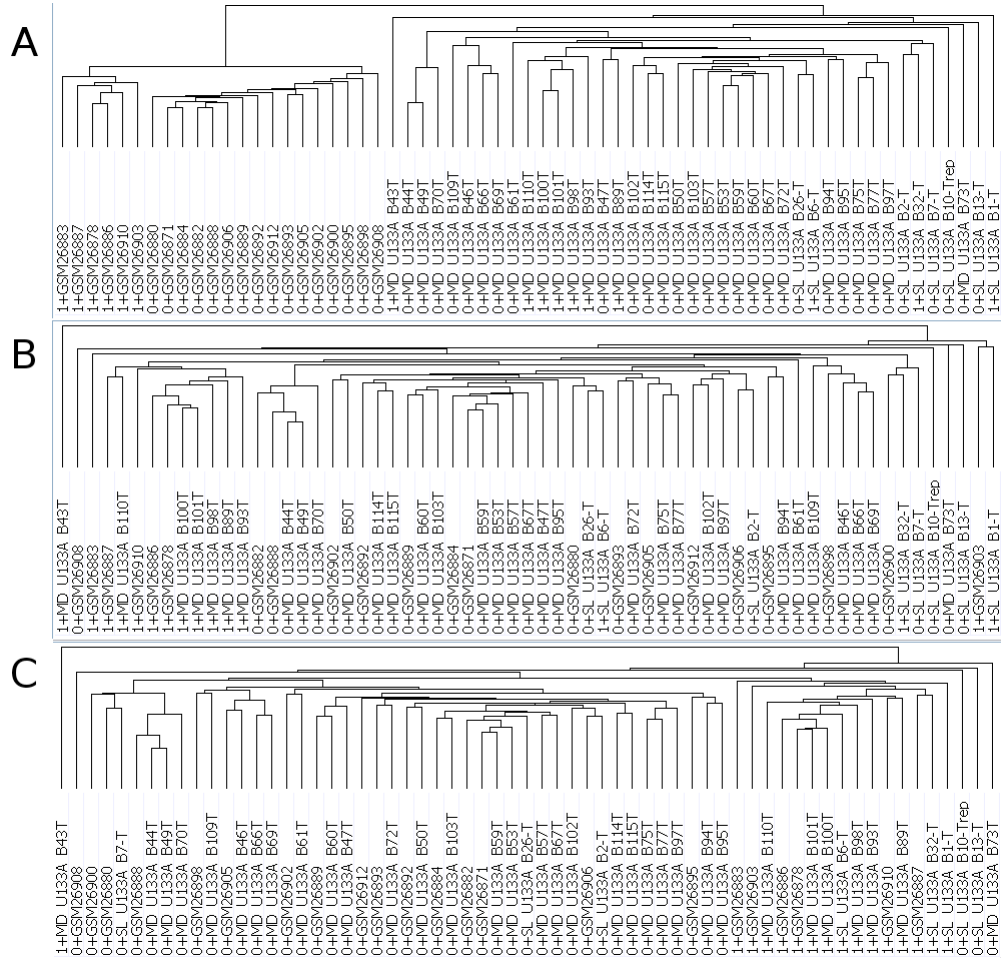


Figure 6.2: Hierarchical Clustering of Doane *et al.* and Farmer *et al.* Data.

(A) Hierarchical clustering of the quantile normalized, natural-log transformed data from dChip using original probes sequence information provided by Affymetrix. (B) Hierarchical clustering of the quantile normalized, XPN normalized, natural-log transformed data using original probe sequence information provided by Affymetrix. (C) Hierarchical clustering of the quantile normalized, XPN normalized, natural-log transformed data using updated probe sequence information provided by AffyProbeMiner. **Information:** Clustering was performed using the Pairwise-Average Linkage method and measures distance using Euclidian Distance. The samples cluster according to institution and not by their “molecular apocrine” or “non-molecular apocrine” phenotype as indicated by 1 and 0, respectively, preceding the sample ID.

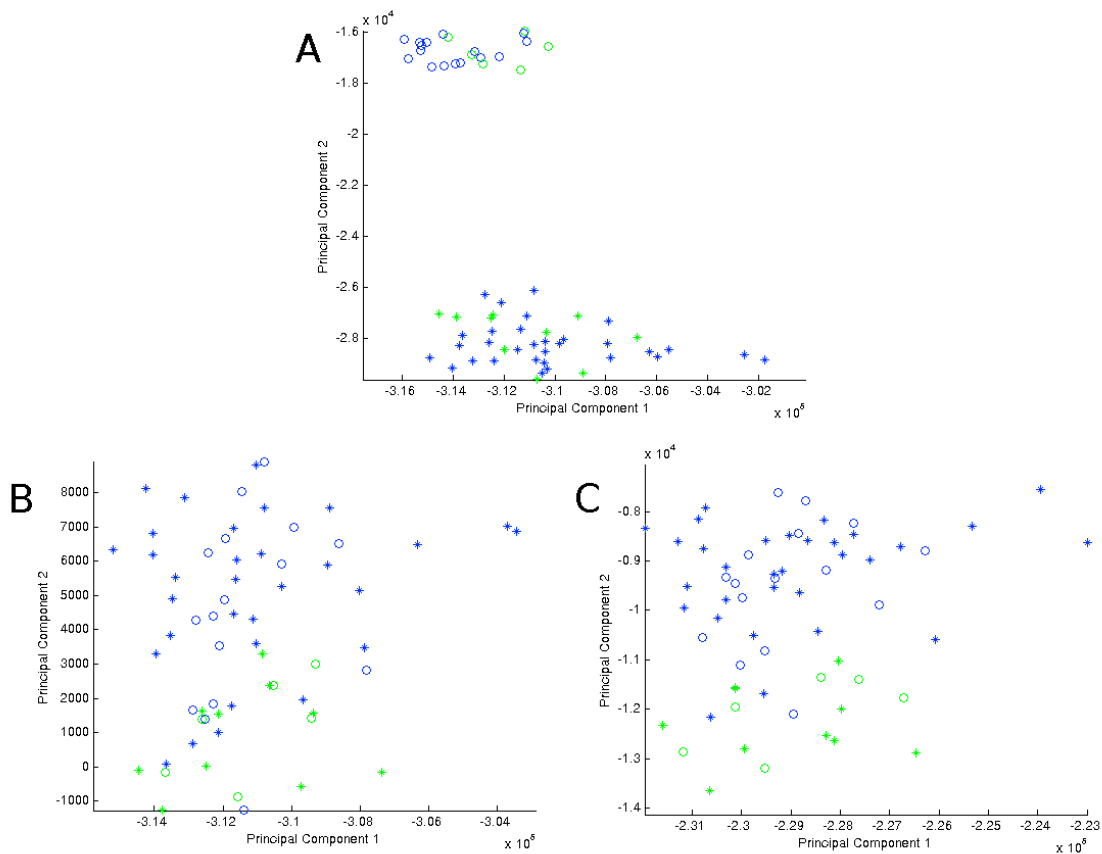


Figure 6.3: PCA of Combined Doane *et al.* and Farmer *et al.* Cohorts

(A) Principal Components Analysis (PCA) of the combined cohorts from Farmer *et al.* (Farmer, *et al.*, 2005) and Doane *et al.* (Doane, *et al.*, 2006) on natural-log scaled data quantile-normalized using the Affymetrix-provided chip definition file (CDF) before cross-study normalization and XPN; p-value = 0.369. (B) PCA of the combined cohorts on natural-log scaled data quantile normalized and XPN normalized using the Affymetrix-provided CDF; p-value < 0.0001. (C) PCA of the combined cohorts on natural-log scaled data quantile normalized and XPN normalized using the AffyProbeMiner-provided CDF (H. Liu, *et al.*, 2007); p-value < 0.0001. Qualitative and quantitative comparison of (A) – (C) systematically shows the improvement of bringing the data from multiple institutions into uniformity using quantile normalization, XPN, and updated probeset definitions. With each frame, the data becomes more well-defined by molecular phenotype rather than by institution, which is confirmed by the decreasing p-values. **Information:** The plots use the first two principal components. Circles represent the Doane *et al.* cohort and asterisks (*) represent the Farmer *et al.* cohort. The hypothesized “molecular apocrine” phenotype is represented by the color green and “non-molecular apocrine” phenotype by the color blue. The p-values are calculated using the Fasano & Franceschini statistical test with the null hypothesis that the two phenotypes cannot be differentiated on the basis of two-dimensional, principal components coordinates.

Comparison for Molecular Equivalence

When we perform XPN in addition to QN, we see significant improvement in the removal of systematic bias using both unsupervised hierarchical clustering (HC) and principal components analysis (PCA). We use routines in the GenePattern software package (Kuehn, Liberzon, Reich, & Mesirov, 2008; Reich, *et al.*, 2006) and separately we compute a p-value using the Fasano & Franceschini statistical test (Fasano & Franceschini, 1987). Figures 6.2A and 6.3A demonstrate that the data clusters by institution with QN alone. The addition of XPN to the normalization scheme (Figures 6.2B and 6.3B) tapers the institutional systematic bias and reveals a single molecular apocrine cluster. The dendrogram from the HC results shows 12 of 16 (75%) samples defined previously as molecular apocrine in a single cluster ($p < 0.0001$). Updating the probeset definitions (Figures 6.2C and 6.3C) brings the molecular apocrine hierarchical cluster membership to 15 of 16 (94%) samples across the combined index cohorts with improved separation by phenotype ($p < 0.0001$). We note that median-centering per probeset by institution also results in statistically significant separation ($p < 0.0001$, see Appendix Figure A.6.1) with 13 of 16 (81%) molecular apocrine samples clustering together in the HC dendrogram (Appendix Figure A.6.2) compared to 15 of 16 (94%) samples using XPN. We note that this difference is not statistically significant.

We evaluate our second proposed criterion for determining molecular equivalence by using Significance Analysis of Microarrays (SAM) (Tusher, *et al.*, 2001) to identify the top 100 statistically significant probesets in each of the index cohorts (after normalization) that differentiate the hypothesized molecular apocrine phenotype from the remaining samples. The resulting gene signatures share 76 genes (Supplementary File 1), while the original two studies identified 138-gene (Doane, *et al.*, 2006) and 400-gene

(Farmer, *et al.*, 2005) profiles with 25 genes overlapping. The extent of overlap for both results is statistically significant (both $p < 0.0001$). For comparison, 100-gene signatures derived from a median-centered dataset using manufacture-provided probeset definitions has 25 overlapping genes and from a median-centered dataset using AffyProbeMiner probeset definitions has 33 overlapping genes (both $p < 0.0001$). While there is no notable difference in statistical significance, the larger number of common genes gives us more attributes with which to investigate the networks and gene interactions that define this species.

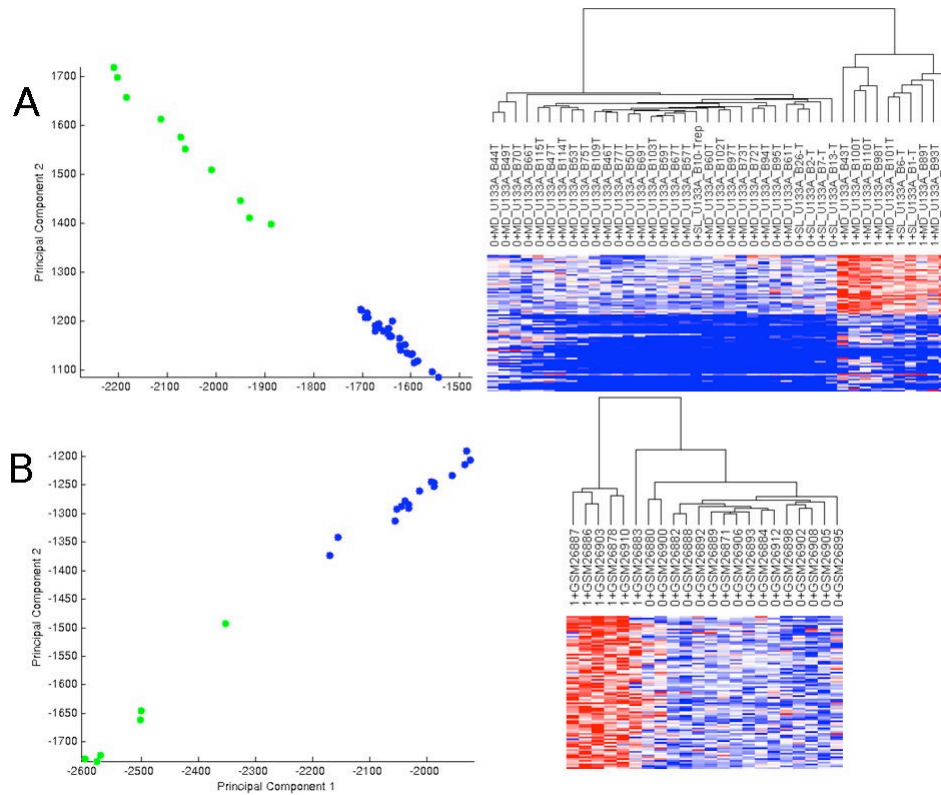


Figure 6.4: PCA and Hierarchical Clustering of Cross-Study Normalized Cohort #1 filtering for 100-Probeset Signature Generated from Analysis of Cohort #2 and Vice Versa

(A) Cohort #1: Doane *et al.* and Cohort #2: Farmer *et al.* (B) Cohort #1: Farmer *et al.* and Cohort #2: Doane *et al.* **Information:** The Principal Components Analysis plot (LEFT) shows that the first component separates the tumor samples according to their hypothesized “non-molecular apocrine” (blue) and

“molecular apocrine” (green) phenotypes using the 300 probesets identified by Significance Analysis of Microarrays (SAM) (Tusher, *et al.*, 2001). Hierarchical clustering (RIGHT) of the cross-study normalized data using the signature also clusters the samples according to phenotype. Phenotype is represented by 0 for “non-molecular apocrine” and 1 for “molecular apocrine”. Clustering was performed using the Pairwise- Average Linkage method and calculates distance using Euclidian Distance. The PCA uses the principal components for the two-dimensional coordinates.

We perform hierarchical clustering and PCA of the Doane *et al.* (Doane, *et al.*, 2006) cohort with the 100-gene signature identified with SAM on the Farmer *et al.* (Farmer, *et al.*, 2005) cohort, and vice versa to test compliance with our third criterion (see Figure 6.4). We compare these results with the performance of the published signatures on the published data as normalized by the submitting institution. While the samples do not group together as tightly (Appendix Figures A.6.3 & A.6.4) as they do with the 100-gene signatures derived using our normalized data, the signatures identified by Farmer *et al.* and Doane *et al.* can indeed predict the molecular apocrine phenotype in the other cohort without the need for combining the data *via* cross-study normalization.

Functional Analysis of the “Molecular Apocrine” Phenotype Using LeFEMiner

Using an approach that builds upon the concept of gene set enrichment, LeFEMiner identifies a set of top-ranked gene ontology (GO) categories in the normalized index cohorts (see Table 6.1). Notably, the “Androgen Up-regulated Genes” (86 genes (Nelson, *et al.*, 2002; Subramanian, *et al.*, 2005)) and “Breast Cancer Estrogen Signaling” (101 genes (Subramanian, *et al.*, 2005)) GO categories both are identified at 0% false discovery rate, with AR presenting as the top ranked gene in the “Breast Cancer Estrogen Signaling” category; three genes overlap between the two profiles used to define the GO categories. Table 6.1 shows that the AR and ER signal based pathways are the top two regulatory signaling pathways, after metabolic and other enzymatic pathways that are listed. These results support the hypothesis that the molecular apocrine subtype has molecular characteristics of a steroid hormone response similar to that of estrogen

response (Doane, *et al.*, 2006; Naderi & Hughes-Davies, 2008; Teschendorff, Naderi, Barbosa-Morais, & Caldas, 2006).

Table 6.1. Top-ranking Gene Ontology categories identified by LeFEminer on the normalized index cohorts.

Rank determined by p-value and false discovery rate. All categories shown in table were identified with 0% false discovery rate. The Category.Size column lists the number of genes in the pre-defined dataset representative of the GO category.

Category.Rank	Category.Name	Category.Size	Median.Category.P.Value
1	carboxylic acid metabolism	138	0
1	fatty acid metabolism	47	0
1	fatty acid metabolism BioCarta	24	0
1	MAP00480 Glutathione metabolism GenMAPP	18	0
5	organic acid metabolism	140	0.001
5	aromatic compound catabolism	6	0.001
5	MAP00350 Tyrosine metabolism GenMAPP	29	0.001
8	oxidoreductase activity, acting on CH-OH group of donors	44	0.0015
9	aromatic amino acid family catabolism	5	0.002
10	electron transporter activity BioCarta	102	0.0025
11	cyclic nucleotide-dependent protein kinase activity	4	0.003
11	ANDROGEN UP GENES na	56	0.003
13	tyrosine catabolism	3	0.004
13	Fatty Acid Synthesis BioCarta	14	0.004
15	regulation of locomotion	6	0.0065
16	acetylgalactosaminyltransferase activity	7	0.007
16	polypeptide N-acetylgalactosaminyltransferase activity	7	0.007
16	MAP00360 Phenylalanine metabolism GenMAPP	16	0.007
19	breast cancer estrogen signalling GEMArray	92	0.0075
20	peroxisome	36	0.008
20	MAP00512 O Glycans biosynthesis GenMAPP	7	0.008
22	amine catabolism	23	0.009
22	cAMP-dependent protein kinase activity	4	0.009
24	regulation of cell migration	6	0.01
24	microtubule cytoskeleton organization and biogenesis	23	0.01

24	mitotic spindle checkpoint	2	0.01
24	electron transporter activity	114	0.01
24	microbody	36	0.01
29	regulation of behavior	6	0.011
29	neuronal lineage restriction	2	0.011

Table 6.2: Transcription regulation analysis by GeneGo's MetaCore generates sub-networks centered on transcription factors given the 17 genes discovered by Backward Chaining Rule Induction (Figure 6.6).

Sub-networks are ranked by a p-value and interpreted in terms of Gene Ontology.

No	Network	GO Processes	Total nodes	Root nodes	p-Value
1	ESR1 (nuclear)	positive regulation of retinoic acid receptor signaling pathway (16.7%; 1.463e-03), negative regulation of mitosis (16.7%; 2.194e-03), epithelial cell maturation (16.7%; 2.559e-03), regulation of retinoic acid receptor signaling pathway (16.7%; 2.925e-03), melanosome localization (16.7%; 4.019e-03)	6	5	3.20E-18
2	HNF4-alpha	negative regulation of protein import into nucleus, translocation (16.7%; 7.318e-04), negative regulation of tyrosine phosphorylation of Stat5 protein (16.7%; 1.463e-03), regulation of protein import into nucleus, translocation (16.7%; 1.829e-03), ornithine metabolic process (16.7%; 2.559e-03), positive regulation of gluconeogenesis (16.7%; 2.559e-03)	6	5	3.20E-18
3	Androgen receptor	prostate gland development (40.0%; 4.905e-06), male somatic sex determination (20.0%; 3.050e-04), somatic sex determination (20.0%; 6.099e-04), gland development (40.0%; 7.323e-04), urogenital system development (40.0%; 7.845e-04)	5	4	1.44E-14
4	HNF1-alpha	glucose homeostasis (66.7%; 5.376e-05), carbohydrate homeostasis (66.7%; 5.376e-05), epithelial cell maturation (33.3%; 1.280e-03), bile acid biosynthetic process (33.3%; 1.646e-03), paraxial mesoderm formation (33.3%; 1.829e-03)	3	2	1.89E-07
5	HNF3-beta	neuron fate specification (66.7%; 2.343e-06), positive regulation of neuron differentiation (66.7%; 6.156e-06), neuron fate commitment (66.7%; 2.409e-05), cell fate specification (66.7%; 3.555e-05), epithelial cell differentiation (66.7%; 6.683e-05)	3	2	1.89E-07

Network Inference Analysis of "Molecular Apocrine" Phenotype Using Backward Chaining Rule Induction and MetaCore

We use See5 as a rule induction method (Rulequest, St. Ives, Australia) and MetaCore, a commercial pathways database with analysis tool distributed by GeneGo (St. Joseph, MI) (Ekins, Nikolsky, Bugrim, Kirillov, & Nikolskaya, 2007) for implementation of the BCRI strategy for network discovery (Edgerton, *et al.*, 2007; Fisher D, *et al.*, 2005; Fisher DH, *et al.*, 2006). The method is similar to one previously used to study yeast networks (Soinov, 2003). We followed the BCRI strategy for six successive iterations to identify 17 genes whose expression could predict threshold expression levels of the genes identified in the previous iteration (see Figure 6.5). A Transcription Regulation Analysis by MetaCore on the 17 genes identifies AR, ESR1 (ER), HNF4-alpha, HNF1-alpha, and HNF3-beta as significant transcription factors regulating the genes identified by BCRI (Table 6.2). The top 3 regulatory pathways listed in Table 2 are ER, HNF4-alpha, and AR. In addition, using Dijkstra's algorithm (MetaCore function) to find the shortest known directed paths within two nodes between the 17 BCRI genes results in a network that clearly shows the close relationship between AR and the BCRI genes (Figure 6.6). The transcription regulation and shortest path analyses (Table 6.2 and Figure 6.6) both also specify ER as having a close network relationship with the BCRI genes.

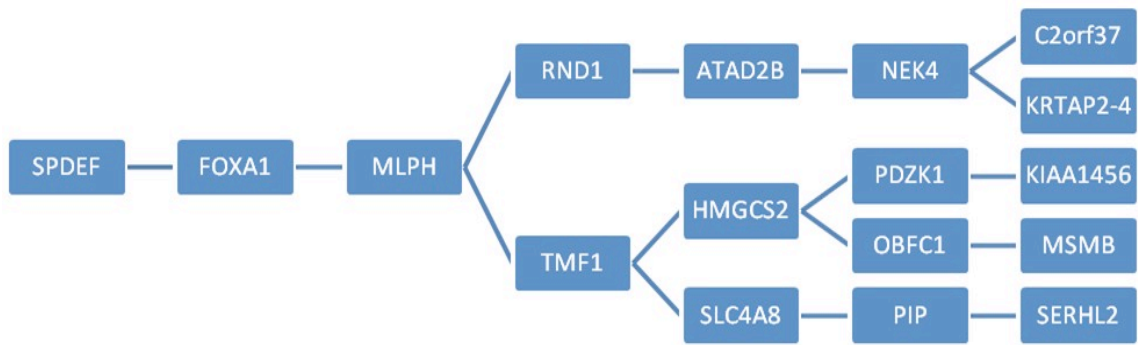


Figure 6.5. Network Inference Using Backward Chaining Rule Induction.

The 17 genes discovered by the Backward Chaining Rule Induction strategy applied to the index cohorts to postulate gene network relationships from the index cohorts' gene expression microarray data.

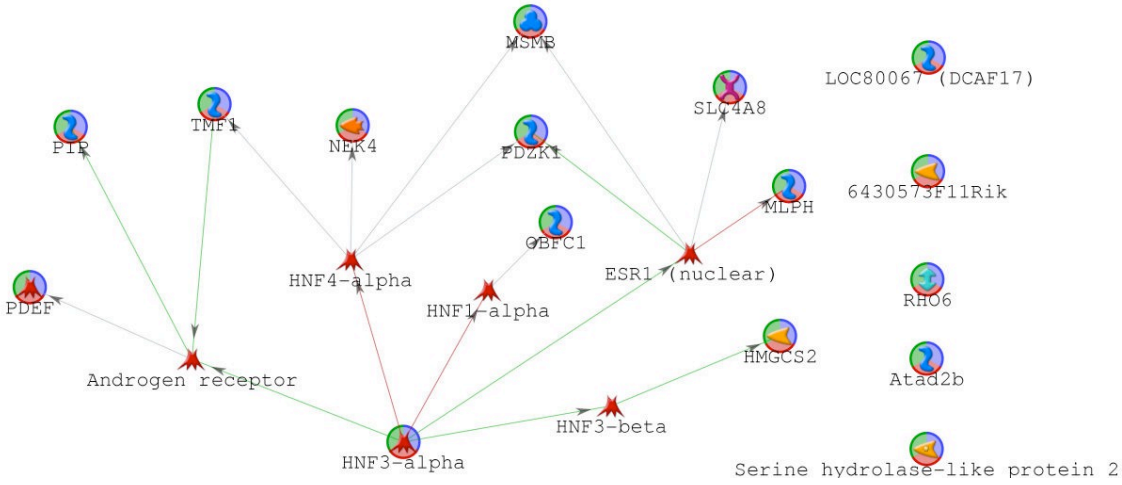


Figure 6.6. Prior Knowledge Support for Gene Relationships Identified by Backward Chaining Rule Induction.

Androgen Receptor is closely connected to the 17 genes identified by the Backward Chaining Rule Induction strategy as indicated by using MetaCore to identify the closest paths connecting the genes.

Persistence of Molecularly-defined Phenotype in Larger Dataset

At this point we have supporting evidence for a role for AR in defining the molecular apocrine subtype using two independent methods of network inference. We now seek to identify the gene network and pathways that interact with AR. Limited

sample sizes hinder this type of analysis (Broom, *et al.*, in preparation). Therefore, we expand our molecular apocrine gene expression data with ER⁻ samples from Ivshina *et al.*, Rouzier *et al.*, and Sotiriou *et al.*, bringing our total to 199 (Doane, *et al.*, 2006; Farmer, *et al.*, 2005; Ivshina, *et al.*, 2006; Rouzier, *et al.*, 2005; Sotiriou, *et al.*, 2006). We normalize with QN + XPN with updated probeset definitions. We apply SAM to the index cohorts and identify a 346-probeset signature at 0% false discovery rate to predict molecular apocrine samples (see Supplementary Table 1). We use these genes to perform PCA on the expanded cohort. These results show a natural demarcation in the larger ER⁻ dataset where the 22 molecular apocrine samples in the index cohort along with an additional 46 samples in the expanded cohort separate from the rest of the data (see Figure 6.7). We refer to these 68 samples as the “model-classified cohort.”

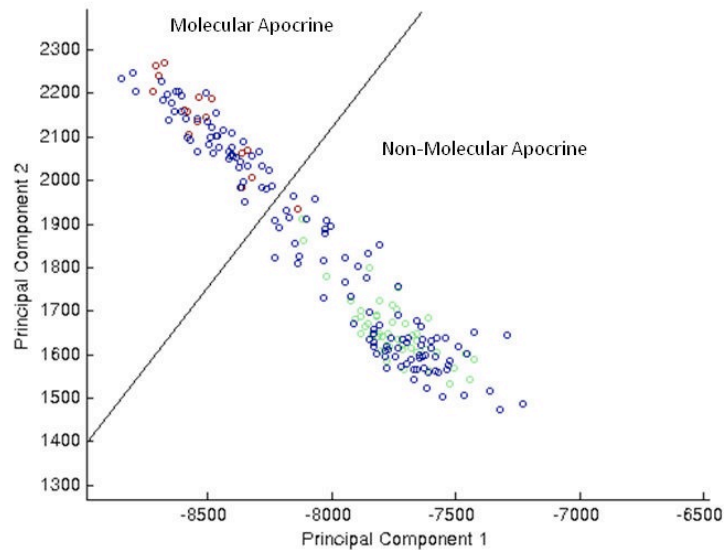


Figure 6.7: Principal Components Analysis of Cross-study Normalized, Five-Cohort Dataset.

The Principal Components Analysis of a 346-probeset signature derived from the combined Doane *et al.* and Farmer *et al.* data qualitatively shows that a large ER⁻ dataset combining the Ivshina *et al.*, Rouzier *et al.*, and Sotiriou *et al.* cohorts (blue) splits into the molecular apocrine (red) and non-molecular apocrine (green) phenotypes as identified by Doane *et al.* and Farmer *et al.* There are a total of 199 samples.

Network Inference Analysis of “Molecular Apocrine” Phenotype Using Gene Shaving & Robust Bayesian Network Analysis

First, we perform a unsupervised gene clustering using Gene Shaving (GS), and subsequently use Robust Bayesian Network Analysis (RBNA) to discover relationships between an AR-based cluster and other gene clusters (Broom, *et al.*, in preparation). Note that we do not seek support for the AR pathway as having a role in the molecular apocrine subtype in the model classified cohort because our gene classifier that predicts membership in the molecular apocrine subtype includes AR. This would have biased the network inferences toward selecting AR.

We identified the top 200 gene clusters using unsupervised GS (Supplementary File 2) ranked according to their internal cluster strength (order of how they were shaved). The cluster containing AR was the 7th ranked cluster (Appendix Figure A.6.5). Clusters 24, 29, and 7 were the top three clusters associated with the molecular apocrine phenotype using Kendall’s tau log rank analysis (see Methods). We selected RBNA as a network inference method for studying interactions between AR, represented as Cluster 7, and other gene clusters because in addition to discovering relationships between the clusters, it provides a “global” perspective on both *interaction* and *prevalence* in the patient population. The top 26 clusters correlating with the molecular apocrine phenotype analysis were used in this analysis. The number of clusters was selected using an absolute value of 0.5 as a cut-off for the Kendall’s tau log rank. This threshold was selected to maintain sufficient correlation with the molecular apocrine phenotype, to allow clusters to be used which will have interactions that may not involve the entire molecular apocrine phenotype or may overlap with other phenotypes, and finally to provide a sufficient number of clusters for RBNA to sample in order to quantify the relative

strengths of interactions in the samples. Network associations amongst 14 of these 26 clusters were identified with RBNA (Figure 6.9). The AR cluster (Cluster 7) is most strongly associated with Cluster 24 and less so with Cluster 71 (Figure 6.8). These are the only two clusters that directly interact with the AR cluster, and we select these two for further characterization.

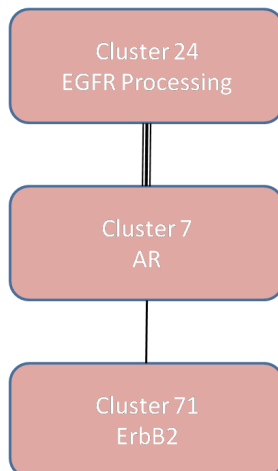


Figure 6.8: Robust Bayesian Network Analysis of Top Apocrine-related Gene Clusters Identified Through Gene Shaving Reveals Interactions with AR.

Robust Bayesian Network Analysis postulates network interactions between the top apocrine-related gene clusters, and their relative strength (indicated by bolder links connecting clusters). Cluster 7 (AR cluster) interacts with Cluster 29 (EGFR processing genes) and Cluster 71 (ErbB2 cluster). This is a subset of all the interacting clusters identified by RBNA (Supplementary Figure 6.6).

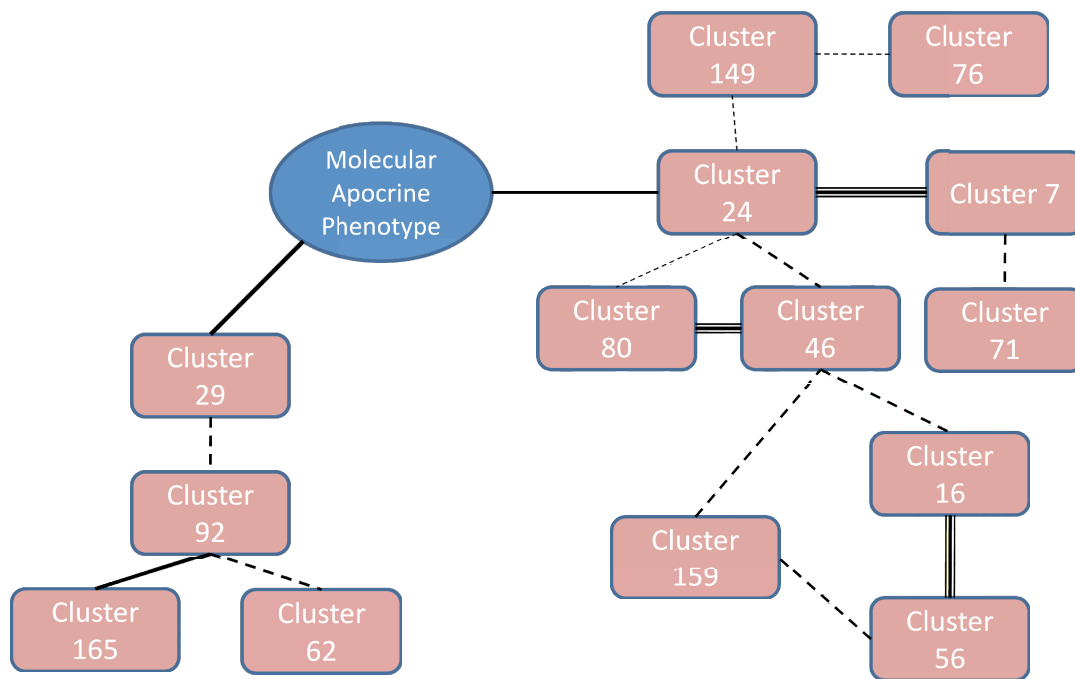


Figure 6.9: Robust Bayesian Network Analysis of Top Apocrine-related Gene Clusters Identified Through Gene Shaving.

Robust Bayesian Network Analysis postulates network interactions between the top apocrine-related gene clusters, with the strength of relationships indicated by boldness of links. The 14 interacting clusters are depicted along with a node illustrating the “molecular apocrine phenotype.” Links to the phenotype node indicate clusters with the strongest association with the molecular apocrine phenotype.

Analysis of the Interacting Gene Clusters

We submit the members of the interacting clusters to both MetaCore (Ekins, *et al.*, 2007) and GeneCards (Rebhan, Chalifa-Caspi, Prilusky, & Lancet, 1997; Stelzer, *et al.*, 2008) to identify associated gene ontologies and known transcription regulation relationships. From GeneCards we identify multiple upregulated species associated with EGFR processing in Cluster 24, which we label as the EGFR processing cluster, but are not necessarily involved in amplifying EGFR. MetaCore shows that Cluster 71, which contains ErbB2, also contains other EGFR-related genes. We call this cluster the ErbB2 cluster. We also analyzed the clusters that indirectly interact with the AR cluster. Cluster

16 is interesting because although AR is not a member, MetaCore reveals a large number of genes whose transcription is regulated by AR. Furthermore, MetaCore analysis of Cluster 16 suggests network relationships related to ER, p53, and Maspin (a tumor suppressor gene associated with breast, prostate, and pancreatic cancer). In addition to Cluster 16, MetaCore identifies relationships between ER and the genes in the AR and ErbB2 clusters along with clusters 56, 62, 71, 76, 80, and 92. Of interest, ErbB3 is present in cluster 62, which has an indirect link to the molecular apocrine subtype (see Figure 6.9).

DISCUSSION & CONCLUSION

Our conclusions are pertinent to both bioinformatics in general and to this particular breast cancer subset.

Observations of Normalization Strategies to Remove Institutional Bias in Meta-Analysis of Gene Expression Array Data

In the course of our investigation, we compared the effectiveness of normalizing data using quantile normalization, conventional median-centering, and a recently published algorithm called XPN. Although the data from the two institutions demonstrated adequate correlation after quantile normalization, results of the hierarchical clustering continued to be affected by institutional bias. This may indicate a particular sensitivity of hierarchical clustering to institutional bias.

Molecular Equivalence of the “ER Subclass A” with “Molecular Apocrine” Breast Cancer

We have proposed three criteria for evaluating molecular equivalence between transcript-defined subsets identified by two or more independently conducted studies: 1)

the majority of molecularly equivalent samples should cluster together and distinctively separate from the remaining samples in unsupervised clustering of the combined data; 2) there should be statistically significant overlap of gene signatures used to define the phenotype in each separate study; and 3) a classifier trained on data from one institution should successfully predict the phenotype in the other institution, and vice versa. We call upon the microarray community to consider these criteria and establish a standard protocol for establishing molecular equivalence.

In the course of our evaluation, we demonstrate that two of the three criteria proposed are met even without combining and normalizing the data together: the 25-gene overlap between the signatures identified by Farmer *et al.* and Doane *et al.* is statistically significant; and the published signatures for each of these studies adequately predicts the hypothesized breast cancer subset in the other index cohort. However, not only were we able to enlarge the extent of overlap in the signatures, but we found that only after appropriate normalization did the samples from the two institutions cluster together by hypothesized phenotype.

Role of AR Signaling in Molecular Apocrine Tumors

Both authors suggest a role for AR signaling in this subtype of breast cancer based on comparison to data generated by cell lines. In addition, Doane *et al.* suggests that there is some overlap of the signatures with known ER⁺ genes. We chose two different network inference methods to explore causal networks in this data. LeFeminer utilizes a gene set enrichment type approach while BCRI functions as a discovery strategy supplemented by pathway information from Metacore. We selected pathways that were common to both strategies as highly supported. The AR and ER signals were the two signaling pathways that were identified by both algorithms as relevant to the molecular apocrine phenotype. Expression of the ER molecular profile in the molecular

apocrine group, in spite of the fact that it is ER⁻ by immunohistochemistry, has been described by other authors (Doane, *et al.*, 2006; Naderi & Hughes-Davies, 2008; Teschendorff, *et al.*, 2006). From a bioinformatics perspective, and since BCRI is a relatively new method of network inference, we see this result as validation of its utility in pathway discovery.

Pathways that Interact with AR in Molecular Apocrine Breast Cancer

Our analysis shows that the molecular apocrine phenotype lacks an overexpression of basal cytokeratins, which is considered to be a defining feature of basal-like breast cancer (Gusterson, Ross, Heath, & Stein, 2005; Livasy, *et al.*, 2006). Thus, we can consider molecular apocrine tumors to be a distinct subset of ER⁻ tumors that includes both triple-negative and ER⁻/PR⁻/ErbB2⁺ tumors. Since we started our research, two other studies have discovered this subgroup (Teschendorff, Miremadi, Pinder, Ellis, & Caldas, 2007; Weigelt, *et al.*, 2008). One study identified it within triple-negative tumors alone while the other identified it to combine AR and ErbB2 signaling. We agree with the originating authors that the molecular apocrine tumors can be either ErbB2⁺ or ErbB2⁻ based on intraction studies that we will discuss below.

Our results reveal a strong interaction between the AR cluster and a cluster with several genes involved in EGFR processing. Several cell lines studies have hypothesized an interaction between EGFR and both AR and ER, suggesting that together they form a complex with Src that enhances EGFR phosphorylation of tyrosine and therefore increases the effectiveness of EGF signaling (Ignar-Trowbridge, *et al.*, 1992; Migliaccio, *et al.*, 2006; Migliaccio, *et al.*, 2000). However, this is the first study of gene expression data using cancer tissue from patients in which this interaction has been detected using gene expression data analysis methods.

A significant relationship is also revealed between the AR cluster and the ErbB2

cluster. The strength of the interaction between this cluster and the AR cluster is weaker than the EGFR processing cluster. In the index studies of molecular apocrine tumors, approximately half of the cases were ErbB2⁺. This is consistent with the less strong, but significant interaction between AR and ErbB2 in our analysis. In addition to simple co-expression, actual cross-talk between ErbB2 and AR pathways has been suggested based on cell line studies in breast (Migliaccio, *et al.*, 2006; Naderi & Hughes-Davies, 2008). These studies demonstrated an additive effect of AR inhibition in reducing ErbB2 signaling, and suggested that tumors that are AR⁺/ErbB2⁺ might need AR inhibition in addition to targeted anti-ErbB2 therapy to completely neutralize the effective of the ErbB2 signal.

In prostate cancer, cell lines studies have led investigators to hypothesize that ErbB family signaling, including EGFR (ErbB1), ErbB2, and ErbB3, can activate AR and therefore is responsible for evolution from androgen dependent to androgen independent tumor growth (Stern, 2008). Thus, tumors with AR transcription profiles might require therapy with ErbB family inhibitors.

Our results with both BCRI and GS combined with RBNA also support the role of FOXA1 interacting with AR in this phenotype. FOXA1 is known to have a role in potentiating steroid receptor transcription regulation, and its association with AR by immunohistochemistry has been reported by several other investigators (Badve, *et al.*, 2007; Habashy, *et al.*, 2008; Lacroix & Leclercq, 2004; Nakshatri & Badve, 2007; Thorat, *et al.*, 2008; Tozlu, *et al.*, 2006; I. Wolf, *et al.*, 2007). FOXA1 is a member of the AR cluster and was also directly identified by BCRI (Appendix Figure 6.5). Three other genes identified directly by BCRI (*i.e.*, SPDEF, MLPH, and SERHL) are also part of the AR cluster, which further emphasizes BCRI as a valid network inference strategy.

Associations between PIK3CA mutations and AR in triple-negative tumors have been reported recently (Gonzalez-Angulo, *et al.*, 2009). Strong associations between a

PIK3CA expressing cluster and AR cluster were not identified. However, given that mutations in PIK3CA may not be picked up on standard gene expression platforms, this association may not be discoverable from the data.

Clinical and Therapeutic Implications for Molecular Apocrine Breast Cancer

We propose that therapies targeting AR activity may present a rational strategy for managing these patients. The concept of introducing AR blockade as a therapeutic option for breast cancer has received more attention recently (Doane, *et al.*, 2006; Farmer, *et al.*, 2005; Moe & Anderson, 2007; Naderi & Hughes-Davies, 2008; Nahleh, 2008; Ogawa, *et al.*, 2008; Swain, 2001; A. R. Tan & Swain, 2008; Weigelt, *et al.*, 2008). Older trials of AR blockade did not select for patients with AR dependent signaling or AR expression and therefore may not have addressed the question with an optimal cohort (Perrault, *et al.*, 1988). Based upon our interaction studies, we also recommend that any therapeutic strategies for the molecular apocrine subgroup consider combinatorial targeted therapy to include ErbB family targets, particularly EGFR targeted therapy for the entire molecular apocrine subtype and ErbB2 therapy for those tumors that overexpress ErbB2.

While there is evidence to support ER response genes in the molecular apocrine subset, anti-estrogen therapy using tamoxifen in ER⁻ women in general has been shown to have too little benefit for clinical use. However, small benefits were reported that point to the need for more study (Swain, 2001). An important question arises - is the presence of ER signaling inferred because AR and ER share a common pathway, or is there cross-talk where AR activation stimulates the ER pathway? Our pathway analyses from BCRI that demonstrate AR and ER as related signals (Figure 6.6), and analysis of Cluster 16 (Supplementary File 3), do not support a common pathway that is activated by AR and ER. While interesting, these results are not conclusive. We note that if cross-talk from

activated AR signaling is the cause of the ER signal activation in ER⁻ tumors, then AR inhibition therapy would be sufficient to interrupt this signal.

There is little known about the survival of molecular apocrine tumors as they have only been recently introduced as a subtype. Farmer *et al.* (Farmer, *et al.*, 2005) describes poor survival in the cohort that they identified from the literature. Weigelt *et al.* (Weigelt, *et al.*, 2008) suggest that apocrine carcinomas can expect a 10-year survival rate of 35-50%, and Teschendorff *et al.* (Teschendorff, *et al.*, 2007) suggest that it has the poorest outcome of all of the ER⁻ tumor types. Other data suggests that AR⁺ tumors that are otherwise triple-negative as defined by immunohistochemistry may have a better prognosis than the basal subtype of tumors (Rakha, *et al.*, 2007). In a recent study of AR protein expression in any type of breast cancer, an improved prognosis was associated with AR expression above a certain threshold in ER⁺ tumors (Gonzalez-Angulo, *et al.*, 2009). It may be that interactions with ErbB family members modify the survival characteristics of AR⁺ tumors. This deserves further study.

Learning the Systems Biology of Cancer Using Network Inference Methods to Analyze Gene Expression Data

Our results support the strength of using network inference to analyze gene expression array data for oncogenic pathways and their interactions. Our results demonstrate that the discovery of oncogenic pathways and their interactions does not have to rely on comparison with signatures from cell lines, but can be discovered using network inference methods. This study demonstrates the rich knowledge resource within gene expression data generated from human tissues.

MATERIALS & METHODS

Data Collection

The raw CEL files from Farmer *et al.* (Farmer, *et al.*, 2005) are available for download at NCBI GEO Datasets under accession GSE1561. The raw CEL files from Doane *et al.* (Doane, *et al.*, 2006) are available for download at the National Cancer Institute caArray database. The raw CEL files from Ivshina *et al.* (Ivshina, *et al.*, 2006) and Sotiriou *et al.* (Sotiriou, *et al.*, 2006) are available for download at NCBI GEO Datasets under accession GSE4922 and GSE2990, respectively. The raw CEL files from Rouzier *et al.* (Rouzier, *et al.*, 2005) are available for download at <http://bioinformatics.mdanderson.org/pubdata.html>.

Microarray Normalization: Removing Systematic and Institutional Bias

The Doane *et al.* and Farmer *et al.* cohorts were first quantile-normalized (Bolstad, *et al.*, 2003) together using the default settings in DNA-Chip Analyzer (dChip), a software package for probe-level analysis of gene expression microarrays (C. Li, 2008). This process was repeated twice: the first time, the original Affymetrix-provided chip definition file (CDF) was used, and the second time, a transcript-consistent Affymetrix-formatted Chip Definition File (CDF) downloaded from AffyProbeMiner (H. Liu, *et al.*, 2007) was used. A recently published cross-study normalization scheme called XPN (Shabalín, *et al.*, 2008) was subsequently implemented to further combine the quantile-normalized datasets into a single, unified datasets with significantly reduced systematic bias; one dataset derives from normalization with Affymetrix's CDF and a second dataset (the primary dataset used for analysis in this study) derives from normalization with AffyProbeMiner's CDF. The details regarding the normalization scheme, referred to as XPN, have been previously described (Shabalín, *et al.*, 2008). In short, the XPN

algorithm is based on linking gene/sample clusters amongst given datasets. Data is scaled and shifted according to the assumption that similar gene-sets cluster together across multiple platforms. XPN has been shown to successfully remove systematic bias, while avoiding the loss of useful biological information due to data over-correction (Shabalín, *et al.*, 2008).

The other cohorts were included to investigate the persistence of the molecular trends identified in the Doane *et al.* and Farmer *et al.* datasets. All five cohorts were quantile-normalized with dChip using a transcript-consistent Affymetrix-formatted CDF provided by AffyProbeMiner (H. Liu, *et al.*, 2007). Then, XPN was used in serial increments to bring the five cohorts into uniform agreement by removing persistent systematic bias between the datasets.

Significance Analysis of Microarrays: Modified T-Test

Significance Analysis of Microarrays (SAM) was performed on the normalized Doane *et al.* (Doane, *et al.*, 2006) and Farmer *et al.* (Farmer, *et al.*, 2005) data individually to identify top 100 probesets that classify between the molecular apocrine samples and the remaining samples. SAM was also performed on the combined Doane *et al.* and Farmer *et al.* subset of the cross-study normalized, five-cohort data to identify a gene signature with 0% false discovery rate for classifying molecular apocrine samples from the remaining samples, and identifying similar molecular trends in the remaining data. SAM is based on a modified T-test; details regarding the algorithm have been previously described (Cui & Churchill, 2003; Tusher, *et al.*, 2001).

Hierarchical Clustering and Principal Components Analysis

Hierarchical Clustering was performed using a Pairwise-Average Linking method and Euclidian Distance as the distance measure. Both Hierarchical Clustering and

Principal Components Analysis were performed on the GenePattern software package provided by the Broad Institute (Kuehn, *et al.*, 2008; Reich, *et al.*, 2006). Visualizations of the Principal Components Analysis were performed with MATLAB (Mathworks, Natick, MA).

Two-Dimensional Kolmogorov-Smirnov Test

The Fasano & Franceschini statistical test (Fasano & Franceschini, 1987), a two-dimensional adaptation of the Kolmogorov-Smirnov test (Lopes, Hobson, & Reid, 2008), was performed on the coordinates derived from the first two principal components using an algorithm provided by Numerical Recipes in Fortran 90 (Press, Teukolsky, Vetterling, & Flannery, 1996).

Statistical Significance of Overlap Between Gene Signatures

The Fasano & Franceschini statistical test (Fasano & Franceschini, 1987), a two-dimensional adaptation of the Kolmogorov-Smirnov test (Lopes, *et al.*, 2008), was performed on the coordinates derived from the first two principal components using an algorithm provided by Numerical Recipes in Fortran 90 (Press, *et al.*, 1996).

Backward Chaining Rule Induction

Backward Chaining Rule Induction (BCRI) is a supervised learning approach for identifying relationships amongst genes that can predict for the molecular apocrine phenotype. In order to initialize the BCRI strategy, we use a classifier method called See5 (Rulequest, St. Ives, Australia) to build a prediction model from the normalized gene expression data for classifying the molecular apocrine phenotype from the remaining samples in the index cohorts. Successive iterations of the BCRI strategy infer gene network relationships by predicting threshold expression of genes from other genes. Further details regarding the BCRI strategy have been previously described (Edgerton, *et*

al., 2007; Frey, *et al.*, 2005).

Gene Shaving

To identify clusters of highly correlated genes, we used unsupervised Gene Shaving (Hastie, *et al.*, 2000). Specifically, we used a high-performance, parallel C implementation of the method that was developed from the GeneClust software package (Do, Broom, & Wen, 2003). Gene Shaving was used independently on both unweighted data and on 127 bootstrap resamples, extracting the first 150 gene clusters in each case. In both cases, the data was first ranked within each sample. To obtain the unweighted data, the ranked data was ranked again, this time across samples within each cohort. For the bootstrap resamples, each sample within a cohort was assigned a random weight chosen from the Bayesian bootstrap distribution (Rubin, 1981) and weighted rankings across samples within each cohort were computed. In both cases, the rank of each sample was scaled by the number of samples in the cohort, so that for each cohort the data is in the range zero to one. Robust clusters were obtained from the combined outputs of the Gene Shave runs by selecting those genes that occur frequently together in the outputs of individual runs. We extracted the first 200 clusters with the largest number of co-clustering genes, weighted by the homogeneity of the clusters to which they belong.

Robust Bayesian Network Analysis

The 200 robust clusters obtained by Gene Shaving were ranked by their correlation with their molecular apocrine phenotype. A cluster meta-gene score was obtained for each sample by computing the signed average mean gene. We note that unlike other gene clustering methods, Gene Shaving clusters may include both correlated and anti-correlated genes. The 26 clusters with the highest absolute Kendall Tau correlation between the cluster meta-gene scores and the molecular apocrine phenotype

status were selected for network analysis.

The network analysis included nodes for the 26 gene clusters most highly correlated with molecular apocrine status and a node for molecular apocrine status. The cluster meta-gene scores were each discretized to three levels: the lowest, middle, and highest thirds of the expression range for each meta-gene. Forty thousand bootstrap resamples of the discretized weights were obtained by randomly weighting each sample according to the Bayesian bootstrap distribution (Rubin, 1981), and a high-scoring network was found for each resample using greedy hill-climbing with random restarts and the sparse candidate algorithm (N. Friedman, Nachman, & Pe'er, 1999). The scoring function used was DPSM with $\lambda = 1$ (Yang & Chang, 2002).

Edges that occurred frequently (in either direction) within the forty thousand best networks thus obtained were selected for the final network. Edges that occurred in at least 97.5% of the networks are drawn with a triple black line, those that occurred in at least 95% of the networks with a black line, and those that occurred in at least 85% of the networks with a dashed line. Gene clusters that are not connected by any path along such edges to the node for molecular apocrine status are not included.

LIST OF KEY ABBREVIATIONS

Estrogen Receptor (ER); Human Epidermal Growth Factor Receptor 2 (ErbB2) commonly referred to as Her2neu; Androgen Receptor (AR); Epidermal Growth Factor Receptor (EGFR); Quantile Normalization (QN); Hierarchical Clustering (HC); Principal Components Analysis (PCA); Significance Analysis of Microarrays (SAM); Backward Chaining Rule Induction (BCRI); Gene Shaving (GS); Robust Bayesian Network Analysis (RBNA)

ACKNOWLEDGEMENTS

Edgerton, Broom, and Cristini were supported by the National Library of Medicine (R01-LM008000) to Edgerton. We thank Andrey Shabalin for helpful discussion regarding cross-study microarray gene expression normalization with XPN. We thank Roland Bassett, Jr. for helpful statistics advice in determining significance of gene signature overlaps. We also thank Douglas H. Fisher, a co-developer of BCRI, for helpful comments. A portion of this material was presented as a platform at the 2009 Annual Meeting of the United States and Canadian Academy of Pathology in Boston, MA (“Does Androgen Receptor Have a Role in a Subgroup of ER Negative Breast Cancer?”). We would like to dedicate this work to the memory of Vivian Cooper, a close friend and colleague of Dr. Fisher, who died of breast cancer on April 10, 2009.

Chapter 7: From Receptor Dynamics to Directed Cell Motion: A Predictive Agent-Based Model for Cell Motility in Complex Microenvironments

PREFACE

It is thought that a significant factor in cancer-related deaths is the ability of tumor cells to spread from primary tumors to form metastases in alternate locations throughout the body. This spreading behavior relies on upon cell motility, which enables cells to invade the host tissues and gain access to the lymphatics and blood vessels, and thus access to the rest of the body. Motility behavior of tumor cells need be analyzed as carefully as the molecular expression pattern of invasive cancer cells (J. Condeelis, *et al.*, 2005).

The tumor microenvironment is an important determinant of the motile behavior of cancer cells, and their desire to invade/metastasize. In the transition of cancer cells towards an invasive phenotype, they alter their interactions with the extracellular matrix, growth factors, and response to local levels of molecular species. In order to develop therapeutic strategies to inhibit invasion and metastasis, an understanding of how molecular patterns (gene expression, protein interactions) and the microenvironment influence cancer cell motility is required.

In this next chapter, we develop an agent-based model of cell motility where the locomotion of cells is the emergent behavior of cytoskeletal reorganization in response to chemical signals in the microenvironment. We calibrate and test the model using experimental cell motility data of neutrophils and breast cancer cells, respectively, in response to controlled gradients of chemical species. This cell-scale model is expected to serve as a method for modeling the motile behavior of cells in our multi-scale cancer

modeling framework, and is an initial step towards linking pathway signaling knowledge with a cell-scale model to predict an emergent behavior, in this case cell motility.

This research was a collaborative effort between Sandeep Sanga, Dr. Paul T. Macklin, and Dr. Vittorio Cristini. Sanga carried out the model development, simulations, analysis, drafted the manuscript, and participated in the study design. Macklin and Cristini conceived the study design, supervised the research, and helped draft the manuscript.

Note: This chapter is currently under preparation as a research manuscript (Sanga, Macklin, & Cristini, In Preparation) to be submitted to a journal whose audience is interested in computational systems biology and theoretical biology.

ABSTRACT

Cellular behaviors are thought to be the consequence of intricate molecular signaling networks influenced by their local microenvironment. Directed cell motility is a behavior relevant to many biological processes including embryogenesis, angiogenesis, inflammatory response, and cancer metastasis. We develop an agent-based model for directed cell motility where motion is an emergent feature of actin cytoskeletal reorganization in response to receptor-ligand binding. We introduce receptor occupancy theory to establish a functional link between receptor-ligand binding events at the cell periphery and downstream intracellular signals leading to actin protrusions causing cell motion. We calibrate the functional relationship using experimental data reported in the literature of dHL-60 neutrophil-like cells responding to controlled fMLP gradients, and subsequently predict the motility of MDA-MB-231 breast carcinoma cells responding to EGF under different gradient conditions. These simulation results support that the underlying mechanisms responsible for directed cell migration are conserved across cell types, and that spatial gradient sensing alone is all that is required to achieve chemotaxis. Furthermore, we demonstrate that in order to recapitulate the experimental data used for

calibration, the curve describing the functional relationship between receptor-ligand dynamics and internal signaling must be sigmoidal in shape. We validate this curve shape using a calibrated protein signaling model describing EGFR signaling from the literature. This result demonstrates a method for deriving functional relationships from pathway signaling models that can be linked to a discrete cell model to predict behavior such as motility.

INTRODUCTION

The cell is the most basic structural and functional unit of living organisms, and is itself an intricate machine whose ultimate behavior derives from a complex set of physical processes reacting to its microenvironment. The fundamental causes of cancer are thought to reside at the genome and protein scales, where a series of mutations and/or altered levels of expression result in aberrant protein function, and enable cells to develop a selective advantage, allowing them to reproduce and mobilize in defiance of normal constraints. At the gross scale, cancer is observed as a mass of cells growing as a bulk tumor and interfering with the function of vital organs. Almost all deaths due to cancer can be attributed to the effects of metastasis (Mehlen & Puisieux, 2006), which is when neoplastic cells also acquire the biological behavior of invasion. Among the critical steps in the invasion-metastasis cascade are the cytoskeletal reorganization affecting cell shape and motility, the cancer cells' ability to form invadopodia that release collagenases and proteinases, invasion of neighboring tissues with escape into the vascular or lymphatic circulation, and then taking up residence and proliferating into tumors that affect vital organs away from the primary tumor (Kedrin, van Rheenen, Hernandez, Condeelis, & Segall, 2007).

Mathematical-Computational Modeling Can Link the Multiple Scales

Without question, cancer is a complex disease with underlying physiological

processes spanning a wide range of interrelated temporal and spatial scales. An ongoing research challenge has been to relate biological information spanning these multiple scales. Mathematical models implemented as computational frameworks have proven to be useful tools for scientifically interrogating biological phenomena (Edelstein-Keshet, 2005), including cancer (Wayne Materi & David S. Wishart, 2007), and can serve as a conduit for connecting gene- and protein-scale intracellular signaling networks with tissue-scale morphology and behavior (Kansal, *et al.*, 2000a; Sanga, Frieboes, Zheng, *et al.*, 2007; Z. Wang, *et al.*, 2007; Zhang, *et al.*, 2007; Zhang, *et al.*, 2009). In this regard, agent-based modeling (ABM) has become a popular computational modeling technique for describing interacting cells, such as diseased cancer cells, because of its well-suited framework for simulating individual cells and the sub-cellular dynamics governing their phenotype in response to external cues from the microenvironment (Peirce, Skalak, & Papin, 2006). The primary value of cellular automata and ABM is to predict higher level emergent patterns by simulating interacting entities, whose individual phenotypic behaviors are governed by a given set of “rules”, through both time and space. ABM in particular enables cells to have arbitrarily complex internals governing their behavior without constraint to grids. Emerging tissue patterns and morphologies have important implications on the biology of the phenomenon in question. In the case of cancer, understanding how and why tumors evolve into the shapes that they do, invade local host tissue, and eventually metastasize to secondary locations throughout the body are processes that can be addressed with ABM in combination with other, more continuum based approaches, in a multi-scale fashion (Chuang, *et al.*, in preparation; Deroulers, Aubert, Badoual, & Grammaticos, 2009; Edgerton, *et al.*, In Review; Macklin, Kim, Tomaiuolo, Edgerton, & Cristini, 2009; Sanga, Frieboes, Zheng, *et al.*, 2007). Recent studies have already combined traditional experimentation with multi-cell agent-based modeling to recapitulate realistic tissue patterning (Macklin, *et al.*, 2009; Thorne, Bailey,

& Peirce, 2007). For example, Macklin *et al.* (2009) develops a model for ductal carcinoma in situ of the breast using an ABM approach, and directly compares simulation results with immunohistochemically-stained pathological slides. These results demonstrate the effectiveness of ABM-type models in predicting multi-cell tissue patterns seen in vivo by tuning parameters governing cellular behaviors such as proliferation, apoptosis, and cell-cell interactions. An intended application of this type of work is to assist the management of breast cancer patient care by more accurately predicting surgical margins (Edgerton, *et al.*, In Review).

From a modeling perspective, a cell's phenotype is a collection of states. The key cell states can be distilled to quiescent, proliferative, apoptotic, adhesiveness, and motile; Cancerous cells have phenotypic alterations such as increased proliferation, decreased apoptosis, decreased adhesiveness, and increased motility. Theoretical research has already implicated the importance of these cell states in understanding tissue scale behavior of tumors (A. R. Anderson, 2005; Macklin, *et al.*, 2009; Sanga, Frieboes, Zheng, *et al.*, 2007). The complex interplay between intracellular biochemical, biomechanical, and bioelectrical signaling, along with microenvironmental stimuli, is responsible for determining the cell's phenotypic properties; thus a modeling framework linking sub-cellular network properties to key cellular activities (*e.g.*, quiescence, proliferation, death, adhesiveness, and motility) can bridge the molecular and cellular scales so as to better understand how events at the molecular scales affect multi-cellular behavior.

Signal Transduction Research Reveals Signaling Relationships Governing Cellular Behavior

In the last decade, we have witnessed a surge in biological data available in the public domain as a result of high-throughput microarray technologies capable of probing

cellular information at the DNA, transcript, and protein scales, as well as advances in experimental techniques such as fluorescent microscopy and mass spectrometry towards studying protein-protein interactions. From these data, bioinformatics and systems biology approaches are utilizing computational algorithms in partnership with experimental systems for reconstructing signaling relationships in order to hypothesize and discover the underlying physiological mechanisms responsible for cellular phenotypes. Often, pathways and signal transduction research begins by defining the architecture of the system in question through identifying a “connection map” of the key genes/proteins and their interacting relationships (Broom, *et al.*, in preparation; Edgerton, *et al.*, 2007; Eungdamrong & Iyengar, 2004a, 2004b; Fisher DH, *et al.*, 2006; Frey, *et al.*, 2005; Sanga, *et al.*, In Press; Soinov, 2003). Following pathway network reconstruction, the next step for investigating the system-level dynamics is to frame a mathematical model for the network, usually in the form of chemical kinetics, compartmental, or reaction-diffusion equations (Bhalla & Iyengar, 1999; Eungdamrong & Iyengar, 2004a, 2004b; Neves & Iyengar, 2002).

Cancer and its progression towards metastasis involves a number of events, with multiple signals from tumor and stromal cells, the extracellular matrix, and soluble growth factors influencing the behavior of cancer cells (Chambers, Groom, & MacDonald, 2002; M. D. Levine, Liotta, & Stracke, 1995; A. Wells, 2000). One of the most important, and best understood, growth factor systems in this regard is the EGF/EGFR system, long implicated in cancer development (Price, Tiganis, Agarwal, Djakiew, & Thompson, 1999; Steeg, 2003b; A. Wells, Kassis, Solava, Turner, & Lauffenburger, 2002). Traditionally associated with tumor cell proliferation and growth, EGFR expression has been found to correlate with the metastatic potential of various cancers (Radinsky, *et al.*, 1995; Wiley, *et al.*, 2003). On the cellular level, EGF was shown to induce chemotaxis of metastatic breast cancer cells, both in vivo (T. Turner,

Epps-Fung, Kassis, & Wells, 1997) and in vitro (Steeg, 2003b; Wyckoff, Segall, & Condeelis, 2000). This is particularly relevant to metastasis, since platelets, smooth muscle cells, monocytes, and macrophages have been shown to produce EGF and related growth factors (Bailly, Yan, Whitesides, Condeelis, & Segall, 1998; Chambers, *et al.*, 2002; Dluz, Higashiyama, Damm, Abraham, & Klagsbrun, 1993; Kume & Gimbrone, 1994; Peoples, *et al.*, 1995). Gradients resulting from the release of these factors may provide chemotactic cues that direct metastatic cell motility towards blood vessels, where they can enter the blood stream and travel to other sites in the body (J. S. Condeelis, *et al.*, 2001; T. Turner, *et al.*, 1997), as well as invasion into and disruption of the surrounding host tissue. In order to understand and subsequently treat metastasis, we need to understand the mechanism of cancer cell motility in response to EGF and other chemoattractants.

Towards the Development of Predictive, Multi-Scale, Computer Representations of Cancer by Incorporating Pathways Knowledge into Each Cell

ErbB signaling pathways, more specifically ErbB1 (EGFR) and ErbB2 (Her2neu), have been the subject of vast amounts of both experimental and theoretical research (W. W. Chen, *et al.*, 2009; A. DeWitt, *et al.*, 2002; A. E. DeWitt, *et al.*, 2001; Harms, Bassi, Horwitz, & Lauffenburger, 2005; Kholodenko, *et al.*, 1999; Lauffenburger, 2000; A. Wells, *et al.*, 2006; Wiley, *et al.*, 2003). Lauffenburger, Wiley, and coworkers have developed, parameterized, and calibrated systems of ordinary differential equations (ODEs) modeling EGFR-ligand binding, trafficking, and downstream signaling via a chemical kinetics approach that has proven useful for studying EGFR-ligand dynamics (W. W. Chen, *et al.*, 2009; A. DeWitt, *et al.*, 2002; Lauffenburger, 2000; Lazzara & Lauffenburger, 2009; A. Wells, *et al.*, 2006; Wiley, *et al.*, 2003). Furthermore, Deisboeck and coworkers have pioneered the use of ABM as a multi-scale cancer modeling tool, and

have recently implemented similar systems of ODEs for tracking gene and protein signaling events resulting from EGFR-ligand binding thought to govern migratory and proliferative behavior (C. Athale, *et al.*, 2005; C. A. Athale & Deisboeck, 2006; L. L. Chen, Zhang, Yoon, & Deisboeck, 2009; Deisboeck, *et al.*, 2001; Z. Wang, *et al.*, 2007; Zhang, *et al.*, 2007; Zhang, *et al.*, 2009).

Cell Motility Modeling Using an Agent-based Approach with Pathways Knowledge

Ultimately, cell movement is an emergent behavior of many dynamic processes regulating the polymerization and depolymerization of actin filaments at the leading and trailing edges. There has been significant progress modeling cell motility by linking observations of actin protrusion/retraction dynamics to theoretical considerations of intracellular signals to capture motion of individual and interacting cells (Flaherty, McGarry, & McHugh, 2007; Keren, *et al.*, 2008; Mogilner, 2006; Mogilner & Edelstein-Keshet, 2002; Mogilner & Rubinstein, 2005; Satulovsky, Lui, & Wang, 2008). These efforts have initiated the framework for modeling cells as “shape machines” whose motility is based on regulation of actin cytoskeletal dynamics. Inspired by these previous efforts, in this report, we describe a multi-scale agent-based model of cancer cell motility built upon emerging biological knowledge of the underlying signaling pathways, triggered by surface receptor and ligand binding dynamics, that regulate cell morphology (Z. Wang, *et al.*, 2007; Wiley, *et al.*, 2003; Zhang, *et al.*, 2009). While previous approaches mostly implement migration velocities (speed and direction) in a rules-based a priori manner, we seek to use a multi-scale ABM approach to more biophysically simulate cell motility where receptor-ligand binding events predict underlying regulation of actin cytoskeletal dynamics, and emergent cell motility. This effort brings the modeling community (*e.g.*, The Center for the Development of a Virtual Tumor (Deisboeck, Zhang, & Martin, 2007)) a step closer towards comprehensive, multi-scale

cancer models which monitor dynamics changes within each cell's molecular networks as a function of their microenvironment. In this study, we build upon the pioneering models presented in the literature (Keren, *et al.*, 2008; Kholodenko, *et al.*, 1999; Macklin, *et al.*, 2009; Maree, Jilkine, Dawes, Grieneisen, & Edelstein-Keshet, 2006; Mogilner, 2006; Satulovsky, *et al.*, 2008; Z. Wang, *et al.*, 2007; Wiley, *et al.*, 2003; Zhang, *et al.*, 2009) and find that we are able to reproduce both random and directed cell motility with respect to chemokine gradients (chemotaxis) as well as with respect to extracellular matrix gradients (haptotaxis).

The Model Building Cycle

Predictive models at the cell or tissue scale are expected to have applications towards model-based predictive and/or preventative medicine (Rodriguez-Fernandez, Mendes, & Banga, 2006). A typical model building cycle consists of 1) defining the purpose of the model, 2) identifying *a priori* knowledge from available data, 3) choosing a suitable modeling framework, 4) performing parameter estimations using available data leading to a *prototype model*, 5) validating the prototype model with new experiments to reveal deficiencies, and 6) refining the model and plan new validation experiments (Rodriguez-Fernandez, *et al.*, 2006; van Riel, 2006). The scope of this work involves developing a *prototype* model for cell motility based on considerations of receptor-ligand binding trained on *in vitro* experimental motility data of neutrophil-like cells and validated with experimental data of breast carcinoma cells. The remainder of this report guides the reader through the steps of the model building cycle. We describe the modeling details, calibrate the functional relationship linking receptor-ligand binding to internal signaling based on experimental data from the literature, and finally validate the model performance by comparing the calibrated model to experimental results from the literature performed under different gradient conditions not used during the calibration

phase.

MODEL DEVELOPMENT

Key Observables of Directed Cell Motility: A Priori Knowledge for Model Development

Cell migration is a well-coordinated, multi-step process where 1) a cell acquires a polarized morphology in response to extracellular cues; 2) the actin cytoskeleton reassembles at the front to form protrusions called lamellipodia (flat-shaped) and filopodia (finger-shaped); 3) these protrusions form focal adhesions to the extracellular matrix acting as anchors; and 4) to move forward, the cell contracts while releasing adhesions towards the rear of the cell. The mechanisms of chemotaxis have been rigorously studied in *Dictyostelium* (in response to cyclic AMP) and neutrophils (in response to fMLP and IL-8), both classic models for studying chemotaxis. Progress has also been made in studying chemotaxis of cancer cells in response to EGF, considered to be a model for metastasis. While the receptors and ligands involved in inducing morphological changes and directed migration varies across cell types, it is generally accepted that the role of actin dynamics in regulating the cytoskeletal structure associated with membrane protrusions and cell-matrix focal adhesions is conserved (Le Clainche & Carlier, 2008; Rafelski & Theriot, 2004).

A generally accepted framework for capturing characteristic morphological behaviors in a cell motility model is the concept of Local Excitation Global Inhibition (LEGI) (Iglesias & Devreotes, 2008; Janetopoulos & Firtel, 2008; Levchenko & Iglesias, 2002; Ma, Janetopoulos, Yang, Devreotes, & Iglesias, 2004; Parent & Devreotes, 1999; Satulovsky, *et al.*, 2008; Skupsky, Losert, & Nossal, 2005). The LEGI framework asserts that a cell's ability to sense gradients derives from an internal balance between signaling

molecules that act locally and those that act globally. The local signaling molecules generate a stimulus focused towards the leading edge of a motile cell, and the global signaling molecules mediate an adaptation to the average stimulus in the cell.

We seek to advance a generalized model for cell motility that helps elucidate the mysteries of chemotaxis. In particular, the cell motility community has not yet definitively identified how cells sense and interpret external cues in the microenvironment, and subsequently morphologically respond to gradients. While it is known that cells build up a polarized spatial pattern of internal signal proteins towards the up-gradient, it remains unclear whether this pattern and eventual response to migrate is a result of temporal or spatial sensing, or perhaps some combination of both. Temporal sensing of gradients has been proposed as a method of gradient sensing for bacteria, where they sense local ambient concentration of chemoattractant at a given moment and compare with previous moments in order to determine their motile behavior. In contrast, spatial sensing has been proposed as a method more complex eukaryotic cells use; rather than compare chemoattractant through time, cells compare chemoattractant levels at different points along their surface, and consequently adjust their migration. Furthermore, it remains a mystery how responses to gradients depend on the magnitude of the gradient at any given point and the ambient attractant concentration at that point (Herzmark, *et al.*, 2007). Experimental investigations with neutrophil-like cells (Herzmark, *et al.*, 2007) have helped elucidate some of these questions. They report chemotactic prowess depends on both the variation in attractant concentration in space and the ambient concentration of attractant. They also observed that cells were able to sense differences in attractant concentrations across their diameters in relatively shallow gradients, and of translating purely spatial cues into a decision of whether or not to migrate. Finally, they report that chemotaxing cells behave as if they interpret gradients primarily with the use of saturable receptors to assess differences in attractant concentration across their own diameters.

A Cell Modeled as a “Shape Machine”

In a recent study, Satulovsky *et al.* (2008) develop a two-dimensional model for cell motility based on the LEGI concept, where each cell is treated as a collection of points representing the cell’s perimeter and the cell’s center is calculated as the centroid of the perimeter points at a given time point. More specifically, the perimeter of each cell is discretized into 360 points, where each point tracks a 1° arc relative to the cell’s center (See Figure 7.1). Each of the 360 perimeter points is free to protrude (extend away from the center) or retract (curtail towards the center) as a function of a net balance between local signals stimulating protrusion and retraction. A noteworthy limitation of this framework is that cell perimeter points are not constrained to conserve volume or mass, thus allowing for the possibility of unrealistic morphologies; a minimum radius is defined as a crude approximation to a cell’s incompressibility and acts as an initial approximation for conservation of mass. However, this modeling structure is relatively flexible, easy to implement, and is able to capture a wide range of realistic two-dimensional cell morphologies while providing a means to provide a framework to efficiently integrate the functional aspects of actin dynamics and their impact on membrane extension-retraction dynamics, from which cell motility emerges. In this study, we develop our agent-based model of cell motility with consideration of receptor-ligand dynamics upon the framework proposed by Satulovsky *et al.* (2008).

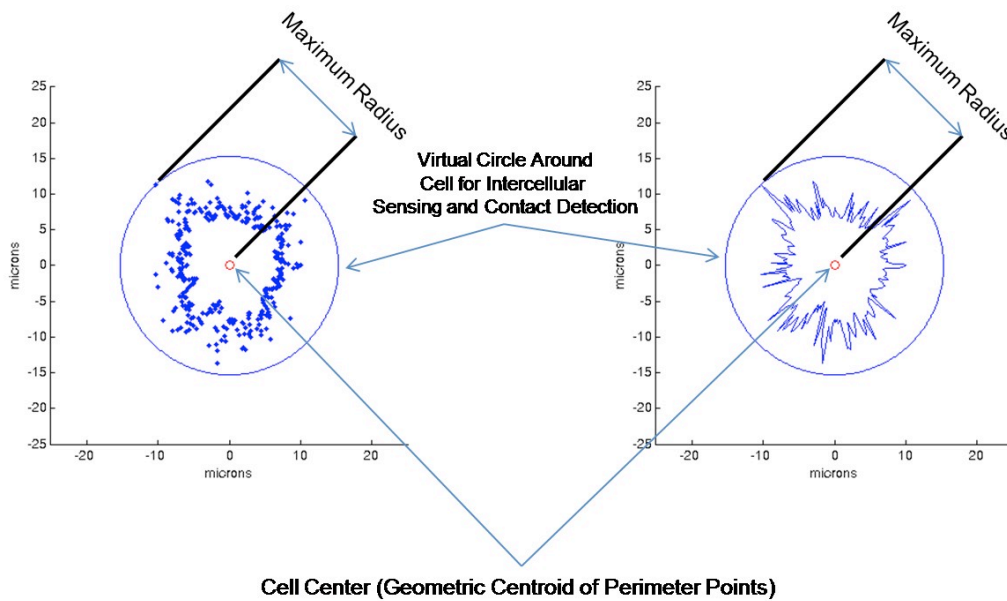


Figure 7.1: Modeling cells as shape machines.

Cell Model Details

Model of Cell Shape as a Function of a Protruding and Retracting Membrane

In this subsection, we comprehensively review the modeling details originally presented by Satulovsky *et al.* (2008) for two-dimensionally treating a cell's shape as a function of protrusions and retractions of the actin cytoskeleton near the cell membrane. At a given time point, t , each cell's perimeter points are represented as vectors \mathbf{r} with the cell center (calculated as the geometric center of the perimeter points) as the origin. A local protrusion signal, $S^+(\mathbf{r}, t)$ is tracked for each perimeter point, and is meant to represent the collective signaling activities stimulating actin polymerization responsible for lamellipodia and filopodia formations. Following the theme of LEGI, retraction signals, $S^-(t)$, are treated as a global cell variable that suppress protrusion signaling and promote retractions

throughout the cell, and effectively represent myosin-dependent contractions that are critical for cells to migrate. The net balance between protrusion and retraction signals determines the evolution each cell's perimeter: retraction of a perimeter point towards the cell center occurs when $S^+(\mathbf{r}, |) < S^-(|)$, and protrusion of a perimeter

point away from the cell center occurs when $S^+(\mathbf{r}, |) > S^-(|)$. The rate of retraction is represented as $\frac{\partial |\mathbf{r}|}{\partial t} = \max\left(\left[|\mathbf{r}| - r_{\min}\right] R^-, 0\right)$ where r_{\min} is a constant minimum radius that accounts for a cell size at which point is incompressible, and R^- is a retraction rate constant. The linear dependence of the retraction rate on $|\mathbf{r}|$ implies elastic behavior of the cytoplasm. The rate of protrusion is represented as $\frac{\partial |\mathbf{r}|}{\partial t} = \max\left(G(R^+), 0\right)$ where $G(R^+)$ is a Gaussian distribution function of average value R^+ and variance R^+ , where R^+ is the average protrusion rate of the cell boundary. The stochastic nature of this protrusion rate is meant to capture variable rates of actin polymerization that depend on local availability of reactants.

Focal adhesions between the cell perimeter and the extracellular matrix (ECM) are modeled as constraints where cell retraction cannot occur as long as the adhesion bond is still intact. This is meant to capture experimental evidence that focal adhesions mediated by integrin receptors form near the leading edge of cells during protrusion, and remain attached as cells move forward. At each protruding point along the cell perimeter, focal adhesions are modeled to form at a probability of P_{FA}^+ , and disassemble with a probability of P_{FA}^- .

Local protrusion signals at a given time point are calculated by

$$\frac{\partial S^+(\mathbf{r}, |)}{\partial t} = \nabla^2 S^+(\mathbf{r}, |)K + \max\left(G\left(\left[\frac{S^+(\mathbf{r}, |)}{S^-(|)} - 1\right], \sigma, | \right) \frac{1}{N}, 0\right)$$

diffuse
decay
baseline burst

, which consists of deterministic terms describing the propagation and decay of

protrusion signals along the cell periphery (approximation of the diffusion of signaling proteins within the cytoplasm), and a stochastic term accounting for local stimulation and generation of new signals via a positive feedback loop. For computational simplicity, the

propagation term (diffusion of signals) is modeled as one-dimensional diffusion along the cell perimeter, though the actual process may occur across the more, if not the entire, cell via multiple signaling pathways affecting actin cytoskeletal reorganizations at the cell surface. The decay term accounts for processes that dampen signals such as dephosphorylation events, diminished energy levels, proteolysis, etc. The feedback loop is implemented to generate more protrusion signals in protruding regions because evidence suggests that signaling cascades for protrusion act as amplifiers of the stimulus, in addition to transducers (Ma, *et al.*, 2004). The protrusion signals are generated with an average magnitude of N_{burst} at an average rate of bursts expressed as $f(S^+(\mathbf{r}, |) - S(|), \odot, |) + P_{baseline}$, where the piecewise linear function $f(x, \odot, |)$ models

the positive feedback: if $x < |$ then $f(x, \odot, |) = 0$; and if $x \geq |$ then $f(x, \odot, |) = [x - |]\odot$ where $|$ represents the takeoff point for the feedback and \odot is the slope of the response. $P_{baseline}$ is a parameter controlling the rate of signal bursts due to internal baseline cell activities. This parameter encompasses the downstream signaling of receptor-ligand binding at the surface which ultimately leads to protrusions. Stochasticity in the generation of signals is controlled using a Gaussian distribution function $G(a)$ where a is both the mean and the variance.

Retraction signals at any given timepoint, $|$, are specified by a global inhibition rule that is a function of the integration of the protrusion signals over the cell perimeter, the cell area (A), and an inhibition constant (C^-): $S^-(|) = C^- A \oint S^+(\mathbf{r}, |) dr$.

While Satulovsky *et al.* (2008) focus the development of their model and calibration of model parameters towards simulating the migratory behavior of Dictyostelium, they also report how changes in their model's parameter values and the additional consideration of focal adhesions exhibited model behavior similar to that of fibroblasts, neurons, and keratocytes. Here, we initiate model parameter values with those suggested to be more representative of mammalian cell types (Satulovsky, *et al.*, 2008). For a complete list of model parameters and their values, refer to Table 7.1.

Table 7.1: Cell Migration Model Parameters.

These parameter values are used in the simulations presented in this study. Note that some of the parameter values were suggested by Satulovsky *et al.* (2008).

Parameter	Definition	Biological Meaning	Suggested Value
$K_{diffuse}$	Diffusion constant of stimulus along membrane; Signal change driven by concentration gradients	Transport of factors that promote actin polymerization and protrusion approximated as a 1D diffusion process along the periphery	$1.91 \times 10^1 \mu\text{m}^2 \cdot \text{sec}^{-1}$ (Satulovsky, <i>et al.</i> , 2008)
K_{decay}	Percentage decay of stimulus per cycle; fraction of signal decrease per unit time	Rate of deactivation of signals for actin polymerization, <i>e.g.</i> , through dephosphorylation, proteolysis, ligand dissociation, or depletion	$2.42 \times 10^{-2} \text{sec}^{-1}$ (Satulovsky, <i>et al.</i> , 2008)
$P_{baseline}$	Average burst rate of signals; Number of pulses per unit time per unit length at maximum fraction of occupied receptors	Average rate of spontaneous generation of signals that promote actin polymerization	1.81×10^{-2} (Satulovsky, <i>et al.</i> , 2008)
N_{burst}	Average magnitude of each generated protrusive pulse	Magnitude of spontaneous signals that promote actin polymerization and protrusion	1.30×10^1 (Satulovsky, <i>et al.</i> , 2008)
C	Inhibitor concentration per unit; concentration of retraction signals generated per unit spread area per unit integrated protrusion signals	Response of signals that promote global cell retraction, such as RhoGTPases and phosphatases like PTP-Pest, in relation to protrusive activities and spreading area	$1.30 \times 10^{-5} \mu\text{m}^{-3}$ (Satulovsky, <i>et al.</i> , 2008)
R^+	Average rate of increase in radius protrusion	Average rate of pseudopodium extension	$1.03 \times 10^{-1} \mu\text{m} \cdot \text{sec}^{-1}$ (Satulovsky, <i>et al.</i> , 2008)
R^-	Rate of fraction decrease in radius during retraction	Rate of retraction, in relation to distance from cell center; Linear dependence on radius implies elastic behavior of cytoplasm	$2.81 \times 10^{-2} \text{sec}^{-1}$ (Satulovsky, <i>et al.</i> , 2008)
\odot	Slope of positive feedback curve	Rate of new signals stimulated by existing	$2.91 \times 10^1 \text{sec}^{-1}$ (Satulovsky, <i>et al.</i> , 2008)

		net signals	
λ	Feedback curve take off point; X-intercept of the positive feedback curve	Minimal protrusive signals required to activate the feedback loop that promotes actin polymerization and protrusion	$3.22 \mu\text{m}^{-1}$ (Satulovsky, <i>et al.</i> , 2008)
R_{Minimum}	Minimum distance that any one perimeter point can retract towards the cell center	Maintains a minimum radius for the cell; though a cell can morphologically alter itself, the minimum radius defines the minimum cell size beyond which the cell is incompressible.	$5 \mu\text{m}$
p^{+FA}	Probability of formation of a focal adhesion per unit time	Probability of the formation of a focal adhesion in an extending perimeter point of the cell	0.015sec^{-1}
P_{FA}^{-}	Probability of detachment of an already formed focal adhesion per unit time	Probability of the detachment of a focal adhesion between the cell perimeter and the environment	$P_{FA}^{-} = 1 - e^{\left(-\frac{\log(2)}{T_{\text{half}}}\right)} = 0.023$ (based on assumption of bond half life, $T_{\text{half}} = 30 \text{sec}$) (Satulovsky, <i>et al.</i> , 2008)
K_D	Receptor-Ligand disassociation constant at chemical equilibrium	Characterizes the dynamics of Receptor-Ligand disassociation and is the ratio between the disassociation rate constant and association rate constant	Neutrophils (FPR-fMLP): 10 nM (Herzmark, <i>et al.</i> , 2007) Breast Carcinoma Cells (EGFR-EGF): 10 nM (Schlessinger, 2002)
A	Phenomenological parameter in Hill-type Sigmoidal Curve	Controls shape of the Sigmoidal curve describing effect of receptor-ligand binding dynamics on internal signaling	0.7
m	Phenomenological parameter in Hill-type Sigmoidal Curve	Controls shape of the Sigmoidal curve describing effect of receptor-ligand binding dynamics on internal signaling	2.46
z	Phenomenological parameter in Hill-type Sigmoidal Curve	Controls displacement of the Sigmoidal curve along the y-coordinates	0.0

Model of the Receptor-Ligand Dynamics Leading to Intracellular Signaling

The development of an all-encompassing protein signaling model governing cellular phenotypic decisions and behaviors, such as cell motility, is a highly complex and daunting task for a number of reasons. Primarily, it is a difficult undertaking because many functional protein interactions are not well-understood, and many more are left to be determined. Therefore, models of this sort are typically developed based on a highly-focused set of key, well-reported interacting molecular species. Deisboeck and coworkers have been developing an agent-based cancer model whose phenotypic decision to migrate or proliferate hinges on the activation of the PLC γ and Raf-MEK-ERK pathways downstream of EGFR-ligand binding events. Experimental evidence in various cancer cell lines indicate that PLC γ and ERK activation downstream of EGFR signaling pathways correlate with a cell's migratory and proliferative behavior; PLC γ has been shown to preferentially activate towards the leading edge of migrating cells (Dittmar, *et al.*, 2002; Mouneimne, *et al.*, 2004), while the involvement of the Ras-MEK-ERK pathways in cell proliferation has been well established (Carey, *et al.*, 2007; Friday & Adjei, 2008; Hilger, Scheulen, & Strumberg, 2002; K. K. Wong, 2009). Deisboeck and coworkers model a cell's response to their microenvironment by way of explicitly modeling the signal transduction network, with a system of differential equations, of interacting genes and proteins stimulated by EGFR-ligand binding and tracking downstream PLC γ and ERK concentrations with respect to time (C. Athale, *et al.*, 2005; Z. Wang, Birch, & Deisboeck, 2008; Z. Wang, *et al.*, 2007).

Rather than explicitly model a signal transduction network of interacting molecules, Satulovsky *et al.* (2008) apply an alternative approach where the intracellular signals responsible for cell migration are modeled as randomized local bursts of

protrusion signals that diffuse along the cell periphery and are globally inhibited by retraction signals. We introduce a representation of receptor-ligand binding mediating directed cell motility by biasing the generation of local protrusion signals at each cell perimeter point as a function of local ligand concentration.

As a step towards considering the effect receptor-ligand dynamics on cell migration, we introduce a functional relationship between the generation of protrusion signals, P_{baseline} , at each cell perimeter point and receptor occupancy by ligand(s) based on the concept of receptor occupancy theory (Limbird, 2005), where a response to receptor-ligand binding events at chemical equilibrium is modeled as a function of occupied receptors (FRO). We use a simple ligand-receptor binding model to calculate the FRO at equilibrium:
$$FRO = \left[\frac{C}{C + k_D} \right]_i$$
 (Herzmark, *et al.*, 2007; Limbird, 2005; S. J. Wang, Saadi,

Lin, Minh-Canh Nguyen, & Li Jeon, 2004) where $i=1$ to 360 for each perimeter point. C , at each perimeter point, is the ambient ligand concentration (*e.g.*, [EGF], [fMLP]) experienced by the entire cell perimeter. In our modeling of neutrophils responding to gradients of Formly-Methionyl-Leucly-Phenylalanine (fMLP) binding to fMLP Receptor (FPR), we use $k_D = 10$ nM according to binding studies from the literature (Herzmark, *et al.*, 2007; Quehenberger, Prossnitz, Cavanagh, Cochrane, & Ye, 1993). Likewise, in our modeling of breast carcinoma cells responding to EGF gradients binding to EGFR, we use $k_D = 10$ nM (Schlessinger, 2002). Biologically, the use of a model based on receptor occupancy theory simulates a cell's receptor-mediated activity, such as motility, acclimating to the immediate microenvironment; Satulovsky *et al.* (2008) report a parameter value for P_{baseline} that simulates the rate of generation of new signals promoting actin protrusions at the cell perimeter, and consequently cell motility; we modify the parameter's value according to FRO, and refer to it as $P_{\text{protrusion}}(\text{FRO})$. According to the

model, protrusions occur at points along the cell periphery where the net protrusive signals (taking into account new signals, diffusing signals, and signals generated due to the positive feedback loop) outweigh the global inhibitory signal. A sensitivity analysis of the model, where the effect of perturbing $P_{\text{protrusion}}$ on the leading cell edge on the motile behavior of simulated cells, demonstrates that local decreases in the parameter value of $P_{\text{protrusion}}$ as a function of FRO lead to a net increase in local protrusion signals. Thus, there is an inverse relationship between the value of $P_{\text{protrusion}}(\text{FRO})$ and protrusive activities.

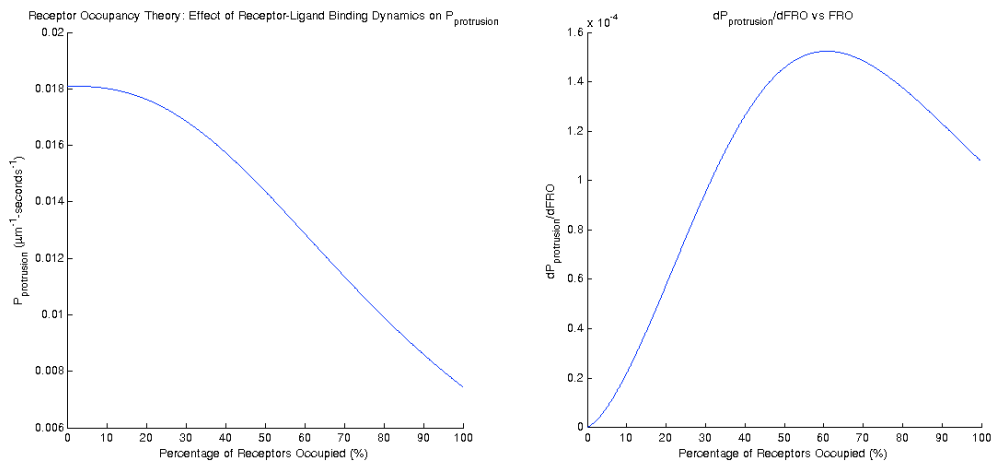


Figure 7.2: Application of Receptor Occupancy Theory: Effect of Receptor-Ligand Binding Dynamics on Signal Generation.

LEFT: Sigmoidal curve that describes how changes in receptor occupancy affect the rate new signal generation, $P_{\text{protrusion}}$, which has an inverse relationship with net protrusive signaling activity. The parameters describing the shape of the curve can be tuned according to a particular cell line. The parameter values used in this report are available in Table 7.1. RIGHT: Plot of the |slope| of the curve from the left panel shows that in order to recapitulate the chemotaxis results from Herzmark *et al.* (2007), the peak difference in signal generation activity across a cell must occur at approximately 50% receptor occupancy, or near an ambient ligand concentration equal to the dissociation constant, K_D .

We take advantage of this inverse relationship by linking FRO with the net activity of protrusion signals at each simulated cell's perimeter points. Intuitively, it follows that with less FRO, the cell's local protrusion signals are lower, and with more

FRO, the cell's local protrusion signals are higher. In order to achieve this relationship, we assume a sigmoidal relationship between FRO and $P_{\text{protrusion}}$ until FRO reaches a high enough capacity, the resulting $P_{\text{protrusion}}$ is relatively low, and once receptor-ligand binding nears saturation, the resulting $P_{\text{protrusion}}$ peaks and does not change significantly with FRO (see Figure 7.2). We apply a Hill-type curve (El-Kareh & Secomb, 2003, 2005) to model this sigmoidal behavior, and may be represented as $P = \frac{P_{\text{baseline}}}{1 + A^{-1}x^m} + z$, where parameters A , m , and z are control the shape and magnitude of the curve describing the functional relationship; we expect that these parameters may be fit according to experiments or theoretical signaling models of comparing receptor-ligand binding with internal signaling activities, perhaps a signaling marker such as PLC β (C. Athale, *et al.*, 2005; Dittmar, *et al.*, 2002; Mouneimne, *et al.*, 2004; Z. Wang, *et al.*, 2007) or PIP3 (Servant, *et al.*, 2000). We choose to maintain the rate of generation of protrusion signals by at P_{baseline} when $\text{FRO} = 0$ to represent baseline cell activity without the presence of ligand in the microenvironment.

Other Modeling Considerations: Agent-Based Framework, Cell-Cell Sensing, and Proliferation

An agent-based framework is well-suited for the object-oriented programming paradigm. Satulovsky *et al.* (2008) have made available to the public a basic implementation of a single cell with their model (source code written in C using a structured programming approach). We built an object-oriented framework atop this source code in C++ so as to readily call for multiple cells for simulating various experimental conditions *in vitro* and *in vivo*. In order to account for cell-to-cell contact, we implemented a contact detection algorithm (Hogue, 1998) to determine whether or not two or more cells are in contact in the virtual space. Without the consideration of cell

contact, the virtual cells would have no sense of one another's presence and simply pass through one another.

At each time step, we iterate through the cell list data structure and exhaustively calculate the Euclidian distance between each of the virtual cells. If $\text{distance}_{ij} > (r_i + r_j)$, where r is the maximum radius of the cell, then the two cells are not interacting with each other. However, if $\text{distance}_{ij} < |r_i - r_j|$, then the two cells in question are overlapping. This situation occurs following mitosis in our model, when the two daughter cells are placed at the same cell center coordinates as their parent cell. This situation also occurs during the initializing of simulations when cells may be randomly seeded close to each other. Also this situation can occur when two motile cells run into each other before avoiding each other. In our model, when two cells overlap, we implement a rule where one of the two cells is randomly chosen to move adjacent to the other cell by a distance equivalent to its maximum radius (at that time point) from the other cell's center; the angular position relative to the first cell's center to which the second cell is moved is stochastically chosen. In order to avoid cells from running over each other, we implement a rule that cells in contact cannot experience protrusion activity, regardless of the state of intracellular signaling, in the cellular region where the two cells are in contact. A virtual circle around each cell (with center at the cell's center and radius defined by the perimeter point farthest away the cell center) is defined and this region of no protrusion activity is treated as the intersection of the virtual circles (see Figure 7.3). In this region, each cell experiences retraction either until there is no more interaction or the perimeter point in question has retracted to the minimum radius, and the inherent bias of protrusion signals on the non-interacting sides of the cells mobilizes the cells away from each other.

We treat mitotic events as a simple probability rule. The HL-60 and MDA-MB-231 cell lines both have a doubling time of approximately 25 hours (Collins, 1987;

Korah, Sysounthone, Golowa, & Wieder, 2000). Since our simulation time steps are on the order of 1 per second, we estimate the probability of mitosis per timestep as $\frac{1 \text{ mitosis}}{(25 \text{ hours}) \left(\frac{3600 \text{ seconds}}{\text{hour}} \right)} = 1.11e-5$. We note that the duration of the simulations in

this report are on the order of few hours, thus mitosis is not an immediate concern. Mitosis has been treated more rigorously in prior agent-based cell models running simulations on the order of days and months, where the state of each cell within the cell cycle is tracked as a probability function of time (Macklin, *et al.*, 2009) or a protein signaling model elicits proliferation by tracking activation of well-reported proliferation-related pathways such as Raf-MEK-ERK pathway (Z. Wang, *et al.*, 2007) and levels of p27 (Alarcon, Byrne, & Maini, 2004).

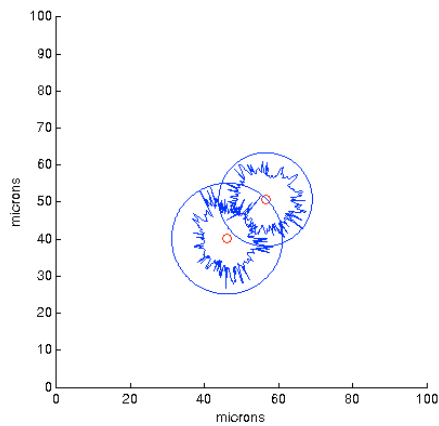


Figure 7.3: Two cells Interacting and Responding via Contact Detection Scheme.

This is an example of two cells interacting with each other during a simulation. The contact detection algorithm determines the region of intersection of the two virtual circles surrounding the cells, and subsequently introduces retraction of the cells' perimeter points in the intersecting region.

Simulation of In Vitro Microenvironments

Experimental biologists utilize various cell-based assays in combination with time-lapse microscopy to monitor cell behavior in controlled environments. A recent trend in the cell motility community has been to perform *in vitro* experimental studies using microfabricated chemotaxis chambers (Herzmark, *et al.*, 2007; Lin, *et al.*, 2005; Mosadegh, Saadi, Wang, & Jeon, 2008; Roman Zantl, 2006; Saadi, *et al.*, 2007; Saadi, Wang, Lin, & Jeon, 2006; S. J. Wang, *et al.*, 2004; C. M. Wells & Ridley, 2005). This type of experimental device in conjunction with time-lapse microscopy is a powerful tool for investigating the response of cells to chemoattractants in real-time. Cells are seeded on a specific portion of a glass slide (we refer to as the channel) where the gradient of a chemoattractant such as EGF or fMLP can be finely controlled and stably maintained. The glass slide can also be coated with adhesion factors such as Collagen IV and Fibronectin to promote adhesion and better simulate *in vivo* extracellular matrix (ECM) conditions.

For the purposes of comparisons to experimental studies from literature (Herzmark, *et al.*, 2007; S. J. Wang, *et al.*, 2004) in this paper, we set up our two-dimensional virtual microenvironment as a lattice mesh N microns x M microns (See Table 7.2 for specifics). Simulated cells are not constrained to move on the lattice mesh points, however we assign the concentration of chemoattractant at each lattice point according to the gradient profile in question. Wang *et al.* (2004) studied chemotaxis of MDA-MB-231 breast cancer cells under uniform, linear ($y = mx$) and polynomial ($y = Cx^{4.2}$) gradients, where m and C are parameters whose value depend on the difference of EGF concentration across the channel (See Table 7.2). Herzmark *et al.* (2007) investigated the chemotactic behavior of differentiated HL60 neutrophil-like cells under

linear ($y = mx$) and exponential ($y = V_0 e^{k(x-x_{\text{source}})}$) gradient conditions, where V_0 is the maximum ligand concentration at the source, k is the decay rate of the ligand, and x_{source} is the x-coordinate of the source of the ligand on one side of the channel. We initiate the cell motility model with the same microenvironmental conditions as these chemotaxis experiments, and compare the results of our simulations with them in order to assess the utility of model as a predictive *in silico* tool for studying cell motility.

Table 7.2: Microenvironmental Conditions in Chemotaxis Chamber.

Simulating the gradient profiles considered in Herzmark *et al.* (2007) with dHL60 neutrophils and Wang *et al.* (2004) with MDA-MB-231 breast cancer cells. The chemotaxis chambers used in Herzmark *et al.* (2007) and Wang *et al.* (2004) is simulated on $N \times M$ lattices with the concentration of ligand calculated at each lattice point.

Profile	Lattice	Model	Parameter Values	Simulating Conditions
Uniform	400 μm x 400 μm	$y = V_0$	$V_0 = [0 \ 3.906 \ 7.812 \ 15.625]$ nM	Herzmark <i>et al.</i> (2007)
Linear	400 μm x 400 μm	$y = (V_0/\Delta x)x$	$V_0 = 45$ nM $\Delta x = 400$ μm	Herzmark <i>et al.</i> (2007)
	350 μm x 350 μm		$V_0 = [0 \ 3.906 \ 7.812 \ 15.625]$ nM $\Delta x = 350$ μm	Wang <i>et al.</i> (2004)
Power	400 μm x 400 μm	$y = Cx^{4.2}$	$C = [0.046\text{e-}9 \ 0.092\text{e-}9 \ 0.184\text{e-}9]$	Wang <i>et al.</i> (2004)
Exponential	1000 μm x 1000 μm	$y = V_0 e^{k(x-x_{\text{source}})}$	$V_0 = 100$ nM $x_{\text{source}} = 400$ μm	Herzmark <i>et al.</i> (2007)

PROTOTYPE MODEL PERFORMANCE: COMPARISON TO IN VITRO EXPERIMENTS FROM LITERATURE

Chemotactic experiments with both neutrophil-like (Herzmark, *et al.*, 2007) and breast carcinoma (S. J. Wang, *et al.*, 2004) cells have reported differential migratory responses to chemoattractants that appear to depend on the steepness of the gradient profile. Cells exhibit weak (or no) chemotactic response under linear chemoattractant

profiles, as measured by Chemotactic Index metric, and exhibit noticeably stronger chemotactic responses under nonlinear chemoattractant profiles. These reports point to the differences in ligand concentration a cell experiences across its exposed surface, ΔC , relative to the average ambient ligand concentration, C , experienced by the cell as indicative of this differential behavior. Under linear profiles, a cell experiences a constant ΔC between its front and rear edges, whereas C increases linearly as the cell moves up the ligand gradient (see Figure 7.4). Thus, the fractional difference in concentration experienced across the cell, $\frac{\Delta C}{C}$, decreases as a cell moves up the linear gradient. Under exponential profiles (Herzmark, *et al.*, 2007), ΔC is proportional to C ; thus $\frac{\Delta C}{C}$ remains constant as a cell moves up the gradient (see Figure 7.5). We use the experimental results of dHL-60 neutrophil-like cells behavior when exposed to linear and exponential gradients of fMLP as the basis for initially training and assessing our prototype cell motility model. We subsequently show that the *trained* model can predict the differential behavior seen in MDA-MB-231 breast cancer cells responding to gradients of EGF.

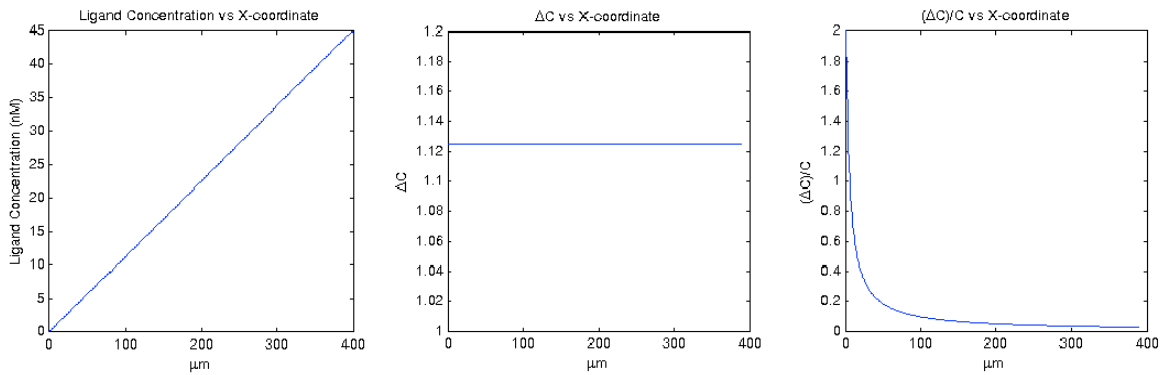


Figure 7.4: Linear Profile in Simulated Chemotaxis Chamber.

LEFT: Plot of linear gradient of fMLP across the simulated chemotaxis chamber. MIDDLE: Difference in ambient fMLP concentration, ΔC , across a “typical” size of diameter 10 μm versus position across the chemotaxis chamber. RIGHT: the fractional difference of fMLP concentration experienced across the cell, $\frac{\Delta C}{C}$, versus position across the chemotaxis chamber.

Calibration of Receptor Occupancy Theory-based Signaling Model

Herzmark *et al.* (2007) investigate the chemotactic behavior of dHL-60 neutrophil-like cells exposed for 10 minutes to a linear fMLP gradient with maximum of 45 nM (see Figure 7.4) and exponential fMLP gradients with a maximum of 100 nM (see Figure 7.5). The results are interpreted as the chemotactic index (ratio of the distance traveled by a cell towards the up gradient to the total distance traveled by the cell during the same time duration) for each cell in relation to the ambient fMLP concentration of the cell's initial position in the channel, C_{initial} . Under the linear conditions, the experimental results show that the chemotactic index (CI), a measure of chemotactic prowess of a cell, steadily decreases from a maximum of 0.4 to zero (*i.e.*, random motility) as C_{initial} increases (see Figure 7.6). Under the exponential conditions, the experimental results show that the CI steadily increases with C_{initial} it peaks near $C_{\text{initial}} = K_D$, and then steady declines (see Figure 7.7). This peak occurring at $C_{\text{initial}} = K_D$ can be predicted using a simple relationship which relates the difference in FRO across the leading and rear edge of a cell ($\sim 10 \mu\text{m}$ in diameter), ΔFRO , to C_{initial} . (see Figure 7.8). Herzmark *et al.* (2007)

model the difference in receptor occupancy across a cell as

$$\frac{\Delta\text{FRO}}{\text{FRO}} = \frac{C_{\text{front}}}{C_{\text{rear}} + K_D} - \frac{C_{\text{rear}}}{C_{\text{rear}} + K_D} = \frac{K_D \Delta C}{(C_{\text{rear}} + K_D)^2}$$

where by assuming $\Delta C \ll (C_{\text{rear}} + K_D)$, the model simplifies to $\frac{\Delta\text{FRO}}{\text{FRO}} \approx \frac{K_D \Delta C}{(C_{\text{rear}} + K_D)^2}$. Under exponential gradients, the

fractional difference in concentration experienced across the cell, $\frac{\Delta C}{C_{\text{rear}}}$, is constant. By

receptor occupancy theory applied to chemotaxis and according to this simplified model,

it follows that the highest CI occurs when $\frac{K_D \phi C}{(C_{rear} + K_D)^2}$ is maximized, which is when

$$C_{\text{initial}} = C_{\text{rear}} = K_D \text{ (Herzmark, et al., 2007).}$$

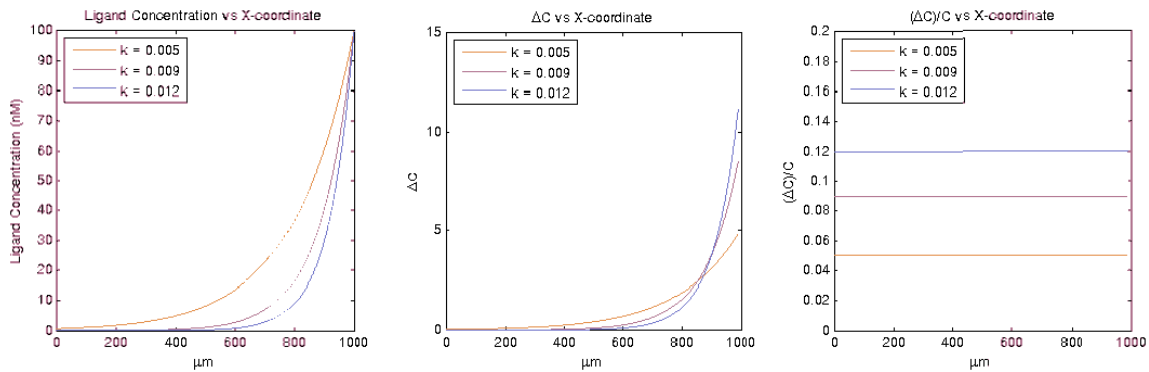


Figure 7.5: Exponential Profile in Simulated Chemotaxis Chamber.

LEFT: Plot of exponential gradient of fMLP across the simulated chemotaxis chamber. MIDDLE: Difference in ambient fMLP concentration, ΔC , across a “typical” size of diameter $10 \mu\text{m}$ versus position across the chemotaxis chamber. RIGHT: the fractional difference of fMLP concentration experienced across the cell, $\frac{\Delta C}{C}$, versus position across the chemotaxis chamber. LEGEND: The k parameter represents

the decay constant in the exponential profile (see Table 7.2).

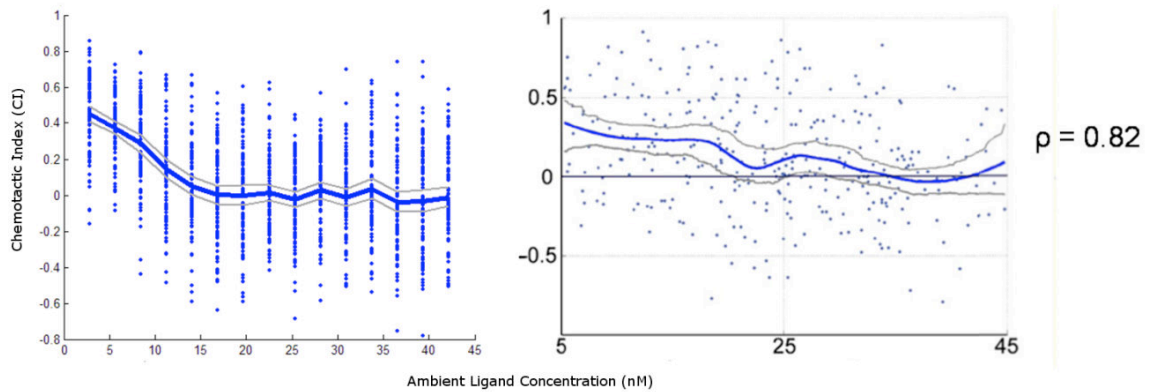


Figure 7.6: Comparison of In Vitro and In Silico Chemotaxis Experiments Under Linear Gradient.

Comparison of the in silico prediction of chemotactic index (CI) vs. starting fMLP concentration (LEFT) against experimental results (Herzmark, et al., 2007) (RIGHT) under the same linear gradient conditions. Note the similar decreasing trend of CI with the initial ligand concentration; this relationship is confirmed

by the Pearson correlation coefficient, ρ , comparing the simulated results from our study with the average trend (blue line) reported by Herzmark *et al.* (2007).. The solid blue line represents the average trend in the data, computed from the CI of individual cells (blue dots). The gray lines represent the 95% confidence intervals, computed from 1000 bootstrap resamples of the data. For the *in silico* results, 50 cells at multiple positions across the chemotaxis chamber corresponding to increasing initial ligand concentrations were simulated. The image in the RIGHT panel was reprinted from Herzmark *et al.* (2007) with kind permission from PNAS.

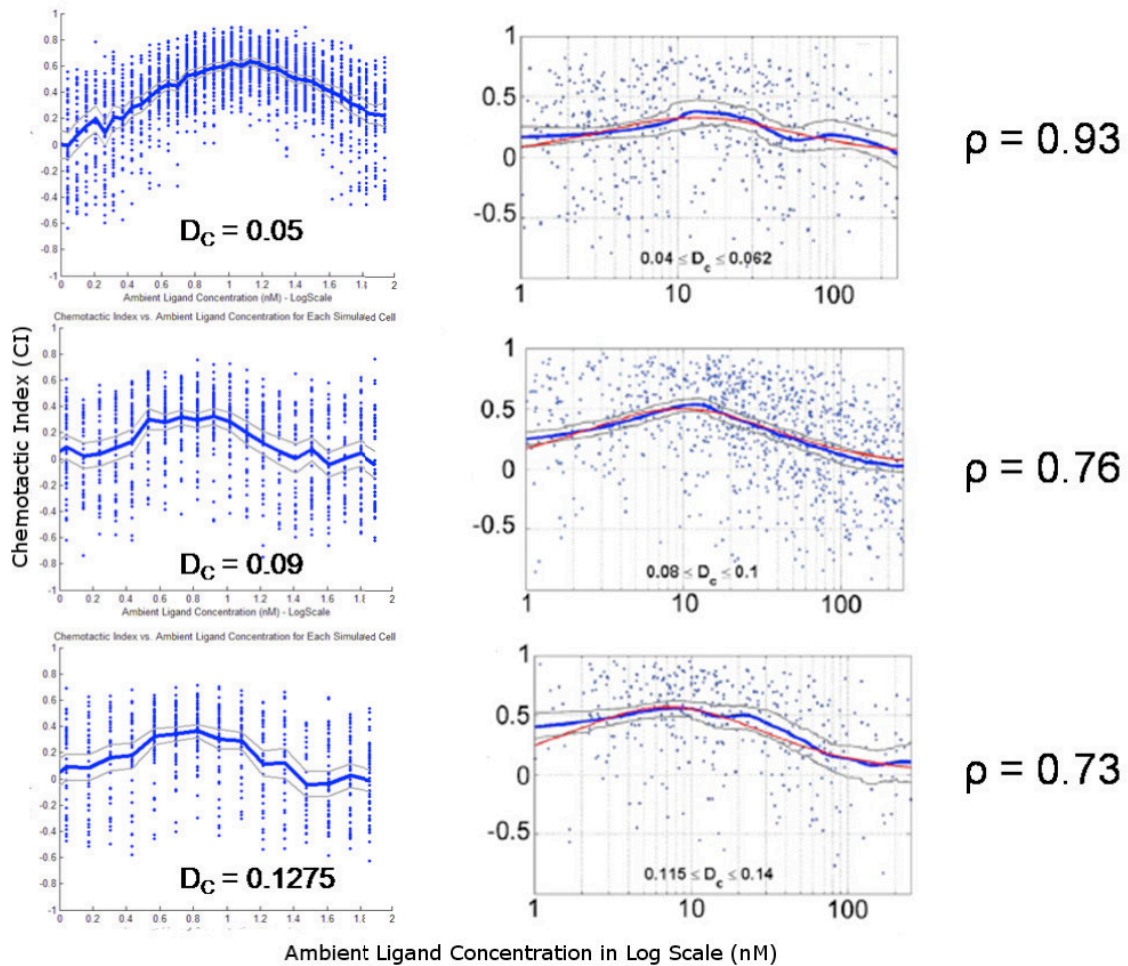


Figure 7.7: Comparison of In Vitro and In Silico Chemotaxis Experiments Under Exponential Gradients.

Comparison of the *in silico* prediction of chemotactic index (CI) vs. starting fMLP concentration (LEFT) against experimental results (Herzmark, *et al.*, 2007) (RIGHT) under the same exponential gradient conditions. Herzmark *et al.* (2007) report their results for three regimes of $D_C = \Delta C/C_{\text{initial}}$. The solid blue line represents the average trend in the data, computed from the CI of individual cells (blue dots). The gray lines represent the 95% confidence intervals, computed from 1000 bootstrap resamples of the data. The red

line is the theoretical relationship, $CI = k(FRO_{front} - FRO_{rear})$, postulated by (Herzmark, *et al.*, 2007), where k is a parameter fit according to experimental data. Note the similar increasing trend of CI as the initial ligand concentration approaches K_D and decreases thereafter; this relationship is confirmed by the Pearson correlation coefficients, ρ , comparing the simulated results from our study with the theoretical relationship (red line) reported by Herzmark *et al.* (2007). For the *in silico* results, 50 cells at multiple positions across the chemotaxis chamber corresponding to increasing initial ligand concentrations were simulated. The x-axis is reported in log10 scale. 476, 1023, and 372 cells were tracked for the three regimes (in order of increasing D_C) reported by Herzmark *et al.* (2007). The images on the RIGHT panel were reprinted from Herzmark *et al.* (2007) with kind permission from PNAS.

Using the experimental data of dHL-60 cells responding to gradients of fMLP as an endpoint for comparison and choosing $K_D = 10$ nM for fMLP-FPR binding dynamics (Herzmark, *et al.*, 2007; Quehenberger, *et al.*, 1993), we vary the shape of the curve linking receptor-ligand binding to intracellular signals leading to actin cytoskeletal reorganization in our model until we recapitulate the cell behavior seen experimentally. Based on dose-response relationships seen in pharmacodynamics (El-Kareh & Secomb, 2003, 2005), we hypothesized that shape of curve to be sigmoidal: at each perimeter point on the cells, intracellular signaling activity is maintained at $FRO = 0\%$ and slowly ramps up as receptors and ligands form complexes, and ultimately peaks as receptors saturate ($FRO \rightarrow 100\%$); this relationship is achieved by manipulating $P_{protrusion}(FRO)$ (see Figure 7.2). Furthermore, in order to recapitulate chemotactic prowess maximizing near $C_{ambient} = K_D$ (as seen experimentally), the slope of the curve must be steepest at this point (see Figure 7.2).

We run chemotaxis simulations using the functional relationship depicted in Figure 7.2 between FRO and the rate of generation of signals driving actin protrusions under linear and exponential gradients (Table 7.2) for 10 minutes. Our results show we are able to recapitulate CI vs. $C_{initial}$ profiles similar to those reported by Herzmark *et al.* (2007), thus supporting the functional link (and its shape) between FRO and intracellular signaling we introduced into our cell motility model (see Figures 7.6 and 7.7) and indicating that our model calibration was successful. We found that the chemotactic

behavior of the cells was sensitive to the shape of the curve, and that the sigmoidal curve shape depicted in Figure 7.2 is necessary to reproduce the dHL-60 chemotactic prowess in response to fMLP gradients.

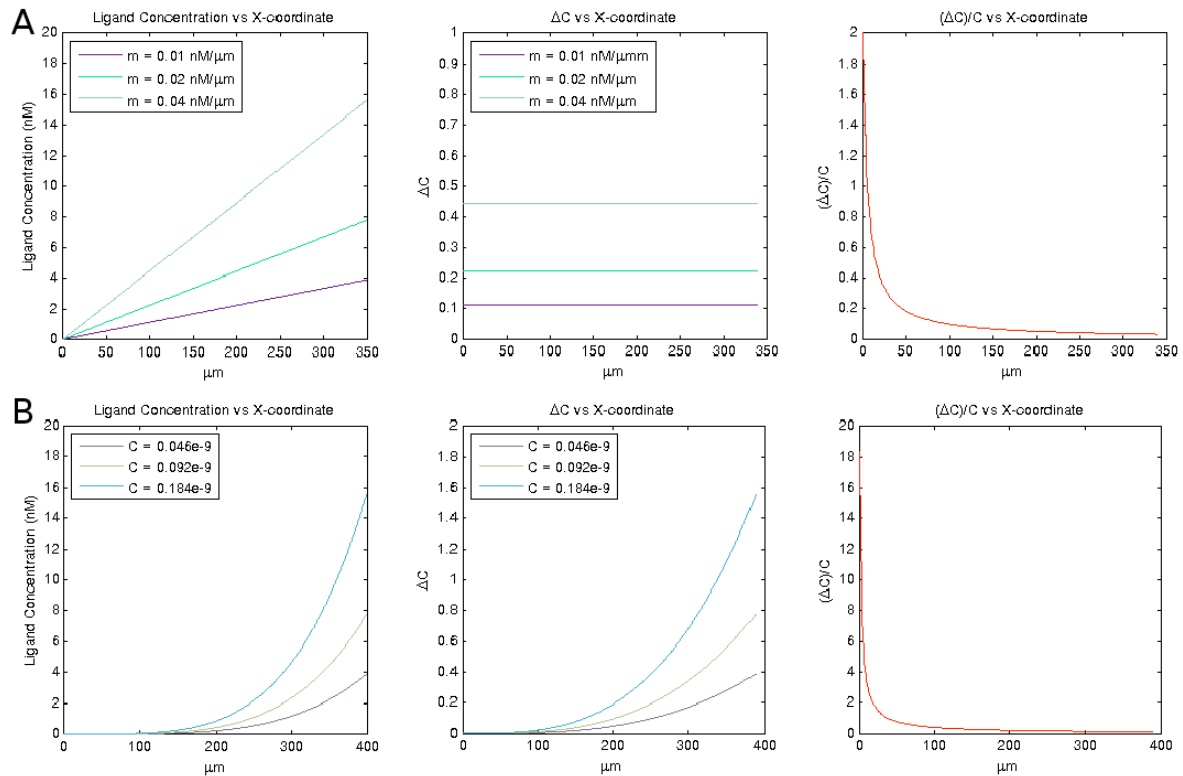


Figure 7.8: Linear and Nonlinear-Power Profiles in Simulated Chemotaxis Chamber for MDA-MB-231 Breast Cancer Cells Responding to EGF.

A) LEFT: Plot of linear gradient of EGF across the simulated chemotaxis chamber. MIDDLE: Difference in ambient EGF concentration, ΔC , across a “typical” size of diameter $10 \mu\text{m}$ versus position across the chemotaxis chamber. RIGHT: the fractional difference of EGF concentration experienced across the cell, $\frac{\Delta C}{C}$, versus position across the chemotaxis chamber; the plot is consistent for this profile regardless of the value of the constant, C. LEGEND: The m parameter represents the slope of the linear profile (see Table 7.2).

B) LEFT: Plot of nonlinear-power gradient of EGF across the simulated chemotaxis chamber. MIDDLE: Difference in ambient EGF concentration, ΔC , across a “typical” size of diameter $10 \mu\text{m}$ versus position across the chemotaxis chamber. RIGHT: the fractional difference of EGF concentration experienced across the cell, $\frac{\Delta C}{C}$, versus position across the chemotaxis chamber; the plot is consistent for this profile regardless of the value of the constant, C. LEGEND: The C parameter represents the constant in the power profile (see Table 7.2).

Validation: Breast Carcinoma Cells Responding to EGF Gradients

Wang *et al.* (2004) performed chemotaxis experiments with MDA-MB-231 breast carcinoma cells under linear and nonlinear gradient conditions of EGF (Table 7.2). As an external test for the motility model, whose link between receptor-ligand binding and intracellular signaling activity we calibrated in the previous section, we simulate the chemotaxis chamber conditions presented in Wang *et al.* (2004) (see Table 7.2), and evaluate the performance of our model in comparison to the experimental results.

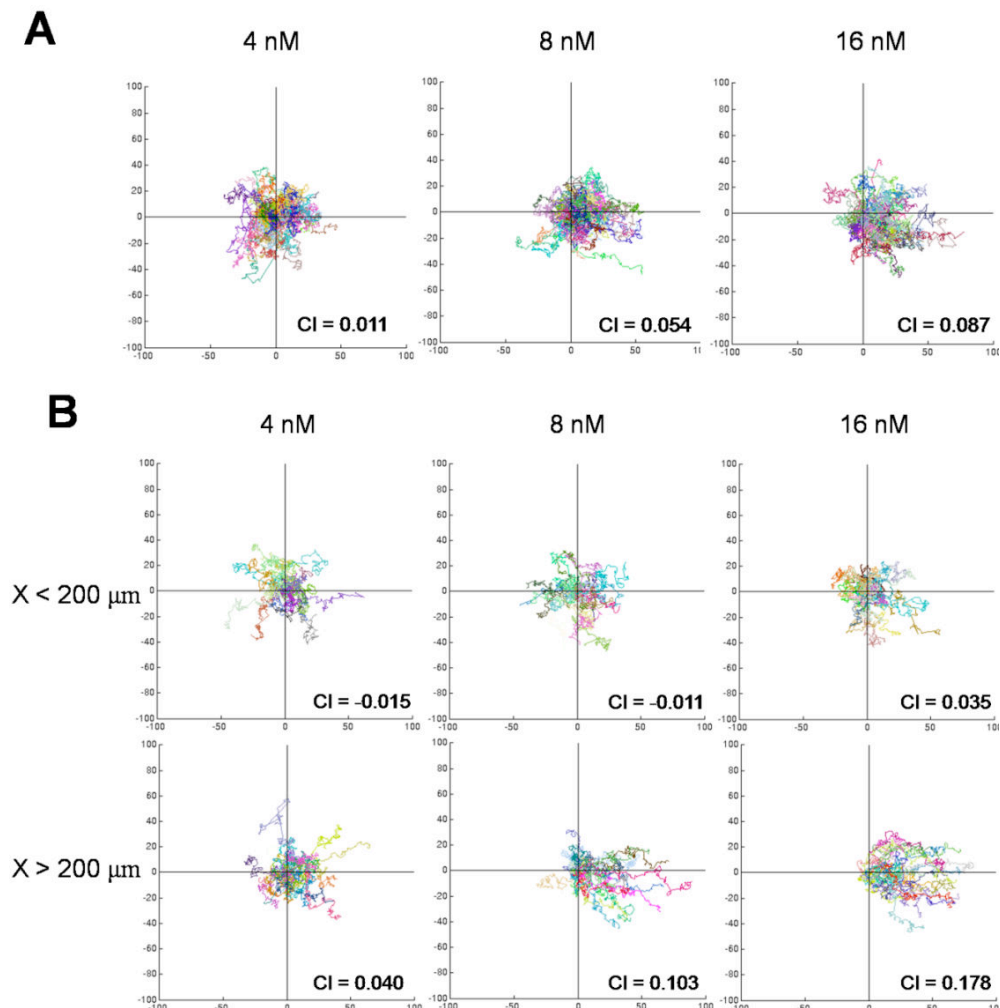


Figure 7.9: Prediction of Differential Motile Behavior of MDA-MB-231 Breast Cancer Cells Responding to EGF.

A) Trajectories of cells experiencing linear gradients with maximum [EGF] of 4, 8, and 16 nM in chemotaxis chamber. B) Trajectories of cells experiencing nonlinear gradients (polynomial) with maximum [EGF] of 4, 8, and 16 nM in chemotaxis chamber. Like Wang *et al.* (2004), the analysis was divided into two regimes: $x < 200 \mu\text{m}$ (where cells experience a shallow gradient) and $x > 200 \mu\text{m}$ (where cells experience a steep gradient). Refer to Table 7.1 and 2 for details regarding the motility parameters and the gradient profiles.

Based on EGF-EGFR binding studies from the literature, we found that K_D ranges from 10 to 20 nM (A. DeWitt, *et al.*, 2002; Kholodenko, *et al.*, 1999; Schlessinger, 2002). In our simulations, we choose $K_D = 10$ nM. Under linear gradient conditions, with varying slopes (see Table 7.2), our simulations agree with the reported experimental results: using CI as the metric for chemotactic prowess, there is no noticeable directed motility and the migration speeds are on the order of $1 \mu\text{m}/\text{min}$ (Figure 7.10). Under nonlinear gradient conditions, with a profile modeled as $y = Cx^{4.2}$ (see Table 7.2), our simulations also agree with the reported experimental results showing the same differential behavior in motility between the first half of the chemotaxis chamber (where the gradient is shallow) and the second half of the chemotaxis chamber (where the gradient is steep) (7.10). We note that unlike linear and exponential gradients used to calibrate the cell motility model's functional link between receptor occupancy and internal signaling, the gradients established by the power relationship $y = Cx^{4.2}$ have neither a constant ΔC (linear) or $\frac{\Delta C}{C}$ (exponential); rather both vary throughout the chemotaxis chamber as a function of distance from the EGF source (see Figure 7.8). By only considering the dissociation constant, K_D , for EGF-EGFR as the free parameter to set according to values reported in the literature, we show that it is possible to predict the experimental results of a *different* cell type responding to a *different* ligand via a *different* receptor experiencing a *different* gradient profile. This is a significant result because we calibrated the shape of the $P_{\text{protrusion}}(\text{FRO})$ curve according to the motility data from

(Herzmark, *et al.*, 2007) under linear gradient conditions, internally validated the curve on motility data from (Herzmark, *et al.*, 2007) under exponential gradient conditions, and externally validated the performance of the model by comparing to motility data from (S. J. Wang, *et al.*, 2004) of a different cell type responding to linear and nonlinear EGF gradients. This computational result is compelling and supports that the mechanisms underlying receptor-ligand binding leading to chemotaxis are indeed conserved across different cell types and that spatial gradient sensing is all that is necessary for cells to successfully migrate preferentially towards gradients of ligand (Herzmark, *et al.*, 2007).

Using a Protein Signaling Model to Justify the Functional Link Between Receptor-Ligand Binding and Signal Generation Leading to Motility

Rather than presume that the shape of the curve describing the relationship between occupied receptors and intracellular motility signaling to be accurate solely on the basis of fitting it so as to reproduce experimental chemotaxis results, we sought to validate the shape of the functional relationship based on a biochemical understanding of the pathways involved in cell motility. To this end, we use the signaling model reported by Wang *et al.* (2007) to simulate lung cancer cells' decisions to migrate or proliferate, to investigate the relationship between EGF-EGFR binding dynamics and the generation of Phospholipase C gamma (PLC γ) under various microenvironmental EGF concentrations.

PLC γ has been shown to preferentially concentrate at the leading edge of migrating cells,

and is used as the molecular signal in the agent-based model reported by Wang *et al.* (2007) to trigger a migratory phenotype. In keeping with the philosophy of our scientific peers, we chose phosphorylated PLC γ to be a molecular signal of migration; we

assume its generation marks increased net protrusive activity necessary for migration. Based on our results thus far, we hypothesize it would also have an inverse relationship with

$P_{\text{protrusion}}(\text{FRO})$. We ran the system of differential equations describing the mass action kinetics of EGF-EGFR binding and downstream generation of phosphorylated PLC \odot for a simulation time duration of 200 seconds with a microenvironmental $[\text{EGF}] = 1 \text{ nM}$, and kept record of $[\text{Phosphorylated PLC}\odot]$ and $[\text{EGFR-EGF}]$ at steady state; we chose 200 seconds as the time duration for the simulation because it allowed enough time for the signal to peak and reach steady state. We repeat this procedure for $[\text{EGF}] = 2, 3, 4, \dots, 1000 \text{ nM}$ of EGF, each time keeping record of the $[\text{Phosphorylated PLC}\odot]$ and $[\text{EGFR-EGF}]$ at steady state. We plot $[\text{Phosphorylated PLC}\odot]_{\text{steady state}}$ (unit normalized) versus the corresponding $[\text{EGFR-EGF}]_{\text{steady state}}$ (unit normalized) for 1 to 1000 nM of $[\text{EGF}]_{\text{microenvironment}}$ (See Figure 7.10). The general shape of the curve matches the sigmoidal features necessary for our model to recapitulate the chemotaxis results. We consider this result, which is founded on a well-calibrated protein signaling model from the literature (Kholodenko, *et al.*, 1999; Z. Wang, *et al.*, 2007; Wiley, *et al.*, 2003), as biological justification for our functional link between receptor-ligand binding dynamics and intracellular signaling governing actin cytoskeletal reorganization, and consequently motility.

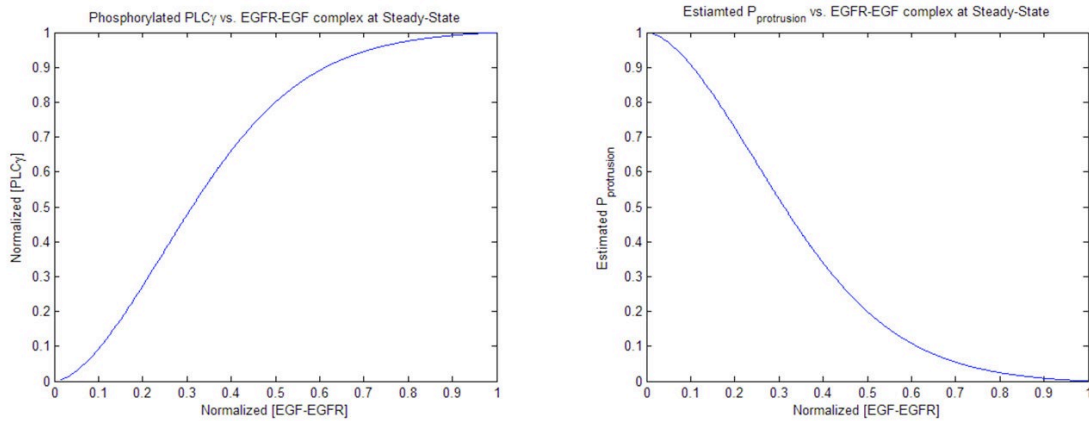


Figure 7.10: Protein Signaling Model Predicts Generation of Phosphorylated PLC γ with respect to EGFR-EGF complex.

LEFT: [Phosphorylated PLC γ] (unit normalized) vs. [EGFR-EGF complex] (unit normalized) at steady state. PLC γ has been shown to preferentially activate towards the leading edge of migrating cells. RIGHT: Due to the inverse relationship between the model parameter $P_{\text{protrusion}}$ and local protrusive activities, we show the perceived relationship between $P_{\text{protrusion}}$ and [EGFR-EGF complex] according to the activation of PLC γ predicted by the protein signaling model (Z. Wang, *et al.*, 2007).

DISCUSSION AND CONCLUSION

Here we have presented a modeling framework for cell motility that takes into account receptor-ligand binding dynamics. We show that it is possible to simulate the chemotactic behavior of neutrophil-like and breast carcinoma cells responding to gradients of fMLP and EGF, respectively, by introducing the concept of receptor occupancy theory to provide a functional link between local chemoattractant levels and internal signaling leading to cytoskeletal restructuring at the cell surface that ultimately drive cell migration. This type of model provides a valuable link for scientists to create and test hypotheses at the molecular scale that impact cellular and tissue scale behavior.

Our simulation results were able to predict, albeit using an abbreviated molecular model for the effect of receptor-ligand binding on internal protrusion signaling, the differential motile behavior of MDA-MB-231 breast cancer cells to polynomial-shaped

EGF gradients, after calibrating the model according to neutrophil-like cells responding to linear- and exponential-shaped fMLP gradients. This is compelling because we calibrated the model on one cell type responding to a linear gradient, and use it to predict behavior under different gradient conditions and in a different cell type. These results provide validation for the capability of the model to predict cell motility under various gradients, and support that the underlying mechanisms of cells responding to gradients of chemokines are conserved across cell types regardless of receptor and ligand so long as an understanding of their binding dynamics is known. Our results also support that gradient sensing is likely achieved in eukaryotic cells by a spatial method, where cells dynamically compare the microenvironment on different locations on their surface.

After testing $P_{\text{protrusion}}(\text{FRO})$ of various shapes, we found that a sigmoidal shaped curve is necessary to recapitulate the correct motility behavior (as defined by chemotactic index vs. ambient ligand concentration profiles) by our model in comparison to our training data. This result led us to hypothesize that the receptor dynamics are indeed sigmoidal. In order to test this hypothesis, we turned to a protein signaling model reported in the literature describing key signaling events downstream of EGFR-EGF binding, where chemical kinetic parameters were experimentally-derived or robustly estimated. Specifically, we compared the shape of the curve describing activity of a marker for cell motility, PLC β , in response to local EGF levels in the microenvironment to the shape of the sigmoidal curve we found necessary to recapitulate the motility behavior seen experimentally. This comparison validates our hypothesis because both curves share the same sigmoidal shape.

We view this study as a first integration of a molecular-scale signaling model into our ABM framework, and as a demonstration of a method for implementing other molecular signaling models in future studies. Furthermore, this study points to the potential of analyzing molecular signaling models to find appropriate “shortcuts” to use

in full agent-based models. This modeling framework provides a tool for testing circumstances and/or hypotheses that may affect cell motility, such as interfering with EGFR dynamics via an inhibitor (*e.g.*, Erlotinib) or the impact of Her2/neu receptors on chemotaxis of cancer cells responding to EGF gradients given that Her2/neu is known to dimerize with the EGFR-EGF complex (W. W. Chen, *et al.*, 2009). Elucidating the role of Her2/neu on motility is a particularly interesting in the case breast cancer, where approximately 30% of invasive breast cancers are known to express the receptor.

MATERIALS & METHODS

We initialized our *in silico* chemotaxis chamber according to the conditions used by Herzmark *et al.* (2007) to investigate chemotaxis prowess of dHL-60 neutrophil-like cells. For each simulation, we introduced fMLP gradients to the microenvironment according to the profiles and parameters listed in Table 7.2. We assumed the simulated cells were surrounded by a bath with sufficient levels of nutrients so to not inhibit the potential for proliferation or induce apoptosis. We simulated the motility of 50 cells initialized along evenly distributed increments across the *in silico* chemotaxis chamber in order to assess the chemotactic prowess of our virtual cells at different initial ambient fMLP concentrations. Simulations were run for 10 simulated minutes with one second timesteps. The chemotactic index (CI) was calculated for each cell as the ratio of its net displacement towards the gradient to its total trajectory. The average trend in the data was computed from the chemotactic index of the individual cells. The 95% confidence intervals depicted in Figures 7.6 and 7.7 were computed from 1000 bootstrap resamples of the data.

Subsequently, we initialized our *in silico* chemotaxis chamber according to the conditions used by Wang *et al.* (2004) to study chemotaxis of MDA-MB-231 breast carcinoma cells responding to EGF gradients. We simulated linear and polynomial

gradients of EGF according to the profile and parameters listed in Table 7.2. We seeded the chemotaxis chamber with 49 cells evenly distributed across the two-dimensional Cartesian plane. Simulations were run for three simulated hours with one second timesteps.

The cell motility model was written in object-oriented C++. Simulations were run on a high performance server with two 3.16 GHz quadcore Intel Xeon processors, 2x6 MB of L2 cache and Core2 architecture, 1333 MHz FSB, and 16 GB of RAM. On average, the ratio of processing time to simulation time is $1.2e-3$ per cell.

The data analysis and plotting were performed using MATLAB 7.1.

ACKNOWLEDGEMENTS

This work was funded by the Cullen Trust for Healthcare and institutional funds from the University of Texas Health Science Center in Houston, TX, USA. We also thank Mr. Charles Tate for his kind support.

Appendix

CHAPTER 6 SUPPLEMENTARY FIGURES AND TABLES

Figures

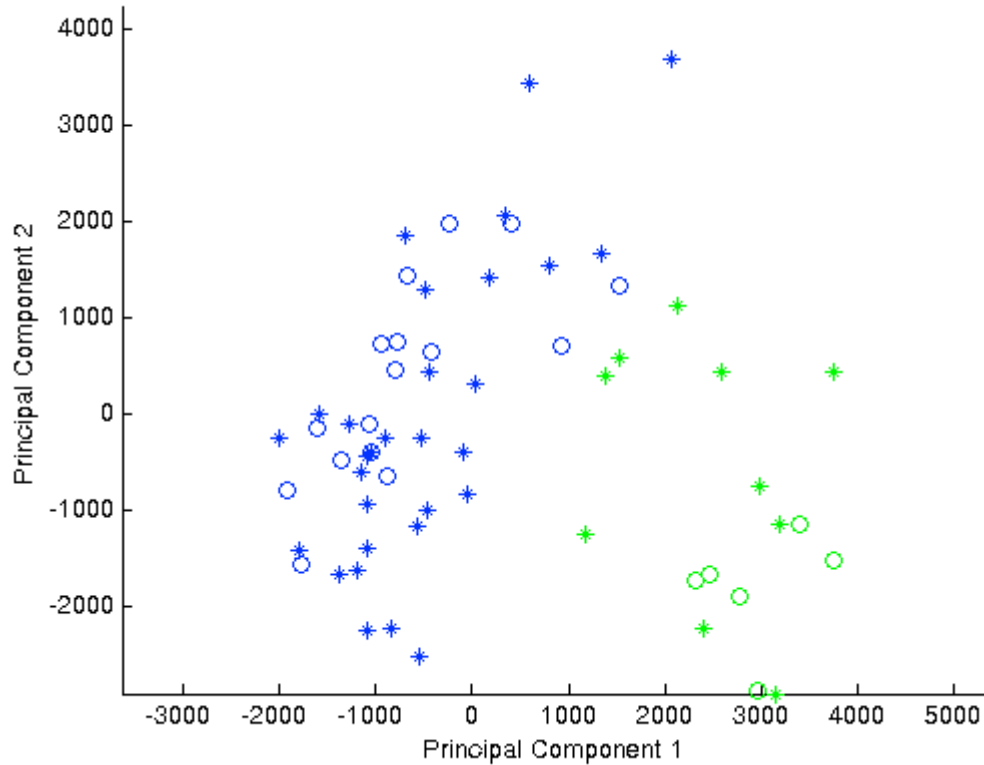


Figure A.6.1: PCA of Combined Doane *et al.* and Farmer *et al.* Cohorts where Data is Median-Centered per Probeset by Institution.

Principal Components Analysis (PCA) of the combined cohorts from Farmer *et al.* (Farmer, *et al.*, 2005) and Doane *et al.* (Doane, *et al.*, 2006) on natural-log scaled data quantile-normalized using the AffyProbeMiner-provided chip definition file (CDF) with median-centering per probeset by institution; p -value < 0.0001 . Information: The plot uses the first two principal components. Circles represent the Doane *et al.* cohort and asterisks (*) represent the Farmer *et al.* cohort. The hypothesized “molecular apocrine” phenotype is represented by the color green and “non-molecular apocrine” phenotype by the color blue. The p -values are calculated using the Fasano & Franceschini statistical test with the null hypothesis that the two phenotypes cannot be differentiated on the basis of two-dimensional, principal components coordinates.

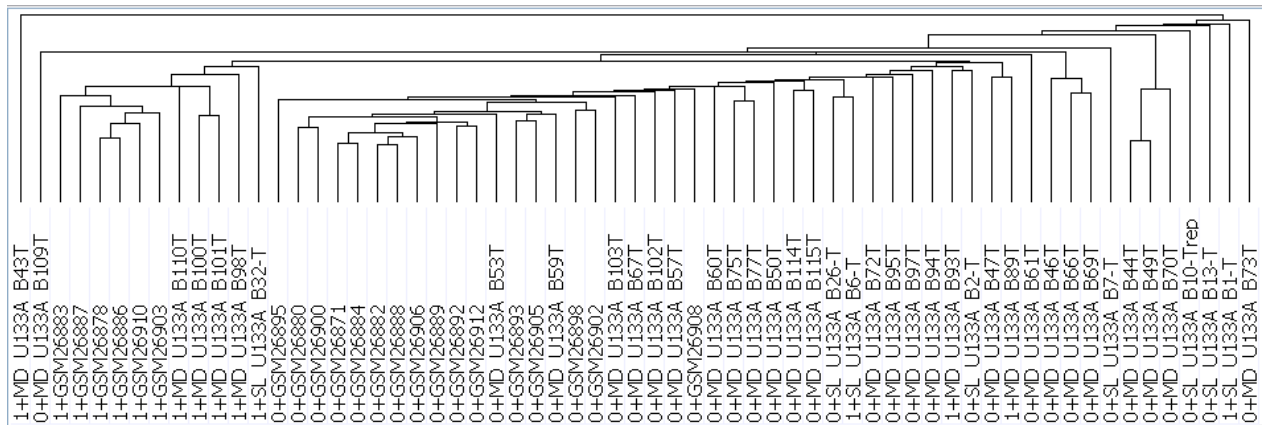


Figure A.6.2: Hierarchical Clustering of Combined Doane *et al.* and Farmer *et al.*

Cohorts where Data is Median-Centered per Probeset by Institution. of the combined cohorts from Farmer *et al.* (Farmer, *et al.*, 2005) and Doane *et al.* (Doane, *et al.*, 2006) on natural-log scaled data quantile-normalized using the AffyProbeMiner-provided chip definition file (CDF) with median-centering per probeset by institution. Information: Clustering was performed using the Pairwise-Average Linkage method and measures distance using Euclidian Distance. The samples cluster according to institution and not by their “molecular apocrine” or “non-molecular apocrine” phenotype as indicated by 1 and 0, respectively, preceding the sample ID.

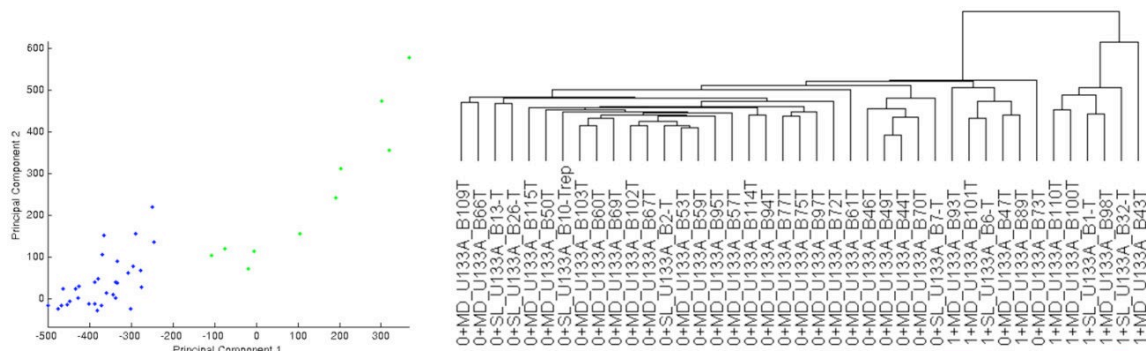


Figure A.6.3: PCA and Hierarchical Clustering on Doane *et al.* Cohort filtered for 400-gene signature by Farmer *et al.*

The 400-gene signature was identified by Farmer *et al.* to discriminate the molecular apocrine from the basal phenotype. The original data normalized by Doane *et al.* was filtered for the 400-gene signature. (LEFT) PCA plot of the filtered data. (RIGHT) Hierarchical Clustering of the filtered data. Information: (LEFT) The plot uses the first two principal components. The hypothesized “molecular apocrine” phenotype is represented by the color green and “non-molecular apocrine” phenotype by the color blue. (RIGHT) Phenotypes of “molecular apocrine” or “non-molecular apocrine” are indicated by 1 and 0, respectively, preceding the sample ID. Clustering was performed using the Pairwise-Average Linkage method and calculates distance using Euclidian Distance.

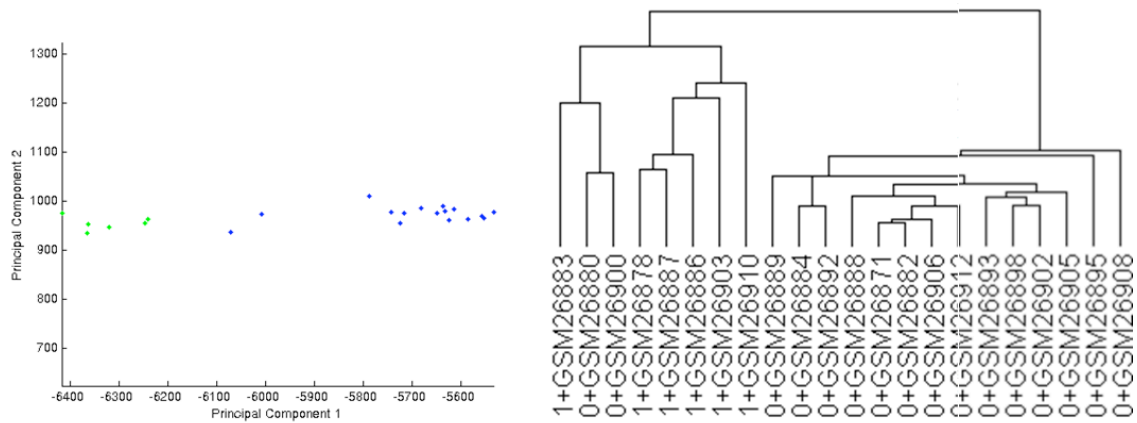


Figure A.6.4: PCA and Hierarchical Clustering on Farmer *et al.* Cohort filtered for 138-gene signature by Doane *et al.*

The 138-gene signature was identified by Doane *et al.* to discriminate the molecular apocrine from the basal phenotype. The original data normalized by Farmer *et al.* was filtered for the 138-gene signature. (LEFT) PCA plot of the filtered data. (RIGHT) Hierarchical Clustering of the filtered data. Information: (LEFT) The plot uses the first two principal components. The hypothesized “molecular apocrine” phenotype is represented by the color green and “non-molecular apocrine” phenotype by the color blue. (RIGHT) Phenotypes of “molecular apocrine” or “non-molecular apocrine” are indicated by 1 and 0, respectively, preceding the sample ID. Clustering was performed using the Pairwise-Average Linkage method and calculates distance using Euclidian Distance.

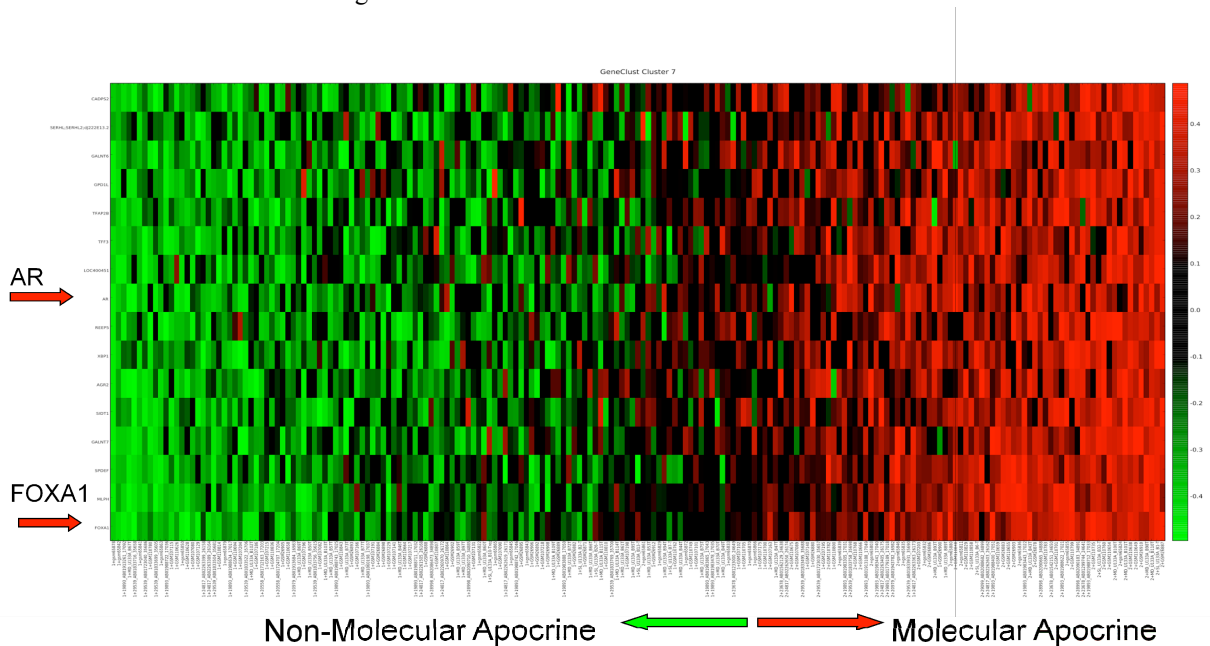


Figure A.6.5: Heat-map of Gene Cluster 7 Identified by Unsupervised Gene Shaving.

This heatmap of gene cluster 7 identified by unsupervised Gene Shaving includes both the AR and FOXA1 genes. This 16-gene cluster alone is able to act as a gene signature for separating the molecular apocrine samples in the 199 ER⁺ from the remaining samples.

Tables

Table A.6.1: The 100-probeset signatures for differentiating molecular apocrine and non-molecular apocrine phenotypes derived strictly from Doane *et al.* cohort.

This gene signatures was derived from the Doane *et al.* samples following normalization using XPN on the combined Doane *et al.* and Farmer *et al.* samples, with updated probeset definitions. There are 76 overlapping genes between this signature and that from Table A.6.2. This signatures was derived using Significance Analysis of Microarrays (Tusher, *et al.*, 2001).

Rank from SAM	HUGO Gene ID
1	FOXA1
2	TFAP2B
3	MLPH
4	SPDEF
5	SERHL;SERHL2;dJ222E13.2
6	PIP
7	RND1
8	AGR2
9	SERHL2
10	HMGCS2
11	AR
12	TFF3
13	SIDT1
14	XBP1
15	ALDH3B2
16	ZDHHC7
17	DHRS2
18	APOD
19	AZGP1;LOC646282
20	HGD
21	FAM5C
22	LASS4
23	LOC400451
24	REEP5
25	GALNT7

26	CYB5A
27	HRASLS2
28	FASN
29	KIAA0644
30	AZGP1
31	ALCAM
32	DCXR
33	CLCA2
34	KIAA1467
35	GPD1L
36	KMO
37	CRAT
38	MAGED2
39	UCP2
40	FGFR4
41	KCNMA1
42	SPRED2
43	GALNT6
44	LOC257407
45	CRISP3
46	CROT
47	FMO5
48	MSX2
49	MPHOSPH6
50	C1orf115
51	CIRBP
52	ZNF552
53	GJB1
54	TRIM36
55	SERHL;SERHL2
56	CLDN8
57	TSPAN15
58	C2orf54
59	DHCR24
60	CKMT1A;CKMT1B;LOC649970+731391:LOC731391
61	SCUBE2
62	SLC22A5
63	LARGE

64	G6PD
65	ECHDC2
66	GHR
67	SEPHS2
68	C7orf24
69	SEPP1
70	ARFIP2
71	SLC9A3R1
72	TBC1D9
73	FAH
74	SCNN1A
75	RETSAT
76	QPRT
77	TSC22D3
78	MAOA
79	YIPF1
80	SLC2A10
81	SPG11
82	ZNF652
83	IQGAP2
84	KIAA0310
85	FYCO1
86	CORO1B
87	PMAIP1
88	ANP32E
89	KRT16
90	SLC16A1
91	SERPINB5
92	FOXC1
93	FBXO17
94	NRTN
95	SPP1
96	CTSL2
97	NCAPH
98	LDHB
99	TUBB2B
100	LRFN4

Table A.6.2: The 100-probeset signatures for differentiating molecular apocrine and non-molecular apocrine phenotypes derived strictly from Farmer *et al.* cohort.

This gene signatures was derived from the Farmer *et al.* samples following normalization using XPN on the combined Doane *et al.* and Farmer *et al.* samples, with updated probeset definitions. There are 76 overlapping genes between this signature and that from Table A.6.1. This signatures was derived using Significance Analysis of Microarrays (Tusher, *et al.*, 2001).

Rank from SAM	HUGO Gene ID
1	PIP
2	SERHL;SERHL2;dJ222E13.2
3	FOXA1
4	DHRS2
5	SERHL2
6	TFAP2B
7	SPDEF
8	MLPH
9	HGD
10	APOD
11	HMGCS2
12	ALDH3B2
13	LASS4
14	AR
15	AGR2
16	RND1
17	AZGP1;LOC646282
18	GALNT7
19	ZDHHC7
20	PMAIP1
21	LOC400451
22	SIDT1
23	KRT16
24	TFF3
25	HRASLS2
26	CRAT
27	FASN
28	XBP1
29	REEP5
30	FAM5C

31	DCXR
32	CYB5A
33	AZGP1
34	KIAA0644
35	KIAA1467
36	ALCAM
37	GPD1L
38	CROT
39	LOC257407
40	SERHL;SERHL2
41	CLDN8
42	MAGED2
43	TRIM36
44	FMO5
45	DHCR24
46	C1orf115
47	RETSAT
48	KMO
49	KCNMA1
50	GJB1
51	IQGAP2
52	CLCA2
53	SLC22A5
54	SPRED2
55	CAB39L
56	MKNK2
57	FGFR4
58	ECHDC2
59	ZNF552
60	MPHOSPH6
61	SLC9A3R1
62	APBB2
63	CIRBP
64	UCP2
65	C2orf54
66	SCNN1A
67	TMEM62
68	LARGE

69	GALNT10
70	CYP4F8
71	MSX2
72	MAOA
73	TSC22D3
74	YIPF1
75	SEPHS2
76	CRISP3
77	TRPV6
78	SCGN
79	C7orf24
80	CES1;LOC652708
81	LOC651278;SERHL;SERHL2;dJ222E13.2
82	SPG11
83	CLN3
84	TM7SF2
85	AMACR
86	ALOX15B
87	NR2F1
88	TSPAN15
89	TMEM134
90	AKR1A1
91	BLVRB
92	ACSM1
93	VPS37C
94	FLJ20273
95	FAH
96	CYB561
97	GHR
98	CYP2J2
99	GALNT6
100	ABCA12

Table A.6.3: The 346-probeset signature for differentiating molecular apocrine and non-molecular apocrine phenotypes.

The signature was identified using Significance Analysis of Microarrays software on the Doane *et al.* and Farmer *et al.* cohorts at a false discovery rate of 0%.

Rank from SAM	HUGO Gene ID
1	SPDEF
2	FOXA1
3	MLPH
4	LOC651278;SERHL;SERHL2;dJ222E13.2
5	SERHL;SERHL2;dJ222E13.2
6	TFAP2B
7	RND1
8	AGR2
9	PIP
10	LOC400451
11	FASN
12	ALDH3B2
13	GALNT7
14	AR
15	SERHL2
16	SCNN1A
17	TFF3
18	CRAT
19	HGD
20	XBP1
21	SIDT1
22	CYB5A
23	PMAIP1
24	DHRS2
25	GPD1L
26	DHCR24
27	LASS4
28	APOD
29	LDHB
30	REEP5
31	SLC16A1
32	SLC44A4
33	SERPINB5
34	SLC9A3R1
35	KRT16
36	FOXC1
37	C1orf115

38	ATAD4
39	FLJ20273
40	ANP32E
41	AZGP1;LOC646282
42	KIAA0644
43	MYC
44	HRASLS2
45	TUBB6
46	TSPAN15
47	TOX3
48	PTX3
49	KMO
50	MAGED2
51	FSCN1
52	DUSP4
53	ARFIP2
54	ALCAM
55	KIAA1467
56	RHOB
57	DCXR
58	TBC1D9
59	NPDC1
60	BIK
61	KCNMA1
62	C20orf42
63	HMGCS2
64	ERBB2
65	GALNT6
66	MSX2
67	TSC22D3
68	SLC2A10
69	MTMR2
70	C12orf24
71	CRISP3
72	SERHL;SERHL2
73	DEK
74	AZGP1
75	ARHGAP8;LOC553158

76	C7orf24
77	CLN3
78	ZNF552
79	MKNK2
80	FZD4
81	HMGN4
82	HOXB2
83	SLC22A5
84	PLS3
85	FAH
86	SGCB
87	YIPF1
88	FGFR4
89	TMEM158
90	CNTNAP2
91	G6PD
92	CYB561
93	PCTK3
94	DGKD
95	GABRP
96	C1orf34
97	CLCA2
98	ECHDC1
99	ACACA
100	GATA3
101	NQO1
102	HPX
103	EN1
104	TYMS
105	TSPAN13
106	TM7SF2
107	CTSL2
108	FMO5
109	C10orf38
110	DUSP5
111	RETSAT
112	METRN
113	KIF23

114	GPSM2
115	BCL11A
116	GSTM3
117	MSN
118	ZNF467
119	REEP1
120	MIA
121	NRTN
122	PERLD1
123	WFS1
124	PEX11A
125	C2orf54
126	IQGAP2
127	FHOD1
128	PAPSS1
129	RET
130	C3orf14
131	DALRD3
132	SH3BP4
133	SPG11
134	RGL2
135	CEACAM6
136	MPHOSPH6
137	BLVRB
138	RAB17
139	SCARB1
140	DUSP10
141	CRIP1
142	SLC7A8
143	CDK10
144	TTK
145	GPX7
146	LRBA
147	LRRRC23
148	VPS37C
149	UQCRCQ
150	SYT17
151	SLC37A1

152	CYP4F8
153	ALDH4A1
154	QPRT
155	RNF103
156	PKP1
157	ABCA12
158	TOP2A
159	TUBB2B
160	NRAS
161	NNT
162	IQCG
163	CADPS2
164	SLC19A2
165	ALOX15B
166	SLC9A1
167	CPD
168	GPM6B
169	YEATS2
170	TEGT
171	NCAPH
172	LRFN4
173	ERBB3
174	UGDH
175	SULT1A1
176	BCL11A
177	SPP1
178	SEPHS2
179	FAM5C
180	ELOVL5
181	ZMYND8
182	TIMP2
183	QKI
184	ZIC1
185	HPRT1
186	TMEM62
187	GALE
188	KIAA0310
189	UCP2

190	FLJ20920
191	PNMT
192	C10orf116
193	SLC39A14
194	C16orf45
195	BLVRA
196	EFHD1
197	ACTR3B
198	KRT17
199	GRB7
200	ROPN1;ROPN1B
201	LOC257407
202	LRP12
203	CA12
204	ADCY6
205	ERBB4
206	CHIC2
207	MMP7
208	SLMO1
209	TJP3
210	ACOX2
211	CRIP2
212	ANKRD15
213	HSPA4L
214	LOC389842;RANBP1
215	ABHD11
216	RGS5
217	S100A2
218	CALU
219	CLGN
220	SUOX
221	ART3
222	CIRBP
223	ABCC6;LOC730013
224	KRT23
225	CANT1
226	NUDT4
227	FYCO1

228	LIMA1
229	NQO1
230	MICALL1
231	POF1B
232	PRKX;PRKY
233	SLC9A6
234	TMEM135
235	AKR1A1
236	EMP2
237	SLC38A1
238	SCARB1
239	GHR
240	MID1
241	DAPK1
242	LOC731069;PHF8
243	SCD
244	DOPEY2
245	C10orf57
246	SACS
247	ACVR1B
248	CLIC4
249	EGFL7
250	KRT6B
251	PIK3C2B
252	LOC648629;SYNGR2
253	LYN
254	MAPT
255	LRRC31
256	LOC643634;TPM4
257	LRIG1
258	MUC1
259	GJB1
260	GINS4
261	PSIP1
262	LRP10
263	MIPEP
264	MSX2;MSX2P
265	AP1G2

266	SFT2D2
267	SELENBP1
268	MUC1
269	LGALS8
270	FAAH
271	SLC7A8
272	KIAA0141
273	ETFA
274	KIAA0984
275	MCCC2
276	TRPM4
277	SC4MOL
278	CIB1
279	SPTLC2
280	CORO1B
281	CES1;LOC652708
282	ATOX1
283	GDF15
284	RABEP2
285	LASS6
286	AMACR
287	KYNU
288	SQRDL
289	COQ4
290	SYNGR2
291	LOC649917;LOC731354;ZDHHC11
292	FBP1
293	APBA2BP
294	ECHDC2
295	KRT19
296	C20orf3
297	MSMB
298	DUSP6
299	CAMK2N1
300	CLDN8
301	TRIM36
302	BCAM
303	RAB26

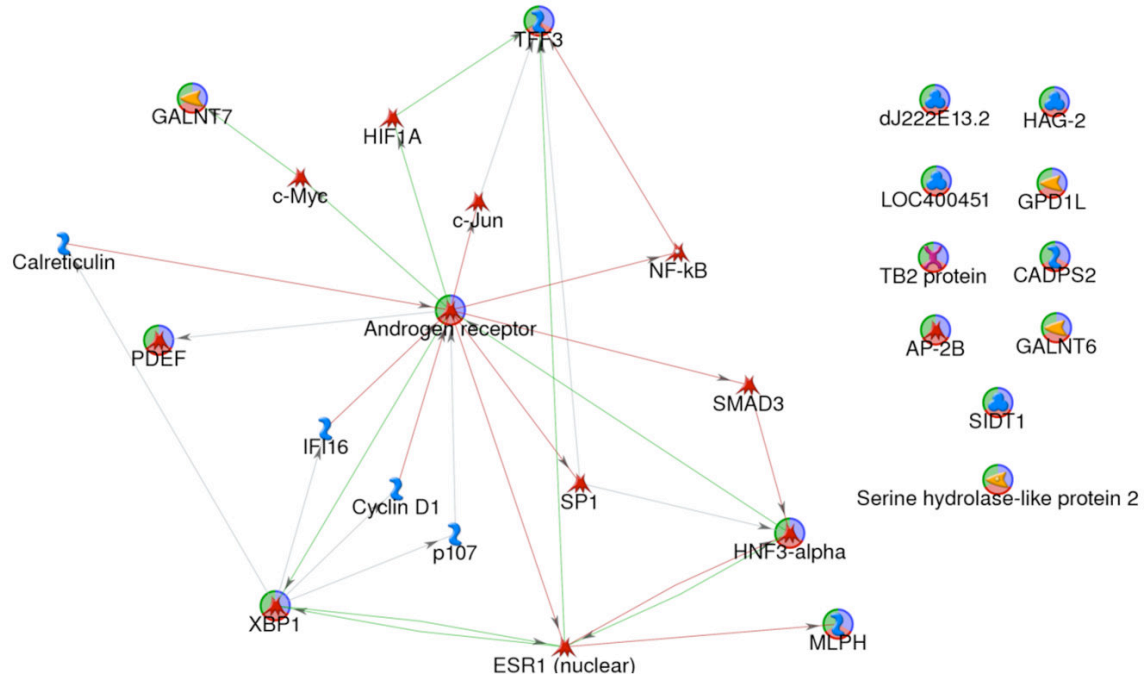
304	GPX4
305	CYB5R1
306	SCUBE2
307	TSPAN8
308	LOC645313;LOC651109;SCD
309	SLC22A18
310	INPP4B
311	KIAA0182
312	PBLD
313	NEIL1
314	PDLIM1
315	IDH2
316	LASP1
317	RAB1B
318	TOB1
319	CYB561D2
320	KIF13B
321	SRPK3
322	NME3
323	TSTA3
324	C20orf116
325	PPCS
326	SPRED2
327	CCNU
328	RBM35B
329	FEM1C
330	ROGDI
331	ABAT
332	LOH11CR2A
333	DNAJC12
334	HMGCS1
335	GUSB
336	PLEKHF2
337	ZNF652
338	FLJ20366
339	KRT7
340	NR2F1
341	BSCL2

342	ATP2C2
343	TMEM87A
344	PECI
345	BMP4
346	MEGF9

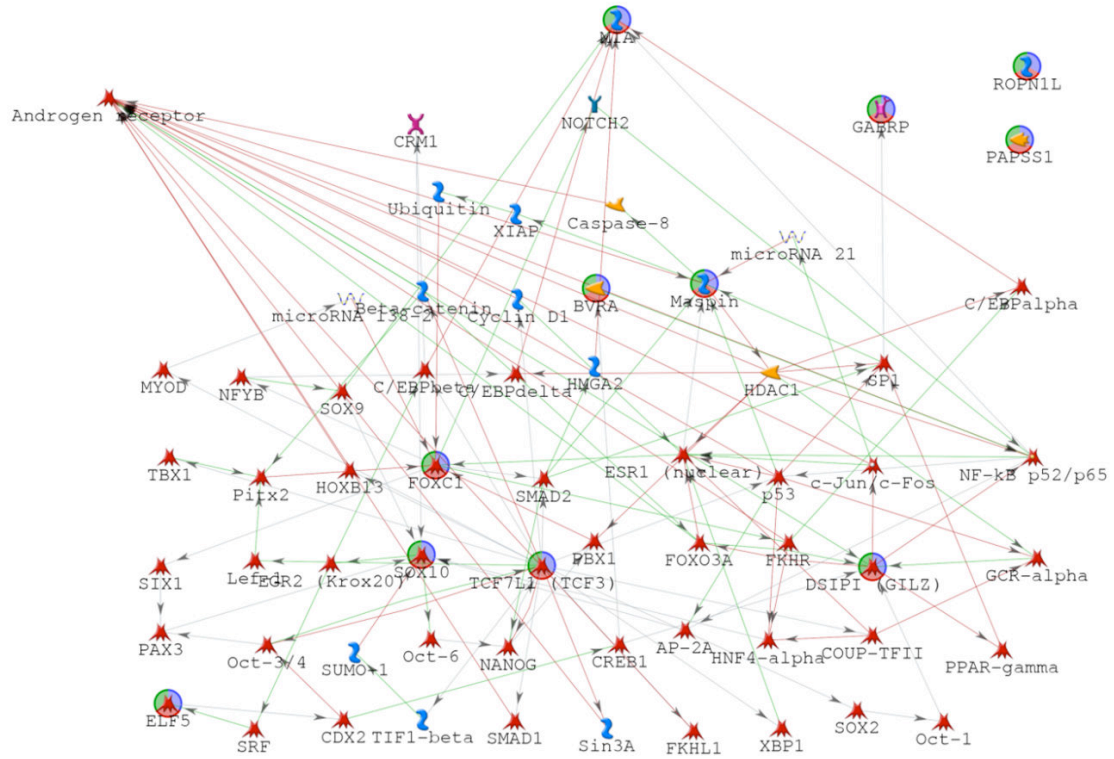
Close Network Interactions Identified by MetaCore for the Select Interacting Gene Clusters

Robust Bayesian Network Analysis indicates interactions amongst 14 of the top 26 gene clusters identified by Gene Shaving (Figure 6.9). MetaCore identifies close network relationships amongst the genes in Clusters 7, 16, 24, and 71 using Dijkstra's shortest path algorithm and prior knowledge:

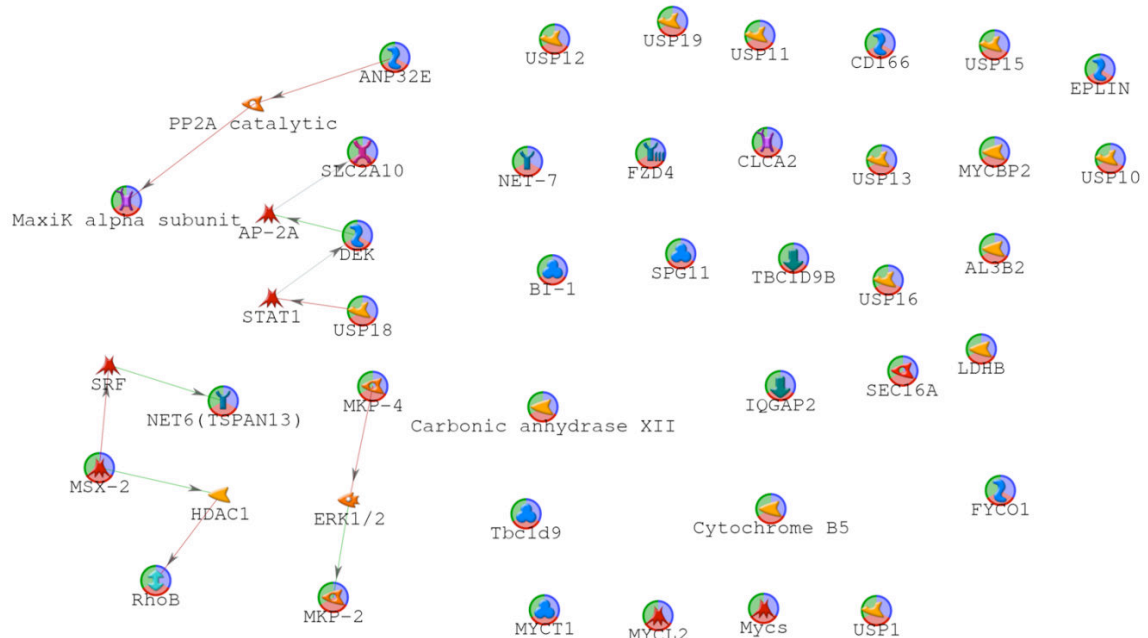
Cluster 07



Cluster 16



Cluster 24



References

- Adam, J. (1996). General aspects of modeling tumor growth and the immune response. In J. Adam & N. Bellomo (Eds.), *A survey of models on tumor immune systems dynamics* (pp. 15-87). Boston, MA: Birkhauser.
- Akutsu, T., Miyano, S., & Kuhara, S. (1999). Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac Symp Biocomput*, 17-28.
- Alarcon, T., Byrne, H. M., & Maini, P. K. (2003). A cellular automaton model for tumour growth in inhomogeneous environment. *J Theor Biol*, 225(2), 257-274.
- Alarcon, T., Byrne, H. M., & Maini, P. K. (2004). A mathematical model of the effects of hypoxia on the cell-cycle of normal and cancer cells. *J Theor Biol*, 229(3), 395-411.
- Alarcon, T., Byrne, H. M., & Maini, P. K. (2005). A multiple scale model for tumor growth. *Multi-scale Model. Simul.*, 3(2), 440-475.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1994). *Molecular Biology of the Cell* (3rd ed.). New York, NY: Garland Publishing.
- Ambrosi, D., & Preziosi, L. (2002). On the closure of mass balance models for tumor growth. *Mathematical Models & Methods in Applied Sciences*, 12(5), 737-754.
- Anderson, A., Zheng, X., & Cristini, V. (2005). Adaptive unstructured volume remeshing - I: the method. *J. Comp. Physics*, 156(1-2), 191-206.
- Anderson, A. R. (2005). A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Math Med Biol*, 22(2), 163-186.
- Anderson, A. R., & Chaplain, M. A. (1998). Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull Math Biol*, 60(5), 857-899.
- Anderson, A. R., & Quaranta, V. (2008). Integrative mathematical oncology. *Nat Rev Cancer*, 8(3), 227-234.
- Anderson, A. R. A. (2005). A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Mathematical Medicine and Biology-a Journal of the Ima*, 22(2), 163-186.
- Anderson, D. M., McFadden, G. B., & Wheeler, A. A. (1998). Diffuse-interface methods in fluid mechanics. *Annual Review of Fluid Mechanics*, 30, 139-165.
- Arancia, G., Calcabrini, A., Meschini, S., & Molinari, A. (1998). Intracellular distribution of anthracyclines in drug resistant cells. *Cytotechnology*, 27(1-3), 95-111.
- Araujo, R. P., & McElwain, D. L. (2004). A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bull Math Biol*, 66(5), 1039-1091.
- Athale, C., Mansury, Y., & Deisboeck, T. S. (2005). Simulating the impact of a molecular 'decision-process' on cellular phenotype and multicellular patterns in brain tumors. *J Theor Biol*, 233(4), 469-481.
- Athale, C. A., & Deisboeck, T. S. (2006). The effects of EGF-receptor density on multi-scale tumor growth patterns. *J Theor Biol*, 238(4), 771-779.

- Atkinson, E. J., Fridley, B. L., Goode, E. L., McDonnell, S. K., Liu-Mares, W., Rabe, K. G., *et al.* (2007). Linkage analysis using principal components of gene expression data. *BMC Proc*, 1 Suppl 1, S79.
- Ayati, B. P., Webb, G. F., & Anderson, A. R. A. (2006). Computational methods and results for structured multi-scale models of tumor invasion. *Multi-scale Modeling & Simulation*, 5(1), 1-20.
- Badve, S., Turbin, D., Thorat, M. A., Morimiya, A., Nielsen, T. O., Perou, C. M., *et al.* (2007). FOXA1 expression in breast cancer--correlation with luminal subtype A and survival. *Clin Cancer Res*, 13(15 Pt 1), 4415-4421.
- Bailly, M., Yan, L., Whitesides, G. M., Condeelis, J. S., & Segall, J. E. (1998). Regulation of protrusion shape and adhesion to the substratum during chemotactic responses of mammalian carcinoma cells. *Exp Cell Res*, 241(2), 285-299.
- Baish, J. W., Gazit, Y., Berk, D. A., Nozue, M., Baxter, L. T., & Jain, R. K. (1996). Role of tumor vascular architecture in nutrient and drug delivery: an invasion percolation-based network model. *Microvasc Res*, 51(3), 327-346.
- Bangs, A. L., & Paterson, T. S. (2003). Finding value in in silico biology. *Biosilico*, 1, 18-22.
- Barnes, M., Freudenberg, J., Thompson, S., Aronow, B., & Pavlidis, P. (2005). Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. *Nucleic Acids Res*, 33(18), 5914-5923.
- Barra, V. (2004). Analysis of gene expression data using functional principal components. *Comput Methods Programs Biomed*, 75(1), 1-9.
- Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat Genet*, 37(4), 382-390.
- Bearer, E. L., Lowengrub, J. S., Frieboes, H. B., Chuang, Y. L., Jin, F., Wise, S. M., *et al.* (2009). Multiparameter computational modeling of tumor invasion. *Cancer Res*, 69(10), 4493-4501.
- Belani, C. P. (2002). Creating a molecular classification for lung cancer. *Clin Lung Cancer*, 4(3), 139.
- Bello, L., Lucini, V., Costa, F., Pluderi, M., Giussani, C., Acerbi, F., *et al.* (2004). Combinatorial administration of molecules that simultaneously inhibit angiogenesis and invasion leads to increased therapeutic efficacy in mouse models of malignant glioma. *Clinical Cancer Research*, 10(13), 4527-4537.
- Bellomo, N., & Preziosi, L. (2000). Modelling and mathematical problems related to tumor evolution and its interaction with the immune system. *Mathematical and Computer Modelling*, 32(3-4), 413-452.
- Berger, M. J., & Colella, P. (1989). Local Adaptive Mesh Refinement for Shock Hydrodynamics. *Journal of Computational Physics*, 82(1), 64-84.
- Bernsen, H. J. J. A., & Van der Kogel, A. J. (1999). Antiangiogenic therapy in brain tumor models. *Journal of Neuro-Oncology*, 45(3), 247-255.
- Bhalla, U. S., & Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science*, 283(5400), 381-387.

- Bicciato, S., Pandin, M., Didone, G., & Di Bello, C. (2003). Pattern identification and classification in gene expression data using an autoassociative neural network model. *Biotechnol Bioeng*, *81*(5), 594-606.
- Bickel, D. R., Montazeri, Z., Hsieh, P. C., Beatty, M., Lawit, S. J., & Bate, N. J. (2009). Gene network reconstruction from transcriptional dynamics under kinetic model uncertainty: a case for the second derivative. *Bioinformatics*, *25*(6), 772-779.
- Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., *et al.* (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*, *439*(7074), 353-357.
- Bloemendal, H. J., Logtenberg, T., & Voest, E. E. (1999). New strategies in anti-vascular cancer therapy. *European Journal of Clinical Investigation*, *29*(9), 802-809.
- Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, *19*(2), 185-193.
- Boushaba, K., Levine, H. A., & Nilsen-Hamilton, M. (2006). A mathematical model for the regulation of tumor dormancy based on enzyme kinetics. *Bulletin of Mathematical Biology*, *68*(7), 1495-1526.
- Bracke, M. E., Maeseneer, D. D., Van Marck, V., Derycke, L., Vanhoecke, B., De Wever, O., *et al.* (2008). Cell motility and breast cancer metastasis. In R. E. Mansel (Ed.), *Metastasis of Breast Cancer* (pp. 47-75): Springer.
- Brandt, A. (1977). Multilevel Adaptive Solutions to Boundary-Value Problems. *Mathematics of Computation*, *31*(138), 333-390.
- Breward, C. J., Byrne, H. M., & Lewis, C. E. (2003). A multiphase model describing vascular tumour growth. *Bull Math Biol*, *65*(4), 609-640.
- Broom, B. M., Rinsurongkawong, W., Pusztai, L., & Do, K.-A. (in preparation). Building networks with microarray data.
- Bru, A., Albertos, S., Subiza, J. L., Garcia-Asenjo, J. L., & Bru, I. (2003). The universal dynamics of tumor growth. *Biophysical Journal*, *85*(5), 2948-2961.
- Burton, A. C. (1966). Rate of growth of solid tumours as a problem of diffusion. *Growth*, *30*(2), 157-176.
- Byrne, H., & Preziosi, L. (2003). Modelling solid tumour growth using the theory of mixtures. *Mathematical Medicine and Biology-a Journal of the Ima*, *20*(4), 341-366.
- Byrne, H. M., Alarcon, T., Owen, M. R., Webb, S. D., & Maini, P. K. (2006). Modelling aspects of cancer dynamics: a review. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, *364*(1843), 1563-1578.
- Byrne, H. M., & Chaplain, M. A. (1995). Growth of nonnecrotic tumors in the presence and absence of inhibitors. *Math Biosci*, *130*(2), 151-181.
- Byrne, H. M., & Chaplain, M. A. J. (1996). Modelling the role of cell-cell adhesion in the growth and development of carcinoma. *Mathematical and Computer Modelling*, *24*(12), 1-17.

- Byrne, H. M., & Chaplain, M. A. J. (1997). Free boundary value problems associated with the growth and development of multicellular spheroids. *European Journal of Applied Mathematics*, 8, 639-658.
- Byrne, H. M., & Chaplin, M. A. (1996). Growth of necrotic tumors in the presence and absence of inhibitors. *Math Biosci*, 135(2), 187-216.
- Carey, A. M., Pramanik, R., Nicholson, L. J., Dew, T. K., Martin, F. L., Muir, G. H., *et al.* (2007). Ras-MEK-ERK signaling cascade regulates androgen receptor element-inducible gene transcription and DNA synthesis in prostate cancer cells. *Int J Cancer*, 121(3), 520-527.
- Carter, S. L., Eklund, A. C., Mecham, B. H., Kohane, I. S., & Szallasi, Z. (2005). Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics*, 6, 107.
- Castro, M., Molina-Paris, C., & Deisboeck, T. S. (2005). Tumor growth instability and the onset of invasion. *Physical Review E*, 72(4), -.
- Chambers, A. F., Groom, A. C., & MacDonald, I. C. (2002). Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer*, 2(8), 563-572.
- Chaplain, M. A., & Anderson, A. (2003). Mathematical modeling of tissue invasion. In L. Preziosi (Ed.), *Cancer modeling and simulation* (pp. 269-297): CRC Press.
- Chaplain, M. A. J. (1996). Avascular growth, angiogenesis and vascular growth in solid tumours: The mathematical modelling of the stages of tumour development. *Mathematical and Computer Modelling*, 23(6), 47-87.
- Chaplain, M. A. J., Graziano, L., & Preziosi, L. (2006). Mathematical modelling of the loss of tissue compression responsiveness and its role in solid tumour development. *Mathematical Medicine and Biology-a Journal of the Ima*, 23(3), 197-229.
- Chaplain, M. A. J., & Lolas, G. (2005). Mathematical modelling of cancer cell invasion of tissue: The role of the urokinase plasminogen activation system. *Mathematical Models & Methods in Applied Sciences*, 15(11), 1685-1734.
- Chaturvedi, R., Huang, C., Kazmierczak, B., Schneider, T., Izaguirre, J. A., Glimm, T., *et al.* (2005). On multi-scale approaches to three-dimensional modelling of morphogenesis. *Journal of The Royal Society Interface*, 2(3), 237-253.
- Chen, L. L., Zhang, L., Yoon, J., & Deisboeck, T. S. (2009). Cancer cell motility: optimizing spatial search strategies. *Biosystems*, 95(3), 234-242.
- Chen, T., He, H. L., & Church, G. M. (1999). Modeling gene expression with differential equations. *Pac Symp Biocomput*, 29-40.
- Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. A., *et al.* (2009). Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol Syst Biol*, 5, 239.
- Chicoine, M. R., & Silbergeld, D. L. (1995). Assessment of Brain-Tumor Cell Motility *in vivo* and *in vitro*. *Journal of Neurosurgery*, 82(4), 615-622.
- Chomyak, O. G., & Sidorenko, M. V. (2001). Multicellular spheroids model in oncology. *Experimental Oncology*, 23(4), 236-241.

- Chuang, Y.-L., Edgerton, M. E., Macklin, P., Wise, S., Lowengrub, J., & Cristini, V. (in preparation). Clinical Predictions of Bulk DCIS Properties Based on a Duct-Scale Mixture Model.
- Citri, A., & Yarden, Y. (2006). EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol*, 7(7), 505-516.
- Clatz, O., Bondiau, P. Y., Delingette, H., Malandain, G., Sermesant, M., Warfield, S. K., *et al.* (2004). In silico tumor growth: application to glioblastoma. In C. Barillot, D. R. Haynor & P. Hether (Eds.), *MICCAI 2004, LNCS 3217* (pp. 337-345): Springer-Verlag.
- Clatz, O., Sermesant, M., Bondiau, P. Y., Delingette, H., Warfield, S. K., Malandain, G., *et al.* (2005). Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. *IEEE Trans Med Imaging*, 24(10), 1334-1346.
- Collins, S. J. (1987). The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression. *Blood*, 70(5), 1233-1244.
- Condeelis, J., Singer, R. H., & Segall, J. E. (2005). The great escape: When cancer cells hijack the genes for chemotaxis and motility. *Annual Review of Cell and Developmental Biology*, 21, 695-718.
- Condeelis, J. S., Wyckoff, J. B., Bailly, M., Pestell, R., Lawrence, D., Backer, J., *et al.* (2001). Lamellipodia in invasion. *Semin Cancer Biol*, 11(2), 119-128.
- Covell, D. G., Wallqvist, A., Rabow, A. A., & Thanki, N. (2003). Molecular classification of cancer: unsupervised self-organizing map analysis of gene expression microarray data. *Mol Cancer Ther*, 2(3), 317-332.
- Crampin, E. J., Halstead, M., Hunter, P., Nielsen, P., Noble, D., Smith, N., *et al.* (2004). Computational physiology and the Physiome Project. *Exp Physiol*, 89(1), 1-26.
- Cristini, V., Blawdziewicz, J., & Lowenberg, M. (2001). An adaptive mesh algorithm for evolving surfaces: simulations of drop breakup and coalescence. *J. Comp. Physics*, 168, 445-463.
- Cristini, V., Frieboes, H. B., Gatenby, R., Caserta, S., Ferrari, M., & Sinek, J. (2005). Morphologic instability and cancer invasion. *Clin Cancer Res*, 11(19 Pt 1), 6772-6779.
- Cristini, V., Li, X., Lowengrub, J. S., & Wise, S. M. (2009). Nonlinear simulations of solid tumor growth using a mixture model: invasion and branching. *J Math Biol*, 58(4-5), 723-763.
- Cristini, V., Lowengrub, J., & Nie, Q. (2003). Nonlinear simulation of tumor growth. *J Math Biol*, 46(3), 191-224.
- Cui, X., & Churchill, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol*, 4(4), 210.
- Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., *et al.* (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res*, 33(20), e175.

- DeGregorio, M. W., Lui, G. M., Macher, B. A., & Wilbur, J. R. (1984). Uptake, metabolism, and cytotoxicity of doxorubicin in human Ewing's sarcoma and rhabdomyosarcoma cells. *Cancer Chemother Pharmacol*, *12*(1), 59-63.
- Deisboeck, T. S., Berens, M. E., Kansal, A. R., Torquato, S., Stemmer-Rachamimov, A. O., & Chiocca, E. A. (2001). Pattern of self-organization in tumour systems: complex growth dynamics in a novel brain tumour spheroid model. *Cell Prolif*, *34*(2), 115-134.
- Deisboeck, T. S., Zhang, L., & Martin, S. (2007). Advancing Cancer Systems Biology: Introducing the Center for the Development of a Virtual Tumor, CViT. *Cancer Inform*, *5*, 1-8.
- Deisboeck, T. S., Zhang, L., Yoon, J., & Costa, J. (2009). In silico cancer modeling: is it ready for prime time? *Nat Clin Pract Oncol*, *6*(1), 34-42.
- Demant, E. J., & Friche, E. (1998). Kinetics of anthracycline accumulation in multidrug-resistant tumor cells: relationship to drug lipophilicity and serum albumin binding. *Biochem Pharmacol*, *56*(9), 1209-1217.
- Deroulers, C., Aubert, M., Badoual, M., & Grammaticos, B. (2009). Modeling tumor cell migration: From microscopic to macroscopic models. *Physical Review E*, *79*(3), -.
- Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., *et al.* (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res*, *14*(16), 5158-5165.
- DeWitt, A., Iida, T., Lam, H. Y., Hill, V., Wiley, H. S., & Lauffenburger, D. A. (2002). Affinity regulates spatial range of EGF receptor autocrine ligand binding. *Dev Biol*, *250*(2), 305-316.
- DeWitt, A. E., Dong, J. Y., Wiley, H. S., & Lauffenburger, D. A. (2001). Quantitative analysis of the EGF receptor autocrine system reveals cryptic regulation of cell response by ligand capture. *J Cell Sci*, *114*(Pt 12), 2301-2313.
- Dickinson, R. B., & Tranquillo, R. T. (1993). A Stochastic-Model for Adhesion-Mediated Cell Random Motility and Haptotaxis. *Journal of Mathematical Biology*, *31*(6), 563-600.
- Dimilla, P. A., Barbee, K., & Lauffenburger, D. A. (1991). Mathematical-Model for the Effects of Adhesion and Mechanics on Cell-Migration Speed. *Biophysical Journal*, *60*(1), 15-37.
- Dittmar, T., Husemann, A., Schewe, Y., Nofer, J. R., Niggemann, B., Zanker, K. S., *et al.* (2002). Induction of cancer cell migration by epidermal growth factor is initiated by specific phosphorylation of tyrosine 1248 of c-erbB-2 receptor via EGFR. *FASEB J*, *16*(13), 1823-1825.
- Dluz, S. M., Higashiyama, S., Damm, D., Abraham, J. A., & Klagsbrun, M. (1993). Heparin-binding epidermal growth factor-like growth factor expression in cultured fetal human vascular smooth muscle cells. Induction of mRNA levels and secretion of active mitogen. *J Biol Chem*, *268*(24), 18330-18334.
- Do, K.-A., Broom, B. M., & Wen, S. (2003). *The analysis of gene expression data: methods and software*. New York: Springer.

- Doane, A. S., Danso, M., Lal, P., Donaton, M., Zhang, L., Hudis, C., *et al.* (2006). An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene*, 25(28), 3994-4008.
- Dordal, M. S., Ho, A. C., Jackson-Stone, M., Fu, Y. F., Goolsby, C. L., & Winter, J. N. (1995). Flow cytometric assessment of the cellular pharmacokinetics of fluorescent drugs. *Cytometry*, 20(4), 307-314.
- dos Reis, A. N., Mombach, J. C. M., Walter, M., & de Avila, L. F. (2003). The interplay between cell adhesion and environment rigidity in the morphology of tumors. *Physica a-Statistical Mechanics and Its Applications*, 322(1-4), 546-554.
- Durand, R. E. (1986). Chemosensitivity testing in V79 spheroids: drug delivery and cellular microenvironment. *J Natl Cancer Inst*, 77(1), 247-252.
- Durand, R. E. (1990). Slow penetration of anthracyclines into spheroids and tumors: a therapeutic advantage? *Cancer Chemother Pharmacol*, 26(3), 198-204.
- Edelstein-Keshet, L. (2005). *Mathematical models in biology*. Philadelphia: Society of Industrial and Applied Mathematics.
- Edgerton, M. E., Fisher, D. H., Tang, L., Frey, L. J., & Chen, Z. (2007). Data Mining for Gene Networks Relevant to Poor Prognosis in Lung Cancer Via Backward-Chaining Rule Induction. *Cancer Inform*, 2, 93114.
- Edgerton, M. E., Macklin, P., Chuang, Y.-L., Tamaiolo, G., Kim, J., Sanga, S., *et al.* (In Review). Ductal carcinoma in silico: A multi-scale model of ductal carcinoma in situ.
- Eichler, G. S., Reimers, M., Kane, D., & Weinstein, J. N. (2007). The LeFE algorithm: embracing the complexity of gene expression in the interpretation of microarray data. *Genome Biol*, 8(9), R187.
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25), 14863-14868.
- Ekins, S., Nikolsky, Y., Bugrim, A., Kirillov, E., & Nikolskaya, T. (2007). Pathway mapping tools for analysis of high content data. *Methods Mol Biol*, 356, 319-350.
- El-Kareh, A. W., & Secomb, T. W. (2003). A mathematical model for cisplatin cellular pharmacodynamics. *Neoplasia*, 5(2), 161-169.
- El-Kareh, A. W., & Secomb, T. W. (2005). Two-mechanism peak concentration model for cellular pharmacodynamics of Doxorubicin. *Neoplasia*, 7(7), 705-713.
- ElBayoumi, T. A., & Torchilin, V. P. (2009). Tumor-targeted nanomedicines: enhanced antitumor efficacy in vivo of doxorubicin-loaded, long-circulating liposomes modified with cancer-specific monoclonal antibody. *Clin Cancer Res*, 15(6), 1973-1980.
- Elvin, P., & Garner, A. P. (2005). Tumour invasion and metastasis: challenges facing drug discovery. *Current Opinion in Pharmacology*, 5(4), 374-381.
- Eungdamrong, N. J., & Iyengar, R. (2004a). Computational approaches for modeling regulatory cellular networks. *Trends Cell Biol*, 14(12), 661-669.

- Eungdamrong, N. J., & Iyengar, R. (2004b). Modeling cell signaling networks. *Biol Cell*, 96(5), 355-362.
- Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., *et al.* (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene*, 24(29), 4660-4671.
- Fasano, G., & Franceschini, A. (1987). A multidimensional version of the Kolmogorov-Smirnov test. *Mon. Not. R. Astron. Soc.*, 225, 155-170.
- Feinberg, A. P. (2004). The epigenetics of cancer etiology. *Semin Cancer Biol*, 14(6), 427-432.
- Fernandez, E. A., & Balzarini, M. (2007). Improving cluster visualization in self-organizing maps: application in gene expression data analysis. *Comput Biol Med*, 37(12), 1677-1689.
- Ferreira, S. C., Martins, M. L., & Vilela, M. J. (2002). Reaction-diffusion model for the growth of avascular tumor. *Physical Review E*, 65(2), -.
- Finkenstadt, B., Heron, E. A., Komorowski, M., Edwards, K., Tang, S., Harper, C. V., *et al.* (2008). Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*, 24(24), 2901-2907.
- Fisher D, Edgerton M, Tang L, Frey L, & Chen Z (2005). Searching for Meaningful Feature Interactions with Backward Chaining Rule Induction. In Famili AF, Kok JN, Pena JM, Siebes A & Feelders A (Eds.), *Advances in Intelligent Data Analysis VI* (Vol. 3646, pp. 86-96). Berlin/Heidelberg: Springer.
- Fisher DH, Edgerton ME, Chen Z, Tang L, & Frey LJ (2006). Backward Chaining Rule Induction. [Journal]. *Intelligent Data Analysis*, 10(5), 397-417.
- Flaherty, B., McGarry, J. P., & McHugh, P. E. (2007). Mathematical models of cell motility. *Cell Biochem Biophys*, 49(1), 14-28.
- Folkman, J. (1971). Tumor angiogenesis: therapeutic implications. *N Engl J Med*, 285(21), 1182-1186.
- Folkman, J., & Klagsbrun, M. (1987). Angiogenic factors. *Science*, 235(4787), 442-447.
- Frey, L., Edgerton, M. E., Fisher, D. H., Tang, L., & Chen, Z. (2005). Using prior knowledge and rule induction methods to discover molecular markers of prognosis in lung cancer. *AMIA Annu Symp Proc*, 256-260.
- Friday, B. B., & Adjei, A. A. (2008). Advances in targeting the Ras/Raf/MEK/Erk mitogen-activated protein kinase cascade with MEK inhibitors for cancer therapy. *Clin Cancer Res*, 14(2), 342-346.
- Frieboes, H., Wise, S. M., Lowengrub, J., & Cristini, V. (in press). Three-dimensional diffuse interface simulation of multispecies tumor growth-II: investigation of tumor invasion. *Bull Math Biol*.
- Frieboes, H. B., Lowengrub, J. S., Wise, S., Zheng, X., Macklin, P., Bearer, E. L., *et al.* (2007). Computer simulation of glioma growth and morphology. *Neuroimage*, 37 Suppl 1, S59-70.
- Frieboes, H. B., Lowengrub, J. S., Wise, S., Zheng, X., Macklin, P., Elaine, L. B. D., *et al.* (2007). Computer simulation of glioma growth and morphology. *Neuroimage*, 37, S59-S70.

- Frieboes, H. B., Sinek, J., Nalcioglu, O., Fruehauf, J., & Cristini, V. (2006). Nanotechnology in cancer drug therapy: a biocomputational approach *BioMEMS and Biomedical Nanotechnology* (Vol. 1, pp. 441-466): Springer-Verlag.
- Frieboes, H. B., Zheng, X., Sun, C. H., Tromberg, B., Gatenby, R., & Cristini, V. (2006). An integrated computational/experimental model of tumor invasion. *Cancer Res*, 66(3), 1597-1604.
- Friedl, P. (2004). Prespecification and plasticity: shifting mechanisms of cell migration. *Current Opinion in Cell Biology*, 16(1), 14-23.
- Friedl, P., Hegerfeldt, Y., & Tilisch, M. (2004). Collective cell migration in morphogenesis and cancer. *International Journal of Developmental Biology*, 48(5-6), 441-449.
- Friedl, P., & Wolf, K. (2003). Tumour-cell invasion and migration: Diversity and escape mechanisms. *Nature Reviews Cancer*, 3(5), 362-374.
- Friedman, A. (2004). A hierarchy of cancer models and their mathematical challenges. *Discrete and Continuous Dynamical Systems-Series B*, 4(1), 147-159.
- Friedman, N., Nachman, I., & Pe'er, D. (1999). *Learning bayesian network structure from massive datasets: the 'sparse candidate' algorithm*. Paper presented at the Proc. 15th Conference on Uncertainty in Artificial Intelligence.
- Fruehauf, J. P. (2002). In vitro assay-assisted treatment selection for women with breast or ovarian cancer. *Endocr Relat Cancer*, 9(3), 171-182.
- Fruehauf, J. P., & Bosanquet, A. G. (1993). In vitro determination of drug response: a discussion of clinical applications. *Principles and Practice Oncology*, 7, 1-16.
- Garcke, H., Nestler, B., & Stinner, B. (2004). A diffuse interface model for alloys with multiple components and phases. *Siam Journal on Applied Mathematics*, 64(3), 775-799.
- Garner, A. L., Lau, Y. Y., Jackson, T. L., Uhler, M. D., Jordan, D. W., & Gilgenbach, R. M. (2005). Incorporating spatial dependence into a multicellular tumor spheroid growth model. *Journal of Applied Physics*, 98(12), -.
- Gatenby, R. A. (1998). Mathematical models of tumor-host interactions. *Cancer J*, 11, 289-293.
- Gatenby, R. A., & Gawlinski, E. T. (1996). A reaction-diffusion model of cancer invasion. *Cancer Research*, 56(24), 5745-5753.
- Gatenby, R. A., & Gawlinski, E. T. (2003a). The glycolytic phenotype in carcinogenesis and tumor invasion: Insights through mathematical models. *Cancer Research*, 63(14), 3847-3854.
- Gatenby, R. A., & Gawlinski, E. T. (2003b). The glycolytic phenotype in carcinogenesis and tumor invasion: insights through mathematical models. *Cancer Res*, 63(14), 3847-3854.
- Gonzalez-Angulo, A. M., Stemke-Hale, K., Palla, S. L., Carey, M., Agarwal, R., Meric-Bertram, F., et al. (2009). Androgen receptor levels and association with PIK3CA mutations and prognosis in breast cancer. *Clin Cancer Res*, 15(7), 2472-2478.
- Goutte, C. (1997). Note on free lunches and cross-validation. *Neural Computation*, 9(6), 1245-1249.

- Greenspan, H. P. (1972). Models for the growth of a solid tumor by diffusion. *Stud. Appl. Math.*, 52, 317-340.
- Greenspan, H. P. (1976). Growth and Stability of Cell-Cultures and Solid Tumors. *Journal of Theoretical Biology*, 56(1), 229-242.
- Gulledge, C. J., & Dewhirst, M. W. (1996). Tumor oxygenation: a matter of supply and demand. *Anticancer Res*, 16(2), 741-749.
- Gusterson, B. A., Ross, D. T., Heath, V. J., & Stein, T. (2005). Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res*, 7(4), 143-148.
- Habashy, H. O., Powe, D. G., Rakha, E. A., Ball, G., Paish, C., Gee, J., *et al.* (2008). Forkhead-box A1 (FOXA1) expression in breast cancer and its prognostic significance. *Eur J Cancer*, 44(11), 1541-1551.
- Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1), 57-70.
- Harms, B. D., Bassi, G. M., Horwitz, A. R., & Lauffenburger, D. A. (2005). Directional persistence of EGF-induced cell migration is associated with stabilization of lamellipodial protrusions. *Biophys J*, 88(2), 1479-1488.
- Haroon, Z., Peters, K. G., Greenberg, C. S., & Dewhirst, M. W. (1999). Angiogenesis and blood flow in the solid tumors *Antiangiogenic agents in cancer therapy* (pp. 3-21). Totowa, NJ: Humana Press.
- Hartwell, L. H., & Kastan, M. B. (1994). Cell cycle control and cancer. *Science*, 266(5192), 1821-1828.
- Hastie, T., Tibshirani, R., Eisen, M. B., Alizadeh, A., Levy, R., Staudt, L., *et al.* (2000). 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol*, 1(2), RESEARCH0003.
- Hatzikirou, H., Deutsch, A., Schaller, C., Simon, M., & Swanson, K. (2005). Mathematical modelling of glioblastoma tumour development: A review. *Mathematical Models & Methods in Applied Sciences*, 15(11), 1779-1794.
- Heiser, L. M., Wang, N. J., Talcott, C. L., Laderoute, K. R., Knapp, M., Guan, Y., *et al.* (2009). Integrated analysis of breast cancer cell lines reveals unique signaling pathways. *Genome Biol*, 10(3), R31.
- Herrero, J., & Dopazo, J. (2002). Combining hierarchical clustering and self-organizing maps for exploratory analysis of gene expression patterns. *J Proteome Res*, 1(5), 467-470.
- Herzmark, P., Campbell, K., Wang, F., Wong, K., El-Samad, H., Groisman, A., *et al.* (2007). Bound attractant at the leading vs. the trailing edge determines chemotactic prowess. *Proc Natl Acad Sci U S A*, 104(33), 13349-13354.
- Hilger, R. A., Scheulen, M. E., & Strumberg, D. (2002). The Ras-Raf-MEK-ERK pathway in the treatment of cancer. *Onkologie*, 25(6), 511-518.
- Hogea, C. S., Murray, B. T., & Sethian, J. A. (2006). Simulating complex tumor dynamics from avascular to vascular growth using a general level-set method. *Journal of Mathematical Biology*, 53(1), 86-134.
- Hogue, C. (1998). Shape representation and contact detection for discrete element simulations of arbitrary geometries. *Engineering Computations*, 15(3), 374-390.

- Holz, M., & Fahr, A. (2001). Compartment modeling. *Adv Drug Deliv Rev*, 48(2-3), 249-264.
- Hopkins, A. L., & Groom, C. R. (2002). The druggable genome. *Nat Rev Drug Discov*, 1(9), 727-730.
- Huang, J., Shimizu, H., & Shioya, S. (2003). Clustering gene expression pattern and extracting relationship in gene network based on artificial neural networks. *J Biosci Bioeng*, 96(5), 421-428.
- Huang, S., & Ingber, D. E. (1999). The structural and mechanical complexity of cell-growth control. *Nature Cell Biology*, 1(5), E131-E138.
- Hunter, P. J., Kohl, P., & Noble, D. (2001). Integrative models of the heart: achievements and limitations. *Philos. Trans. Roy. Soc. London Ser. A*, 359, 1049-1054.
- Hurwitz, S. J., Terashima, M., Mizunuma, N., & Slapak, C. A. (1997). Vesicular anthracycline accumulation in doxorubicin-selected U-937 cells: participation of lysosomes. *Blood*, 89(10), 3745-3754.
- Iglesias, P. A., & Devreotes, P. N. (2008). Navigating through models of chemotaxis. *Curr Opin Cell Biol*, 20(1), 35-40.
- Ignar-Trowbridge, D. M., Nelson, K. G., Bidwell, M. C., Curtis, S. W., Washburn, T. F., McLachlan, J. A., *et al.* (1992). Coupling of dual signaling pathways: epidermal growth factor action involves the estrogen receptor. *Proc Natl Acad Sci U S A*, 89(10), 4658-4662.
- Inoue, S., Ohnuma, T., Holland, J. F., & Wasserman, L. R. (1985). Susceptibility of multicellular tumor spheroids (MTS) to doxorubicin (DXR) and cisplatin. *Proceedings of the American Association for Cancer Research*, 26, 341.
- Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T. C., Smeds, J., *et al.* (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res*, 66(21), 10292-10301.
- Jackson, T. L. (2003). Intracellular accumulation and mechanism of action of doxorubicin in a spatio-temporal tumor model. *J Theor Biol*, 220(2), 201-213.
- Jackson, T. L. (2004). A mathematical model of prostate tumor growth and androgen-independent relapse. *Discrete and Continuous Dynamical Systems-Series B*, 4(1), 187-201.
- Jackson, T. L., & Byrne, H. M. (2002). A mechanical model of tumor encapsulation and transcapsular spread. *Mathematical Biosciences*, 180, 307-328.
- Jacqmin, D. (1999). Calculation of two-phase Navier-Stokes flows using phase-field modeling. *Journal of Computational Physics*, 155(1), 96-127.
- Jain, R. K. (1990). Physiological Barriers to Delivery of Monoclonal-Antibodies and Other Macromolecules in Tumors. *Cancer Research*, 50(3), S814-S819.
- Jain, R. K. (2001a). Delivery of molecular medicine to solid tumors: lessons from in vivo imaging of gene expression and function. *J Control Release*, 74(1-3), 7-25.
- Jain, R. K. (2001b). Normalizing tumor vasculature with anti-angiogenic therapy: a new paradigm for combination therapy. *Nat Med*, 7(9), 987-989.
- Janetopoulos, C., & Firtel, R. A. (2008). Directional sensing during chemotaxis. *FEBS Lett*, 582(14), 2075-2085.

- Jekunen, A. P., Shalinsky, D. R., Hom, D. K., Albright, K. D., Heath, D., & Howell, S. B. (1993). Modulation of cisplatin cytotoxicity by permeabilization of the plasma membrane by digitonin in vitro. *Biochem Pharmacol*, *45*(10), 2079-2085.
- Jiang, Y., Pjesivac-Grbovic, J., Cantrell, C., & Freyer, J. P. (2005). A multi-scale model for avascular tumor growth. *Biophys J*, *89*(6), 3884-3894.
- Jones, A. F., Byrne, H. M., Gibson, J. S., & Dold, J. W. (2000). A mathematical model of the stress induced during avascular tumour growth. *Journal of Mathematical Biology*, *40*(6), 473-499.
- Kamangar, F., Dores, G. M., & Anderson, W. F. (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol*, *24*(14), 2137-2150.
- Kansal, A. R., Torquato, S., Harsh, G. I., Chiocca, E. A., & Deisboeck, T. S. (2000a). Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J Theor Biol*, *203*(4), 367-382.
- Kansal, A. R., Torquato, S., Harsh, I. G., Chiocca, E. A., & Deisboeck, T. S. (2000b). Cellular automaton of idealized brain tumor growth dynamics. *Biosystems*, *55*(1-3), 119-127.
- Kedrin, D., van Rheenen, J., Hernandez, L., Condeelis, J., & Segall, J. E. (2007). Cell motility and cytoskeletal regulation in invasion and metastasis. *J Mammary Gland Biol Neoplasia*, *12*(2-3), 143-152.
- Keller, P. J., Pampaloni, F., & Stelzer, E. H. K. (2006). Life sciences require the third dimension. *Current Opinion in Cell Biology*, *18*(1), 117-124.
- Keren, K., Pincus, Z., Allen, G. M., Barnhart, E. L., Marriott, G., Mogilner, A., *et al.* (2008). Mechanism of shape determination in motile cells. *Nature*, *453*(7194), 475-480.
- Khain, E., & Sander, L. M. (2006). Dynamics and pattern formation in invasive tumor growth. *Physical Review Letters*, *96*(18), -.
- Khain, E., Sander, L. M., & Stein, A. M. (2005). A model for glioma growth. *Complexity*, *11*(2), 53-57.
- Kharait, S., Hautaniemi, S., Wu, S., Iwabu, A., Lauffenburger, D. A., & Wells, A. (2007). Decision tree modeling predicts effects of inhibiting contractility signaling on cell motility. *BMC Syst Biol*, *1*, 9.
- Kholodenko, B. N., Demin, O. V., Moehren, G., & Hoek, J. B. (1999). Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem*, *274*(42), 30169-30181.
- Kim, J., Kang, K. K., & Lowengrub, J. (2004). Conservative multigrid methods for Cahn-Hilliard fluids. *Journal of Computational Physics*, *193*(2), 511-543.
- Kim, J. B. (2005). Three-dimensional tissue culture models in cancer biology. *Seminars in Cancer Biology*, *15*(5), 365-377.
- Kitano, H. (2002). Systems biology: a brief overview. *Science*, *295*(5560), 1662-1664.

- Kohno, N., Ohnuma, T., Kaneko, M., & Holland, J. F. (1988). Interactions of doxorubicin and cis-platin in squamous carcinoma cells in culture. *Br J Cancer*, *58*(3), 330-334.
- Kohno, N., Ohnuma, T., & Truog, P. (1994). Effects of hyaluronidase on doxorubicin penetration into squamous carcinoma multicellular tumor spheroids and its cell lethality. *J Cancer Res Clin Oncol*, *120*(5), 293-297.
- Kopfstein, L., & Christofori, G. (2006). Metastasis: cell-autonomous mechanisms versus contributions by the tumor microenvironment. *Cellular and Molecular Life Sciences*, *63*(4), 449-468.
- Korah, R. M., Sysounthone, V., Golowa, Y., & Wieder, R. (2000). Basic fibroblast growth factor confers a less malignant phenotype in MDA-MB-231 human breast cancer cells. *Cancer Res*, *60*(3), 733-740.
- Krishna, R., & Mayer, L. D. (2000). Multidrug resistance (MDR) in cancer. Mechanisms, reversal using modulators of MDR and the role of MDR modulators in influencing the pharmacokinetics of anticancer drugs. *Eur J Pharm Sci*, *11*(4), 265-283.
- Kuehn, H., Liberzon, A., Reich, M., & Mesirov, J. P. (2008). Using GenePattern for gene expression analysis. *Curr Protoc Bioinformatics*, Chapter 7, Unit 7 12.
- Kuiper, R. A. J., Schellens, J. H. M., Blijham, G. H., Beijnen, J. H., & Voest, E. E. (1998). Clinical research on antiangiogenic therapy. *Pharmacological Research*, *37*(1), 1-16.
- Kume, N., & Gimbrone, M. A., Jr. (1994). Lysophosphatidylcholine transcriptionally induces growth factor gene expression in cultured human endothelial cells. *J Clin Invest*, *93*(2), 907-911.
- Kunkel, P., Ulbricht, U., Bohlen, P., Brockmann, M. A., Fillbrandt, R., Stavrou, D., *et al.* (2001a). Inhibition of glioma angiogenesis and growth in vivo by systemic treatment with a monoclonal antibody against vascular endothelial growth factor receptor-2. *Cancer Res*, *61*(18), 6624-6628.
- Kunkel, P., Ulbricht, U., Bohlen, P., Brockmann, M. A., Fillbrandt, R., Stavrou, D., *et al.* (2001b). Inhibition of glioma angiogenesis and growth in vivo by systemic treatment with a monoclonal antibody against vascular endothelial growth factor receptor-2. *Cancer Research*, *61*(18), 6624-6628.
- Kunz-Schughart, L. A., Kreutz, M., & Knuechel, R. (1998). Multicellular spheroids: a three-dimensional in vitro culture system to study tumour biology. *Int J Exp Pathol*, *79*(1), 1-23.
- Lacroix, M., & Leclercq, G. (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. *Mol Cell Endocrinol*, *219*(1-2), 1-7.
- Lagana, K., Balossino, R., Migliavacca, F., Pennati, G., Bove, E. L., de Leval, M. R., *et al.* (2005). Multi-scale modeling of the cardiovascular system: application to the study of pulmonary and coronary perfusions in the univentricular circulation. *J Biomech*, *38*(5), 1129-1141.

- Lah, T. T., Alonso, M. B. D., & Van Noorden, C. J. F. (2006). Antiprotease therapy in cancer: hot or not? *Expert Opinion on Biological Therapy*, 6(3), 257-279.
- Lamszus, K., Kunkel, P., & Westphal, M. (2003). Invasion as limitation to anti-angiogenic glioma therapy. *Acta Neurochir Suppl*, 88, 69-77.
- Lankelma, J., Fernandez Luque, R., Dekker, H., Schinkel, W., & Pinedo, H. M. (2000). A mathematical model of drug transport in human breast cancer. *Microvasc Res*, 59(1), 149-161.
- Lauffenburger, D. A. (2000). Cell signaling pathways as control modules: complexity for simplicity? *Proc Natl Acad Sci U S A*, 97(10), 5031-5033.
- Lazzara, M. J., & Lauffenburger, D. A. (2009). Quantitative modeling perspectives on the ErbB system of cell regulatory processes. *Exp Cell Res*, 315(4), 717-725.
- Le Clainche, C., & Carlier, M. F. (2008). Regulation of actin assembly associated with protrusion and adhesion in cell migration. *Physiol Rev*, 88(2), 489-513.
- Lee, H. G., Lowengrub, J. S., & Goodman, J. (2002). Modeling pinchoff and reconnection in a Hele-Shaw cell. I. The models and their calibration. *Physics of Fluids*, 14(2), 492-513.
- Leo, P. H., Lowengrub, J. S., & Jou, H. J. (1998). A diffuse interface model for microstructural evolution in elastically stressed solids. *Acta Materialia*, 46(6), 2113-2130.
- Leung, Y. F., & Cavalieri, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends Genet*, 19(11), 649-659.
- Levasseur, L. M., Slocum, H. K., Rustum, Y. M., & Greco, W. R. (1998). Modeling of the time-dependency of in vitro drug cytotoxicity and resistance. *Cancer Res*, 58(24), 5749-5761.
- Levchenko, A., & Iglesias, P. A. (2002). Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. *Biophys J*, 82(1 Pt 1), 50-63.
- Levine, H. A., Pamuk, S., Sleeman, B. D., & Nilsen-Hamilton, M. (2002). Mathematical modeling of capillary formation and development in tumor angiogenesis: Penetration into the stroma. (vol 63, pg 801, 2001). *Bulletin of Mathematical Biology*, 64(2), 423-423.
- Levine, M. D., Liotta, L. A., & Stracke, M. L. (1995). Stimulation and regulation of tumor cell motility in invasion and metastasis. *EXS*, 74, 157-179.
- Leyrat, A., Duperray, A., & Verdier, C. (2003). Adhesion mechanisms in cancer metastasis. In L. Preziosi (Ed.), (pp. 221-242): CRC Press.
- Li, C. (2008). Automating dChip: toward reproducible sharing of microarray data analysis. *BMC Bioinformatics*, 9, 231.
- Li, X. R., Cristini, V., Nie, Q., & Lowengrub, J. S. (2007). Nonlinear three-dimensional simulation of solid tumor growth. *Discrete and Continuous Dynamical Systems-Series B*, 7(3), 581-604.
- Liang, K. C., & Wang, X. (2008). Gene regulatory network reconstruction using conditional mutual information. *EURASIP J Bioinform Syst Biol*, 253894.

- Lim, W. K., Wang, K., Lefebvre, C., & Califano, A. (2007). Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics*, 23(13), i282-288.
- Limbird, L. E. (2005). Introduction to Receptor Theory *Cell Surface Receptors: A Short Course on Theory and Methods* (pp. 1-28). New York, NY: Springer Science and Business Media, Inc.
- Lin, F., Nguyen, C. M., Wang, S. J., Saadi, W., Gross, S. P., & Jeon, N. L. (2005). Neutrophil migration in opposing chemoattractant gradients using microfluidic chemotaxis devices. *Ann Biomed Eng*, 33(4), 475-482.
- Liu, H., Zeeberg, B. R., Qu, G., Koru, A. G., Ferrucci, A., Kahn, A., *et al.* (2007). AffyProbeMiner: a web resource for computing or retrieving accurately redefined Affymetrix probe sets. *Bioinformatics*, 23(18), 2385-2390.
- Liu, Y. (2004). Active learning with support vector machine applied to gene expression data for cancer classification. *J Chem Inf Comput Sci*, 44(6), 1936-1941.
- Livasy, C. A., Karaca, G., Nanda, R., Tretiakova, M. S., Olopade, O. I., Moore, D. T., *et al.* (2006). Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol*, 19(2), 264-271.
- Lopes, R. H., Hobson, P. R., & Reid, I. D. (2008). *Computationally efficient algorithms for the two-dimensional Kolmogorov-Smirnov test*. Paper presented at the International Conference on Computing in High Energy and Nuclear Physics.
- Lowengrub, J., & Truskinovsky, L. (1998). Quasi-incompressible Cahn-Hilliard fluids and topological transitions. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, 454(1978), 2617-2654.
- Lu, Y., Lu, S., Fotouhi, F., Deng, Y., & Brown, S. J. (2004). Incremental genetic K-means algorithm and its application in gene expression data analysis. *BMC Bioinformatics*, 5, 172.
- Ma, L., Janetopoulos, C., Yang, L., Devreotes, P. N., & Iglesias, P. A. (2004). Two complementary, local excitation, global inhibition mechanisms acting in parallel can explain the chemoattractant-induced regulation of PI(3,4,5)P3 response in dictyostelium cells. *Biophys J*, 87(6), 3764-3774.
- Macklin, P., Kim, J., Tomaiuolo, G., Edgerton, M. E., & Cristini, V. (2009). Agent-Based Modeling of Ductal Carcinoma in Situ: Application to Patient-Specific Breast Cancer Modeling. In T. Pham (Ed.), *Computational Biology: Issues and Applications in Oncology*: Springer.
- Macklin, P., & Lowengrub, J. (2005). Evolving interfaces via gradients of geometry-dependent interior Poisson problems: application to tumor growth. *Journal of Computational Physics*, 203(1), 191-220.
- Macklin, P., & Lowengrub, J. (2007a). Erratum to "Nonlinear simulation of the effect of microenvironment on tumor growth" [J. Theor. Biol. 245 (2007) 677-704]. *Journal of Theoretical Biology*, 247(3), 581-581.
- Macklin, P., & Lowengrub, J. (2007b). Nonlinear simulation of the effect of microenvironment on tumor growth. *Journal of Theoretical Biology*, 245(4), 677-704.

- Maggelakis, S. A., & Adam, J. A. (1990). Mathematical model of prevascular growth of a spherical carcinoma. *Math. Comput. Modelling*, *13*, 23-38.
- Mantzaris, N. V., Webb, S., & Othmer, H. G. (2004). Mathematical modeling of tumor-induced angiogenesis. *J. Math. Biol.*, *49*(2), 111-187.
- Maree, A. F., Jilkine, A., Dawes, A., Grieneisen, V. A., & Edelstein-Keshet, L. (2006). Polarization and movement of keratocytes: a multi-scale modelling approach. *Bull Math Biol*, *68*(5), 1169-1211.
- Mareel, M., & Leroy, A. (2003). Clinical, cellular, and molecular aspects of cancer invasion. *Physiological Reviews*, *83*(2), 337-376.
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., *et al.* (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, *7 Suppl 1*, S7.
- Markowitz, F., & Spang, R. (2007). Inferring cellular networks--a review. *BMC Bioinformatics*, *8 Suppl 6*, S5.
- MartinezZaguilan, R., Seftor, E. A., Seftor, R. E. B., Chu, Y. W., Gillies, R. J., & Hendrix, M. J. C. (1996). Acidic pH enhances the invasive behavior of human melanoma cells. *Clinical & Experimental Metastasis*, *14*(2), 176-186.
- Materi, W., & Wishart, D. S. (2007). Computational Systems Biology in Cancer: Modeling Methods and Applications. *Gene Regulation and Systems Biology*, *1*(GRSB-1-Materi-et-al), 91-110.
- Materi, W., & Wishart, D. S. (2007). Computational systems biology in drug discovery and development: methods and applications. *Drug Discov Today*, *12*(7-8), 295-303.
- McDougall, S. R., Anderson, A. R., & Chaplain, M. A. (2006). Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. *J Theor Biol*, *241*(3), 564-589.
- McDougall, S. R., Anderson, A. R., Chaplain, M. A., & Sherratt, J. A. (2002). Mathematical modelling of flow through vascular networks: implications for tumour-induced angiogenesis and chemotherapy strategies. *Bull Math Biol*, *64*(4), 673-702.
- McGhee, J. D., & von Hippel, P. H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J Mol Biol*, *86*(2), 469-489.
- Mehlen, P., & Puisieux, A. (2006). Metastasis: a question of life or death. *Nat Rev Cancer*, *6*(6), 449-458.
- Mellor, H. R., Ferguson, D. J., & Callaghan, R. (2005). A model of quiescent tumour microregions for evaluating multicellular resistance to chemotherapeutic drugs. *Br J Cancer*, *93*(3), 302-309.
- Migliaccio, A., Castoria, G., Di Domenico, M., Ciociola, A., Lombardi, M., De Falco, A., *et al.* (2006). Crosstalk between EGFR and extranuclear steroid receptors. *Ann N Y Acad Sci*, *1089*, 194-200.

- Migliaccio, A., Castoria, G., Di Domenico, M., de Falco, A., Bilancio, A., Lombardi, M., *et al.* (2000). Steroid-induced androgen receptor-oestradiol receptor beta- Src complex triggers prostate cancer cell proliferation. *EMBO J*, *19*(20), 5406-5417.
- Miranda, E., Destro, A., Malesci, A., Balladore, E., Bianchi, P., Baryshnikova, E., *et al.* (2006). Genetic and epigenetic changes in primary metastatic and nonmetastatic colorectal cancer. *Br J Cancer*, *95*(8), 1101-1107.
- Mischel, P. S., Cloughesy, T. F., & Nelson, S. F. (2004). DNA-microarray analysis of brain cancer: molecular classification for therapy. *Nat Rev Neurosci*, *5*(10), 782-792.
- Moe, R. E., & Anderson, B. O. (2007). Androgens and androgen receptors: a clinically neglected sector in breast cancer biology. *J Surg Oncol*, *95*(6), 437-439.
- Mogilner, A. (2006). On the edge: modeling protrusion. *Curr Opin Cell Biol*, *18*(1), 32-39.
- Mogilner, A., & Edelstein-Keshet, L. (2002). Regulation of actin dynamics in rapidly moving cells: a quantitative analysis. *Biophys J*, *83*(3), 1237-1258.
- Mogilner, A., & Rubinstein, B. (2005). The physics of filopodial protrusion. *Biophys J*, *89*(2), 782-795.
- Mols, F., Vingerhoets, A. J., Coebergh, J. W., & van de Poll-Franse, L. V. (2005). Quality of life among long-term breast cancer survivors: a systematic review. *Eur J Cancer*, *41*(17), 2613-2619.
- Moreira, J., & Deutsch, A. (2002). Cellular automaton models of tumor development: a critical review. *Advances in Complex Systems*, *5*(2-3), 247-267.
- Mosadegh, B., Saadi, W., Wang, S. J., & Jeon, N. L. (2008). Epidermal growth factor promotes breast cancer cell chemotaxis in CXCL12 gradients. *Biotechnol Bioeng*, *100*(6), 1205-1213.
- Mouneimne, G., Soon, L., DesMarais, V., Sidani, M., Song, X., Yip, S. C., *et al.* (2004). Phospholipase C and cofilin are required for carcinoma cell directionality in response to EGF stimulation. *J Cell Biol*, *166*(5), 697-708.
- Mueller-Klieser, W. (1984). Microelectrode measurements of oxygen tension distributions in multicellular spheroids cultured in Spinner flasks *Spheroids in Cancer Research* (pp. 134-149). New York, NY: Springer-Verlag.
- Mueller-Klieser, W. (2000). Tumor biology and experimental therapeutics. *Critical Reviews in Oncology Hematology*, *36*(2-3), 123-139.
- Naderi, A., & Hughes-Davies, L. (2008). A functionally significant cross-talk between androgen receptor and ErbB2 pathways in estrogen receptor negative breast cancer. *Neoplasia*, *10*(6), 542-548.
- Nadon, R., & Shoemaker, J. (2002). Statistical issues with microarrays: processing and analysis. *Trends Genet*, *18*(5), 265-271.
- Nagy, J. D. (2005). The ecology and evolutionary biology of cancer: A review of mathematical models of necrosis and tumor cell diversity. *Mathematical Biosciences and Engineering*, *2*(2), 381-418.
- Nahleh, Z. (2008). Androgen receptor as a target for the treatment of hormone receptor-negative breast cancer: an uncharted territory. *Future Oncol*, *4*(1), 15-21.

- Nakshatri, H., & Badve, S. (2007). FOXA1 as a therapeutic target for breast cancer. *Expert Opin Ther Targets*, 11(4), 507-514.
- Nelson, P. S., Clegg, N., Arnold, H., Ferguson, C., Bonham, M., White, J., *et al.* (2002). The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci U S A*, 99(18), 11890-11895.
- Nestorov, I. S., Hadjitodorov, S. T., Petrov, I., & Rowland, M. (1999). Empirical versus mechanistic modelling: comparison of an artificial neural network to a mechanistically based model for quantitative structure pharmacokinetic relationships of a homologous series of barbiturates. *AAPS PharmSci*, 1(4), E17.
- Neves, S. R., & Iyengar, R. (2002). Modeling of signaling networks. *Bioessays*, 24(12), 1110-1117.
- Nikkila, J., Toronen, P., Kaski, S., Venna, J., Castren, E., & Wong, G. (2002). Analysis and visualization of gene expression data using self-organizing maps. *Neural Netw*, 15(8-9), 953-966.
- Norris, E. S., King, J. R., & Byrne, H. M. (2006). Modelling the response of spatially structured tumours to chemotherapy: drug kinetics. *Mathematical and Computer Modelling*, 43, 820-837.
- Nugent, L. J., & Jain, R. K. (1984a). Extravascular diffusion in normal and neoplastic tissues. *Cancer Res*, 44(1), 238-244.
- Nugent, L. J., & Jain, R. K. (1984b). Plasma pharmacokinetics and interstitial diffusion of macromolecules in a capillary bed. *Am J Physiol*, 246(1 Pt 2), H129-137.
- Nugent, L. J., & Jain, R. K. (1984c). Two-compartment model for plasma pharmacokinetics in individual blood vessels. *J Pharmacokinet Biopharm*, 12(4), 451-461.
- Ogawa, Y., Hai, E., Matsumoto, K., Ikeda, K., Tokunaga, S., Nagahara, H., *et al.* (2008). Androgen receptor expression in breast cancer: relationship with clinicopathological factors and biomarkers. *Int J Clin Oncol*, 13(5), 431-435.
- Ogino, S., & Goel, A. (2008). Molecular classification and correlates in colorectal cancer. *J Mol Diagn*, 10(1), 13-27.
- Padera, T. P., Stoll, B. R., Tooredman, J. B., Capen, D., di Tomaso, E., & Jain, R. K. (2004). Pathology: cancer cells compress intratumour vessels. *Nature*, 427(6976), 695.
- Page, D. L., Anderson, T. J., & Sakamoto, G. (1987) *Diagnostic histopathology of the breast* (pp. 219-222). New York, NY: Churchill Livingstone.
- Painter, K. J., Maini, P. K., & Othmer, H. G. (2000). Development and applications of a model for cellular response to multiple chemotactic cues. *Journal of Mathematical Biology*, 41(4), 285-314.
- Papin, J. A., Hunter, T., Palsson, B. O., & Subramaniam, S. (2005). Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*, 6(2), 99-111.
- Parent, C. A., & Devreotes, P. N. (1999). A cell's sense of direction. *Science*, 284(5415), 765-770.

- Parkin, D. M., Bray, F., Ferlay, J., & Pisani, P. (2005). Global cancer statistics, 2002. *CA Cancer J Clin*, 55(2), 74-108.
- Patel, A. A., Gawlinski, E. T., Lemieux, S. K., & Gatenby, R. A. (2001). A cellular automaton model of early tumor growth and invasion: The effects of native tissue vascularity and increased anaerobic tumor metabolism. *Journal of Theoretical Biology*, 213(3), 315-331.
- Peirce, S. M., Skalak, T. C., & Papin, J. A. (2006). Multi-scale biosystems integration: Coupling intracellular network analysis with tissue patterning simulations. *IBM J. Res. & Dev.*, 50(6), 601-615.
- Pennacchietti, S., Michieli, P., Galluzzo, M., Mazzone, M., Giordano, S., & Comoglio, P. M. (2003). Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell*, 3(4), 347-361.
- Peoples, G. E., Blotnick, S., Takahashi, K., Freeman, M. R., Klagsbrun, M., & Eberlein, T. J. (1995). T lymphocytes that infiltrate tumors and atherosclerotic plaques produce heparin-binding epidermal growth factor-like growth factor and basic fibroblast growth factor: a potential pathologic role. *Proc Natl Acad Sci U S A*, 92(14), 6547-6551.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797), 747-752.
- Perrault, D. J., Logan, D. M., Stewart, D. J., Bramwell, V. H., Paterson, A. H., & Eisenhauer, E. A. (1988). Phase II study of flutamide in patients with metastatic breast cancer. A National Cancer Institute of Canada Clinical Trials Group study. *Invest New Drugs*, 6(3), 207-210.
- Petalidis, L. P., Oulas, A., Backlund, M., Wayland, M. T., Liu, L., Plant, K., et al. (2008). Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol Cancer Ther*, 7(5), 1013-1024.
- Peterson, L. E. (2003). Partitioning large-sample microarray-based gene expression profiles using principal components analysis. *Comput Methods Programs Biomed*, 70(2), 107-119.
- Pirooznia, M., Yang, J. Y., Yang, M. Q., & Deng, Y. (2008). A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*, 9 Suppl 1, S13.
- Plank, M. J., & Sleeman, B. D. (2003). A reinforced random walk model of tumour angiogenesis and anti-angiogenic strategies. *Mathematical Medicine and Biology - a Journal of the IMA*, 20(2), 135-181.
- Plank, M. J., & Sleeman, B. D. (2004). Lattice and non-lattice models of tumour angiogenesis. *Bulletin of Mathematical Biology*, 66(6), 1785-1819.
- Please, C. P., Pettet, G. J., & McElwain, D. L. (2005). A new approach to modelling the formation of necrotic regions in tumours. *Applied Mathematics Letters*, 411, 89-94.

- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1996). *Numerical Recipes in Fortran 90: The Art of Scientific Computing* (2nd ed.): Cambridge University Press.
- Price, J. T., Tiganis, T., Agarwal, A., Djakiew, D., & Thompson, E. W. (1999). Epidermal growth factor promotes MDA-MB-231 breast cancer cell migration through a phosphatidylinositol 3'-kinase and phospholipase C-dependent mechanism. *Cancer Res*, *59*(21), 5475-5478.
- Pries, A. R., Cornelissen, A. J., Sloot, A. A., Hinkeldey, M., Dreher, M. R., Hopfner, M., *et al.* (2009). Structural adaptation and heterogeneity of normal and tumor microvascular networks. *PLoS Comput Biol*, *5*(5), e1000394.
- Pries, A. R., Reglin, B., & Secomb, T. W. (2001). Structural adaptation of vascular networks: role of the pressure response. *Hypertension*, *38*(6), 1476-1479.
- Pries, A. R., & Secomb, T. W. (2005). Control of blood vessel structure: insights from theoretical models. *Am J Physiol Heart Circ Physiol*, *288*(3), H1010-1015.
- Qu, X., Wan, C., Becker, H. C., Zhong, D., & Zewail, A. H. (2001). The anticancer drug-DNA complex: femtosecond primary dynamics for anthracycline antibiotics function. *Proc Natl Acad Sci U S A*, *98*(25), 14212-14217.
- Quackenbush, J. (2001). Computational analysis of microarray data. *Nat Rev Genet*, *2*(6), 418-427.
- Quaranta, V., Weaver, A. M., Cummings, P. T., & Anderson, A. R. A. (2005). Mathematical modeling of cancer: The future of prognosis and treatment. *Clinica Chimica Acta*, *357*(2), 173-179.
- Quehenberger, O., Prossnitz, E. R., Cavanagh, S. L., Cochrane, C. G., & Ye, R. D. (1993). Multiple domains of the N-formyl peptide receptor are required for high-affinity ligand binding. Construction and analysis of chimeric N-formyl peptide receptors. *J Biol Chem*, *268*(24), 18167-18175.
- Radinsky, R., Risin, S., Fan, D., Dong, Z., Bielenberg, D., Bucana, C. D., *et al.* (1995). Level and function of epidermal growth factor receptor predict the metastatic potential of human colon carcinoma cells. *Clin Cancer Res*, *1*(1), 19-31.
- Rafelski, S. M., & Theriot, J. A. (2004). Crawling toward a unified model of cell mobility: spatial and temporal regulation of actin dynamics. *Annu Rev Biochem*, *73*, 209-239.
- Rakha, E. A., El-Sayed, M. E., Green, A. R., Lee, A. H., Robertson, J. F., & Ellis, I. O. (2007). Prognostic markers in triple-negative breast cancer. *Cancer*, *109*(1), 25-32.
- Rangamani, P., & Iyengar, R. (2007). Modelling spatio-temporal interactions within the cell. *J Biosci*, *32*(1), 157-167.
- Rebhan, M., Chalifa-Caspi, V., Prilusky, J., & Lancet, D. (1997). GeneCards: integrating information about genes, proteins and diseases. *Trends Genet*, *13*(4), 163.
- Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., & Mesirov, J. P. (2006). GenePattern 2.0. *Nat Genet*, *38*(5), 500-501.
- Ribba, B., Colin, T., & Schnell, S. (2006). A multi-scale mathematical model of cancer, and its use in analyzing irradiation therapies. *Theor Biol Med Model*, *3*, 7.

- Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., *et al.* (2003). Cell migration: Integrating signals from front to back. *Science*, 302(5651), 1704-1709.
- Rizzo, V., Sacchi, N., & Menozzi, M. (1989). Kinetic studies of anthracycline-DNA interaction by fluorescence stopped flow confirm a complex association mechanism. *Biochemistry*, 28(1), 274-282.
- Roberts, H. C., Roberts, T. P. L., Lee, T. Y., & Dillon, W. P. (2002). Dynamic, contrast-enhanced CT of human brain tumors: Quantitative assessment of blood volume, blood flow, and microvascular permeability: Report of two cases. *American Journal of Neuroradiology*, 23(5), 828-832.
- Rodriguez-Fernandez, M., Mendes, P., & Banga, J. R. (2006). A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83(2-3), 248-265.
- Rofstad, E. K., & Danielsen, T. (1999). Hypoxia-induced metastasis of human melanoma cells: involvement of vascular endothelial growth factor-mediated angiogenesis. *British Journal of Cancer*, 80(11), 1697-1707.
- Roman Zantl, U. R. E. H. (2006). Chemotaxis in mu-Channels. *Imaging & Microscopy*, 8(1), 30-32.
- Rouzier, R., Perou, C. M., Symmans, W. F., Ibrahim, N., Cristofanilli, M., Anderson, K., *et al.* (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res*, 11(16), 5678-5685.
- Rubenstein, J. L., Kim, J., Ozawa, T., Zhang, M., Westphal, M., Deen, D. F., *et al.* (2000). Anti-VEGF antibody treatment of glioblastoma prolongs survival but results in increased vascular cooption. *Neoplasia*, 2(4), 306-314.
- Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics*, 9(1), 130-134.
- Ruths, D. A., Nakhleh, L., Iyengar, M. S., Reddy, S. A., & Ram, P. T. (2006). Hypothesis generation in signaling networks. *J Comput Biol*, 13(9), 1546-1557.
- Saadi, W., Rhee, S. W., Lin, F., Vahidi, B., Chung, B. G., & Jeon, N. L. (2007). Generation of stable concentration gradients in 2D and 3D environments using a microfluidic ladder chamber. *Biomed Microdevices*, 9(5), 627-635.
- Saadi, W., Wang, S. J., Lin, F., & Jeon, N. L. (2006). A parallel-gradient microfluidic chamber for quantitative analysis of breast cancer cell chemotaxis. *Biomed Microdevices*, 8(2), 109-118.
- Sadowitz, P. D., Hubbard, B. A., Dabrowiak, J. C., Goodisman, J., Tacka, K. A., Aktas, M. K., *et al.* (2002). Kinetics of cisplatin binding to cellular DNA and modulations by thiol-blocking agents and thiol drugs. *Drug Metab Dispos*, 30(2), 183-190.
- Sahai, E. (2005). Mechanisms of cancer cell invasion. *Current Opinion in Genetics & Development*, 15(1), 87-96.
- Sander, L. M., & Deisboeck, T. S. (2002). Growth patterns of microscopic brain tumors. *Physical Review E*, 66(5), -.
- Sanga, S., Broom, B. M., Cristini, V., & Edgerton, M. E. (In Press). Gene expression meta-analysis supports existence of molecular apocrine breast cancer with a role

- for androgen receptor and implies interactions with ErbB family. *BMC Medical Genomics*.
- Sanga, S., Frieboes, H., Sinek, J., & Cristini, V. (2007). A multi-scale approach for computational modeling of biobarriers to cancer chemotherapy via nanotechnology. In H. Nalwa (Ed.), *Cancer Nanotechnology* (pp. 1-21): American Scientific.
- Sanga, S., Frieboes, H. B., Zheng, X., Gatenby, R., Bearer, E. L., & Cristini, V. (2007). Predictive oncology: a review of multidisciplinary, multi-scale in silico modeling linking phenotype, morphology and growth. *Neuroimage*, *37 Suppl 1*, S120-134.
- Sanga, S., Macklin, P., & Cristini, V. (In Preparation). From Receptor Dynamics to Directed Cell Motion: A Predictive Agent-Based Model for Cell Motility in Complex Microenvironments.
- Sanga, S., Sinek, J. P., Frieboes, H. B., Ferrari, M., Fruehauf, J. P., & Cristini, V. (2006). Mathematical modeling of cancer progression and response to chemotherapy. *Expert Rev Anticancer Ther*, *6*(10), 1361-1376.
- Satulovsky, J., Lui, R., & Wang, Y. L. (2008). Exploring the control circuit of cell migration by mathematical modeling. *Biophys J*, *94*(9), 3671-3683.
- Schlappack, O. K., Zimmermann, A., & Hill, R. P. (1991). Glucose Starvation and Acidosis - Effect on Experimental Metastatic Potential, DNA Content and Mtx Resistance of Murine Tumor-Cells. *British Journal of Cancer*, *64*(4), 663-670.
- Schlessinger, J. (2002). Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell*, *110*(6), 669-672.
- Seftor, E. A., Meltzer, P. S., Kirschmann, D. A., Pe'er, J., Maniotis, A. J., Trent, J. M., *et al.* (2002). Molecular determinants of human uveal melanoma invasion and metastasis. *Clinical & Experimental Metastasis*, *19*(3), 233-246.
- Servant, G., Weiner, O. D., Herzmark, P., Balla, T., Sedat, J. W., & Bourne, H. R. (2000). Polarization of chemoattractant receptor signaling during neutrophil chemotaxis. *Science*, *287*(5455), 1037-1040.
- Shabalin, A. A., Tjelmeland, H., Fan, C., Perou, C. M., & Nobel, A. B. (2008). Merging two gene-expression studies via cross-platform normalization. *Bioinformatics*, *24*(9), 1154-1160.
- Shannon, A. M., Bouchier-Hayes, D. J., Condron, C. M., & Toomey, D. (2003). Tumour hypoxia, chemotherapeutic resistance and hypoxia-related therapies. *Cancer Treat Rev*, *29*(4), 297-307.
- Sharp, D. H., Reinitz, J., & Mjolsness, E. (1993). Multi-scale modeling of developmental processes. *Open Systems Information Dynamics*, *2*(1), 67-76.
- Shen, J. W., Subject, J. R., Lock, R. B., & Ross, W. E. (1989). Depletion of topoisomerase II in isolated nuclei during a glucose-regulated stress response. *Mol Cell Biol*, *9*(8), 3284-3291.
- Sherlock, G. (2001). Analysis of large-scale gene expression data. *Brief Bioinform*, *2*(4), 350-362.
- Sierra, A. (2005). Metastases and their microenvironments: linking pathogenesis and therapy. *Drug Resistance Updates*, *8*(4), 247-257.

- Silbergeld, D. L., & Chicoine, M. R. (1997). Isolation and characterization of human malignant glioma cells from histologically normal brain. *Journal of Neurosurgery*, 86(3), 525-531.
- Simon, S. M., & Schindler, M. (1994). Cell biological mechanisms of multidrug resistance in tumors. *Proc Natl Acad Sci U S A*, 91(9), 3497-3504.
- Sinek, J., Frieboes, H., Zheng, X., & Cristini, V. (2004). Two-dimensional chemotherapy simulations demonstrate fundamental transport and tumor response limitations involving nanoparticles. *Biomed Microdevices*, 6(4), 297-309.
- Sinek, J., Frieboes, H. B., Sivaraman, B., Sanga, S., & Cristini, V. (2006). Mathematical and computational modeling: Towards the development and application of nanodevices for drug delivery *Nanotechnologies for the Life Sciences* (Vol. 4, pp. 29-66): Wiley-VCH.
- Sinek, J. P., Sanga, S., Zheng, X., Frieboes, H. B., Ferrari, M., & Cristini, V. (2009). Predicting drug pharmacokinetics and effect in vascularized tumors using computer simulation. *J Math Biol*, 58(4-5), 485-510.
- Siu, W. Y., Arooz, T., & Poon, R. Y. (1999). Differential responses of proliferating versus quiescent cells to adriamycin. *Exp Cell Res*, 250(1), 131-141.
- Skupsky, R., Losert, W., & Nossal, R. J. (2005). Distinguishing modes of eukaryotic gradient sensing. *Biophys J*, 89(4), 2806-2823.
- Smith, N. P., Mulquiney, P. J., Nash, M. P., Braldehy, C. P., Nickerson, D. P., & Hunter, P. J. (2002). mathematical modeling of the heart: cell to organ. *Chaos Solitons Fractals*, 13, 1613-1621.
- Soinov, L. A. (2003). Supervised classification for gene network reconstruction. *Biochem Soc Trans*, 31(Pt 6), 1497-1502.
- Soinov, L. A., Krestyaninova, M. A., & Brazma, A. (2003). Towards reconstruction of gene networks from expression data by supervised learning. *Genome Biol*, 4(1), R6.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98(19), 10869-10874.
- Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., *et al.* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*, 98(4), 262-272.
- Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D., & Levy, S. (2005). A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*, 21(5), 631-643.
- Steeg, P. S. (2003a). Angiogenesis inhibitors: motivators of metastasis? *Nature Medicine*, 9(7), 822-823.
- Steeg, P. S. (2003b). Metastasis suppressors alter the signal transduction of cancer cells. *Nat Rev Cancer*, 3(1), 55-63.
- Stein, A. M., Demuth, T., Mobley, D., Berens, M., & Sander, L. M. (2007). A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment. *Biophysical Journal*, 92(1), 356-365.

- Stelzer, G., Harel, A., Dalah, A., Rosen, N., Shmoish, M., Iny-Stein, T., *et al.* (2008). *GeneCards: One stop site for human gene research*. Paper presented at the The 5th Congress of the Federation of the Israel Societies for Experimental Biology, Eliat, Israel.
- Stephanou, A., McDougall, S. R., Anderson, A. R., & Chaplain, M. A. (2005). Mathematical modeling of flow in 2D and 3D vascular networks: applications to anti-angiogenic and chemotherapeutic drug strategies. *Mathematical and computer Modelling*, *41*, 1137-1156.
- Stephanou, A., McDougall, S. R., Anderson, A. R., & Chaplain, M. A. (2006). Mathematical modeling of the influence of blood rheological properties upon adaptive tumour-induced angiogenesis. *Mathematical and Computer Modelling*, *44*, 96-123.
- Stern, H. (2008). EGFR Family Heterodimers in Cancer Pathogenesis and Treatment. In J. Haley & W. Gullick (Eds.), *EGFR Signaling Networks in Cancer Therapy* (First ed., pp. 15-30). New York City: Humana Press.
- Stetlerstevenson, W. G., Aznavoorian, S., & Liotta, L. A. (1993). Tumor-Cell Interactions with the Extracellular-Matrix during Invasion and Metastasis. *Annual Review of Cell Biology*, *9*, 541-573.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, *102*(43), 15545-15550.
- Sun, S. Y., Wheeler, M. F., Obeyesekere, M., & Patrick, C. W. (2005). A deterministic model of growth factor-induced angiogenesis. *Bulletin of Mathematical Biology*, *67*(2), 313-337.
- Swabb, E. A., Wei, J., & Gullino, P. M. (1974). Diffusion and convection in normal and neoplastic tissues. *Cancer Res*, *34*(10), 2814-2822.
- Swain, S. M. (2001). Tamoxifen for patients with estrogen receptor-negative breast cancer. *J Clin Oncol*, *19*(18 Suppl), 93S-97S.
- Swanson, K. R., Alvord, E. C., & Murray, J. D. (2000). A quantitative model for differential motility of gliomas in grey and white matter. *Cell Proliferation*, *33*(5), 317-329.
- Swanson, K. R., Bridge, C., Murray, J. D., & Alvord, E. C. (2003). Virtual and real brain tumors: using mathematical modeling' to quantify glioma growth and invasion. *Journal of the Neurological Sciences*, *216*(1), 1-10.
- Takata, R., Katagiri, T., Kanehira, M., Tsunoda, T., Shuin, T., Miki, T., *et al.* (2005). Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling. *Clin Cancer Res*, *11*(7), 2625-2636.
- Takemura, Y., Kobayashi, H., Miyachi, H., Hayashi, K., Sekiguchi, S., & Ohnuma, T. (1991). The influence of tumor cell density on cellular accumulation of doxorubicin or cisplatin in vitro. *Cancer Chemother Pharmacol*, *27*(6), 417-422.

- Tan, A. C., & Gilbert, D. (2003). Ensemble machine learning on gene expression data for cancer classification. *Appl Bioinformatics*, 2(3 Suppl), S75-83.
- Tan, A. R., & Swain, S. M. (2008). Therapeutic strategies for triple-negative breast cancer. *Cancer J*, 14(6), 343-351.
- Tan, Y. C., Fisher, J. S., Lee, A. I., Cristini, V., & Lee, A. P. (2004). Design of microfluidic channel geometries for the control of droplet volume, chemical concentration, and sorting. *Lab Chip*, 4(4), 292-298.
- Tanaka, T., Decuzzi, P., Cristofanilli, M., Sakamoto, J. H., Tasciotti, E., Robertson, F. M., *et al.* (2009). Nanotechnology for breast cancer therapy. *Biomed Microdevices*, 11(1), 49-63.
- Tanaka, T., Kaneda, Y., Li, T. S., Matsuoka, T., Zempo, N., & Esato, K. (2001). Digitonin enhances the antitumor effect of cisplatin during isolated lung perfusion. *Ann Thorac Surg*, 72(4), 1173-1178.
- Tannock, I. F., Lee, C. M., Tunggal, J. K., Cowan, D. S., & Egorin, M. J. (2002). Limited penetration of anticancer drugs through tumor tissue: a potential cause of resistance of solid tumors to chemotherapy. *Clin Cancer Res*, 8(3), 878-884.
- Tarasiuk, J., Frezard, F., Garnier-Suillerot, A., & Gattegno, L. (1989). Anthracycline incorporation in human lymphocytes. Kinetics of uptake and nuclear concentration. *Biochim Biophys Acta*, 1013(2), 109-117.
- Teicher, B. A., Holden, S. A., & Jacobs, J. L. (1987). Approaches to defining the mechanism of enhancement by Fluosol-DA 20% with carbogen of melphalan antitumor activity. *Cancer Res*, 47(2), 513-518.
- Teschendorff, A. E., Miremadi, A., Pinder, S. E., Ellis, I. O., & Caldas, C. (2007). An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol*, 8(8), R157.
- Teschendorff, A. E., Naderi, A., Barbosa-Morais, N. L., & Caldas, C. (2006). PACK: Profile Analysis using Clustering and Kurtosis to find molecular classifiers in cancer. *Bioinformatics*, 22(18), 2269-2275.
- Thomlinson, R. H., & Gray, L. H. (1955). The histological structure of some human lung cancers and the possible implications for radiotherapy. *Br J Cancer*, 9(4), 539-549.
- Thorat, M. A., Marchio, C., Morimiya, A., Savage, K., Nakshatri, H., Reis-Filho, J. S., *et al.* (2008). Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. *J Clin Pathol*, 61(3), 327-332.
- Thorne, B. C., Bailey, A. M., & Peirce, S. M. (2007). Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Brief Bioinform*, 8(4), 245-257.
- Tischkowitz, M., Brunet, J. S., Begin, L. R., Huntsman, D. G., Cheang, M. C., Akslén, L. A., *et al.* (2007). Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. *BMC Cancer*, 7, 134.
- Toronen, P., Kolehmainen, M., Wong, G., & Castren, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS Lett*, 451(2), 142-146.

- Tozlu, S., Girault, I., Vacher, S., Vendrell, J., Andrieu, C., Spyrtos, F., *et al.* (2006). Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. *Endocr Relat Cancer*, 13(4), 1109-1120.
- Troger, V., Fischel, J. L., Formento, P., Gioanni, J., & Milano, G. (1992). Effects of prolonged exposure to cisplatin on cytotoxicity and intracellular drug concentration. *Eur J Cancer*, 28(1), 82-86.
- Turner, S., & Sherratt, J. A. (2002). Intercellular adhesion and cancer invasion: A discrete simulation using the extended Potts model. *Journal of Theoretical Biology*, 216(1), 85-100.
- Turner, T., Epps-Fung, M. V., Kassis, J., & Wells, A. (1997). Molecular inhibition of phospholipase cgamma signaling abrogates DU-145 prostate tumor cell invasion. *Clin Cancer Res*, 3(12 Pt 1), 2275-2282.
- Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*, 98(9), 5116-5121.
- Tysnes, B. B., Larsen, L. F., Ness, G. O., Mahesparan, R., Edvardsen, K., GarciaCabrera, I., *et al.* (1996). Stimulation of glioma-cell migration by laminin and inhibition by anti-alpha 3 and anti-beta 1 integrin antibodies. *International Journal of Cancer*, 67(6), 777-784.
- Valafar, F. (2002). Pattern recognition techniques in microarray data analysis: a survey. *Ann N Y Acad Sci*, 980, 41-64.
- van't Veer, L. J., & Bernards, R. (2008). Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*, 452(7187), 564-570.
- van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., *et al.* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 347(25), 1999-2009.
- van Kempen, L. C. L. T., Ruiter, D. J., van Muijen, G. N. P., & Coussens, L. M. (2003). The tumor microenvironment: a critical determinant of neoplastic evolution. *European Journal of Cell Biology*, 82(11), 539-548.
- van Riel, N. A. (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief Bioinform*, 7(4), 364-374.
- Wang, D., & Lippard, S. J. (2005). Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov*, 4(4), 307-320.
- Wang, S. J., Saadi, W., Lin, F., Minh-Canh Nguyen, C., & Li Jeon, N. (2004). Differential effects of EGF gradient profiles on MDA-MB-231 breast cancer cell chemotaxis. *Exp Cell Res*, 300(1), 180-189.
- Wang, Z., Birch, C. M., & Deisboeck, T. S. (2008). Cross-scale sensitivity analysis of a non-small cell lung cancer model: linking molecular signaling properties to cellular behavior. *Biosystems*, 92(3), 249-258.
- Wang, Z., Zhang, L., Sagotsky, J., & Deisboeck, T. S. (2007). Simulating non-small cell lung cancer with a multi-scale agent-based model. *Theor Biol Med Model*, 4, 50.

- Ward, J. P., & King, J. R. (2003). Mathematical modelling of drug transport in tumour multicell spheroids and monolayer cultures. *Mathematical Biosciences*, 181, 177-207.
- Weigelt, B., Horlings, H. M., Kreike, B., Hayes, M. M., Hauptmann, M., Wessels, L. F. A., *et al.* (2008). Refinement of breast cancer classification by molecular characterization of histological special types. *The Journal of Pathology*, 216(2), 141-150.
- Wells, A. (2000). Tumor invasion: role of growth factor-induced cell motility. *Adv Cancer Res*, 78, 31-101.
- Wells, A., Harms, B., Iwabu, A., Koo, L., Smith, K., Griffith, L., *et al.* (2006). Motility signaled from the EGF receptor and related systems. *Methods Mol Biol*, 327, 159-177.
- Wells, A., Kassis, J., Solava, J., Turner, T., & Lauffenburger, D. A. (2002). Growth factor-induced cell motility in tumor invasion. *Acta Oncol*, 41(2), 124-130.
- Wells, C. M., & Ridley, A. J. (2005). Analysis of Cell Migration Using the Dunn Chemotaxis Chamber and Time-Lapse Microscopy. In J.-L. Guan (Ed.), *Cell Migration: Developmental Methods and Protocols* (pp. 31-41). Totowa, New Jersey: Humana Press, Inc.
- Wiley, H. S., Shvartsman, S. Y., & Lauffenburger, D. A. (2003). Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol*, 13(1), 43-50.
- Wise, S. M., Lowengrub, J. S., Frieboes, H. B., & Cristini, V. (2008). Three-dimensional multispecies nonlinear tumor growth--I Model and numerical method. *J Theor Biol*, 253(3), 524-543.
- Wise, S. M., Lowengrub, J. S., Kim, J. S., & Johnson, W. C. (2004). Efficient phase-field simulation of quantum dot formation in a strained heteroepitaxial film. *Superlattices and Microstructures*, 36(1-3), 293-304.
- Wise, S. M., Lowengrub, J. S., Kim, J. S., Thornton, K., Voorhees, P. W., & Johnson, W. C. (2005). Quantum dot formation on a strain-patterned epitaxial thin film. *Applied Physics Letters*, 87(13), -.
- Wolf, I., Bose, S., Williamson, E. A., Miller, C. W., Karlan, B. Y., & Koeffler, H. P. (2007). FOXA1: Growth inhibitor and a favorable prognostic factor in human breast cancer. *Int J Cancer*, 120(5), 1013-1022.
- Wolf, K., & Friedl, P. (2006). Molecular mechanisms of cancer cell invasion and plasticity. *British Journal of Dermatology*, 154, 11-15.
- Wong, K. K. (2009). Recent developments in anti-cancer agents targeting the Ras/Raf/MEK/ERK pathway. *Recent Pat Anticancer Drug Discov*, 4(1), 28-35.
- Wong, Y. F., Selvanayagam, Z. E., Wei, N., Porter, J., Vittal, R., Hu, R., *et al.* (2003). Expression genomics of cervical cancer: molecular classification and prediction of radiotherapy response by DNA microarray. *Clin Cancer Res*, 9(15), 5486-5492.
- Wu, F. X. (2008). Genetic weighted k-means algorithm for clustering large-scale gene expression data. *BMC Bioinformatics*, 9 Suppl 6, S12.

- Wu, F. Y. (1982). The Potts-Model. *Reviews of Modern Physics*, 54(1), 235-268.
- Wyckoff, J. B., Segall, J. E., & Condeelis, J. S. (2000). The collection of the motile population of cells from a living tumor. *Cancer Res*, 60(19), 5401-5404.
- Xie, H., Li, G., Ning, H., Menard, C., Coleman, C. N., & Miller, R. W. (2004). *3D voxel fusion of multi-modality medical images in a clinical treatment planning system*. Paper presented at the Computer-Based Medical Systems.
- Yamaguchi, H., Wyckoff, J., & Condeelis, J. (2005). Cell migration in tumors. *Current Opinion in Cell Biology*, 17(5), 559-564.
- Yang, S., & Chang, K.-C. (2002). Comparison of score metrics for Bayesian network learning. *IEEE Transactions on Systems, Man, and Cybernetics: Part A: Systems and Humans*, 32(3), 419-428.
- Yu, L., Watterson, S., Marshall, S., & Ghazal, P. (2008). Inferring Boolean networks with perturbation from sparse gene expression data: a general model applied to the interferon regulatory network. *Mol Biosyst*, 4(10), 1024-1030.
- Zhang, L., Athale, C. A., & Deisboeck, T. S. (2007). Development of a three-dimensional multi-scale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *J Theor Biol*, 244(1), 96-107.
- Zhang, L., Wang, Z., Sagotsky, J. A., & Deisboeck, T. S. (2009). Multi-scale agent-based cancer modeling. *J Math Biol*, 58(4-5), 545-559.
- Zheng, J. H., Chen, C. T., Au, J. L., & Wientjes, M. G. (2001). Time- and concentration-dependent penetration of doxorubicin in prostate tumors. *AAPS PharmSci*, 3(2), E15.
- Zheng, X., Wise, S. M., & Cristini, V. (2005). Nonlinear simulation of tumor necrosis, neo-vascularization and tissue invasion via an adaptive finite-element/level-set method. *Bull Math Biol*, 67(2), 211-259.
- Zheng, X. M., Lowengrub, J., Anderson, A., & Cristini, V. (2005). Adaptive unstructured volume remeshing - II: Application to two- and three-dimensional level-set simulations of multiphase flow. *Journal of Computational Physics*, 208(2), 626-650.
- Zhu, H. Y., & Rohwer, R. (1996). No free lunch for cross-validation. *Neural Computation*, 8(7), 1421-1426.
- Zutter, M. M., Santoro, S. A., Staats, W. D., & Tsung, Y. L. (1995). Reexpression of the Alpha(2)Beta(1) Integrin Abrogates the Malignant Phenotype of Breast-Carcinoma Cells. *Proceedings of the National Academy of Sciences of the United States of America*, 92(16), 7411-7415.

Vita

Sandeep Sanga was born in 1982 to Madan Mohan Sanga and Kameswari Sanga in San Jose, CA. Sandeep grew up in the Silicon Valley (California) and graduated from Monta Vista High School (Cupertino, CA) in 2000. During his final year of high school, Sandeep concurrently attended De Anza Community College (Cupertino, CA).

Following high school, Sandeep attended Revelle College at the University of California at San Diego and completed his B.S. in Bioengineering in 2004. Sandeep joined the graduate program in Biomedical Engineering at the University of California at Irvine where he completed his M.S. in 2006. Subsequently, Sandeep transferred to the University of Texas at Austin as a graduate student in the joint Biomedical Engineering program between the University of Texas at Austin, the University of Texas Health Science Center in Houston, and the M.D. Anderson Cancer Center.

Sandeep has also spent time teaching and mentoring students. From 2001-2004 as an undergraduate student in San Diego, Sandeep was a teaching assistant for introductory computer programming and general engineering physics courses. From 2005-2006, Sandeep was a graduate teaching assistant for biomedical engineering courses. In 2008 and 2009, Sandeep mentored undergraduate summer interns in cancer research projects.

Complementary to his academic studies, Sandeep has also worked part-time and as an intern in the biomedical industry. Sandeep worked as a laboratory assistant at the Burnham Institute (La Jolla, CA) from 2002-2003. He interned with the *In Vitro* Pharmacology Group at Ferring Pharmaceuticals Research (La Jolla, CA) in 2004. Following the completion of his B.S. in Bioengineering in 2004, Sandeep interned at Biogen IDEC (San Diego, CA) in the Cancer Target Discovery group. He later worked part-time for B. Wilson Partners (Irvine, CA) in 2006 and Tremonti Consulting (Fairfax,

VA) from 2007-2009 as an intellectual property asset management consultant specializing in the life sciences and biomedical devices. He also worked part-time as a scientific & marketing consultant for the Marketing and Communications Department at Affymetrix, Inc. (Santa Clara, CA) from 2006-2008.

List of Publications, Abstracts, Book Chapters, and Manuscripts:

1. **Sanga, S.**, Macklin, P.T., Cristini, V., From receptor dynamics to directed cell motion: A predictive agent-based model of cell motility in complex microenvironments (in preparation)
2. **Sanga, S.**, Broom, B.M., Cristini, V., Edgerton, M.E., Gene expression meta-analysis supports existence of molecular apocrine breast cancer with a role for androgen receptor and implies interactions with ErbB family, *BMC Medical Genomics*, (in press)
3. Edgerton, M.E., Chuang, Y.L., Macklin, P.T., **Sanga, S.**, Kim, J., Tamaiuolo, G., Yang, W., Broom, A., Do, K.A., Cristini, V. Using mathematical models to understand the time dependence of the growth of ductal carcinoma in situ. *Cancer Research* 69, 156s (2009)
4. Edgerton, M.E., **Sanga, S.**, Downs-Kelley, E., Cristini, V. Does androgen receptor have a role in a subgroup of ER-negative breast cancer? *Laboratory Investigation*. 89, 370a (2009) [Also in *Modern Pathology*. 22, 370a (2009)]
5. Edgerton, M.E., Chuang, Y.L., Macklin, P.T., Kim, J., Tomaiuolo, G., Broom, A.D., **Sanga, S.**, Cristini, V. Simulations of Growth of DCIS parameterized from IHC. *Laboratory Investigation*. 89, 37a-38a (2009) [Also in *Modern Pathology*. 22, 37a-38a (2009)]
6. Chen, T.C., **Sanga, S.**, Chou, T.Y., Cristini, V., Edgerton, M.E., Neural networks with K-means clustering via PCA for gene expression profile analysis. *World Congress for Computer Science and Information Engineering conference 2009*
7. Cristini, V., Frieboes, H.B., Li, X., Lowengrub, J.S., Macklin, P., **Sanga, S.**, Wise, S.M., and Zheng, X. Nonlinear modeling and simulation of tumor growth. In: *Modelling and Simulation in Science, Engineering and Technology*. Birkhauser. Boston, MA (In press)
8. Sinek, J.P., **Sanga, S.**, Zheng, X., Cristini, V. Predicting drug pharmacokinetics and effect in vascularized tumors using computer simulation. *J. Math. Biol.* 58, 485-510 (2009)
9. **Sanga, S.**, Frieboes, H.B., Zheng, X., Gatenby, R.A., Bearer, E. Cristini, V. Predictive Oncology: A review of multidisciplinary, multi-scale in silico modeling linking phenotype, morphology, and growth. *NeuroImage*. 31(S1):S120-S134 (2007)
10. **Sanga, S.**, Frieboes, H.B., Sinek, J.P., Cristini, V. A multi-scale approach for computational modeling of biobarriers to cancer chemotherapy via nanotechnology. In: *Nanotechnology Book to be published by American Scientific Publishers*, edited by H.S. Nalwa. (2007)
11. **Sanga, S.**, Sinek, J.P., Frieboes, H.B., Ferrari, M., Fruehauf, J.P., Cristini, V. Mathematical modeling of cancer progression and response to chemotherapy. *Expert Rev. Anticancer Ther.* 6(10):1361-1376 (2006)
12. Sinek, J., Frieboes, H.B., Sivaraman B., **Sanga, S.**, Cristini, V. Mathematical and computational modeling: towards the development and application of nanodevices for drug delivery. In: *Nanotechnologies for the Life Sciences 2005*, Wiley-VCH. 29-66 (2006)
13. Chin, V.I., Taupin, P., **Sanga, S.**, Scheel, J., Gage, F.H., Bhatia, S.N. Microfabricated platform for studying stem cell fates. *Biotechnol. Bioeng.* 88(3), 399-415 (2004)

Permanent address: 10125 Crescent Road, Cupertino, CA 95014

This dissertation was typed by Sandeep Sanga.