

# UC Irvine

## UC Irvine Electronic Theses and Dissertations

### Title

Mathematical modeling of cancer-immune interactions: agent-based and continuous modeling reveal novel, non-monotonic patterns

### Permalink

<https://escholarship.org/uc/item/1qm1v7ww>

### Author

Bergman, Daniel Roy

### Publication Date

2020

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,  
IRVINE

Mathematical modeling of cancer-immune interactions: agent-based and continuous  
modeling reveal novel, non-monotonic patterns

DISSERTATION

submitted in partial satisfaction of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Daniel Bergman

Dissertation Committee:  
Chancellor's Professor Qing Nie, Chair  
Professor German Enciso  
Professor Xing Dai

2020



# DEDICATION

To Caitlin, Caden, and any others who will join our family.

# TABLE OF CONTENTS

	Page
<b>LIST OF FIGURES</b>	<b>vi</b>
<b>LIST OF TABLES</b>	<b>viii</b>
<b>ACKNOWLEDGMENTS</b>	<b>ix</b>
<b>VITA</b>	<b>xi</b>
<b>ABSTRACT OF THE DISSERTATION</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The cancers we have studied . . . . .	3
1.1.1 Epithelial cancers . . . . .	3
1.1.2 Skin cancers . . . . .	5
1.2 The immune components we studied . . . . .	6
1.2.1 Pro-inflammatory immune composition . . . . .	7
1.2.2 Anti-inflammatory immune composition . . . . .	8
1.3 Mathematical preliminaries . . . . .	9
1.3.1 Agent-based modeling . . . . .	10
1.3.2 ODE modeling . . . . .	11
1.3.3 SDE modeling . . . . .	12
1.3.4 Parameter inference . . . . .	19
<b>2 Modeling the competing effects of the immune system and EMT on ep- ithelial cancers</b>	<b>22</b>
2.1 Summary . . . . .	22
2.2 Introduction . . . . .	23
2.3 Methods . . . . .	29
2.3.1 Tumor evolution . . . . .	29
2.3.2 Immune population dynamics . . . . .	30
2.3.3 Periodic cycling inflammation states . . . . .	30
2.3.4 Epithelial-to-mesenchymal transition . . . . .	30
2.3.5 Model simulation . . . . .	31
2.3.6 Parameter estimation and sensitivity analysis . . . . .	32
2.3.7 Analysis of patient survival data from TCGA . . . . .	34

2.4	Results . . . . .	36
2.4.1	A multiscale agent-based model of EMT-immune-tumor cell interactions to study tumor progression . . . . .	36
2.4.2	Identification of key model parameters via global sensitivity analysis . . . . .	37
2.4.3	Mesenchymal properties dramatically alter invasion-free survival times . . . . .	38
2.4.4	A key EMT regime maximizes cancer-free survival time under chronic inflammation . . . . .	40
2.4.5	TCGA data analysis supports predictions highlighting importance of mesenchymal cell proliferation in determining outcome . . . . .	42
2.5	Discussion . . . . .	43
2.6	Supplementary Information . . . . .	47
2.6.1	Model Description . . . . .	47
2.6.2	Definition of parameters specifying the model . . . . .	52
2.6.3	Parameter values used for simulation . . . . .	53
2.6.4	TCGA Analysis . . . . .	56
2.6.5	Cox-PH Tables . . . . .	64
2.6.6	KM Tables . . . . .	71
2.6.7	Extended TCGA Results . . . . .	72
2.7	Supplementary Figures . . . . .	75
<b>3</b>	<b>Modeling interactions among B cells, macrophages, and skin cancers</b>	<b>85</b>
3.1	Summary . . . . .	85
3.2	Introduction . . . . .	86
3.3	Results . . . . .	88
3.3.1	BCC and melanoma exhibit similar responses to checkpoint immunotherapy . . . . .	88
3.3.2	Memory B cells are more anergic in non-responders pre-treatment in both cancers . . . . .	90
3.3.3	Macrophages in BCC have more of an M1 genotype, regardless of responder status . . . . .	92
3.3.4	Plasma B cells are suppressing memory B cells only in BCC . . . . .	95
3.3.5	A dynamical model on interactions among memory B cells, macrophages, and skin tumors . . . . .	98
3.3.6	Multi-stability stemming from immune cell interactions in melanoma and BCC can explain heterogeneous response to checkpoint therapy . . . . .	101
3.3.7	Noise-induced cancer progression and regression potentially account for therapy-resistance in BCC . . . . .	102
3.4	Methods . . . . .	106
3.4.1	The three-component dynamical model . . . . .	106
3.4.2	Energy landscape and transition paths . . . . .	107
3.5	Discussion . . . . .	108
3.5.1	Experimental and clinical implications of the model . . . . .	109
3.6	Supplementary Information . . . . .	110
3.6.1	Literature-based assumptions of the dynamical model . . . . .	110
3.6.2	Non-dimensionalization of the dynamical model and parameter selection	111

3.6.3	Equilibria and their stability of deterministic model . . . . .	113
3.6.4	Adding stochastic effect to the model . . . . .	114
3.6.5	Model simulation methods and analysis . . . . .	115
3.6.6	The simplified gMAM algorithm for transition path . . . . .	117
3.6.7	Other interpretations of immunotherapy . . . . .	118
3.6.8	Pathways to target . . . . .	125
3.6.9	Comparison of melanoma and BCC sensitivity to immunotherapy . .	125
<b>4</b>	<b>Parameter Inference</b>	<b>129</b>
4.1	Summary . . . . .	129
4.2	Introduction . . . . .	129
4.3	Methods . . . . .	131
4.3.1	The ODE model . . . . .	131
4.3.2	Simulating therapy . . . . .	132
4.3.3	Parameter inference . . . . .	132
4.4	Results . . . . .	135
4.4.1	Confirmation of expected relationships and revelation of unexpected relationships between parameters . . . . .	135
4.5	Discussion . . . . .	137
	<b>Bibliography</b>	<b>138</b>

# LIST OF FIGURES

	Page
2.1 Model cartoon, sample trajectory, and sample survival curve . . . . .	35
2.2 Global sensitivity analysis of model parameters . . . . .	37
2.3 Effects of mesenchymal tumor cell properties on the time to invasion . . . . .	39
2.4 Effects of inflammation on the time to invasion under different cycling schemes	41
2.5 Genes predictive of invasiveness in BLCA . . . . .	44
2.6 Genes predictive of invasiveness in UCEC . . . . .	45
2.7 Fig. 2.1BC without targeting cells with pathways mutations . . . . .	75
2.8 Fig. 2.2 without targeting cells with pathways mutations . . . . .	76
2.9 Fig. 2.3 without targeting cells with pathways mutations . . . . .	77
2.10 Fig. 2.4E without targeting cells with pathways mutations . . . . .	78
2.11 Schematic of overall survival (OS) and disease-free interval (DFI) analysis on TCGA data . . . . .	79
2.12 DBSCAN clusters for combined EMT+INFLAM embedding of BLCA patients and corresponding Kaplan-Meier (KM) model . . . . .	80
2.13 DBSCAN clusters for combined EMT+INFLAM embedding of LIHC patients and corresponding Kaplan-Meier (KM) model . . . . .	80
2.14 DBSCAN clusters for combined EMT+INFLAM embedding of UCEC pa- tients and corresponding Kaplan-Meier (KM) model . . . . .	81
2.15 Classification of BLCA patients by DFI . . . . .	82
2.16 Classification of UCEC patients by DFI . . . . .	83
2.17 Classification of LIHC patients by DFI . . . . .	84
3.1 Melanoma and BCC have similar responses to immunotherapy . . . . .	89
3.2 Memory B cells are more activated in non-responders pre response . . . . .	91
3.3 Macrophages in BCC have more of a pro-inflammatory genotype, regardless of responder status . . . . .	94
3.4 Suppression of memory B cells by plasma B cells is only implicated in resis- tance to therapy in BCC . . . . .	97
3.5 ICB-targeted parameters can explain R vs. NR in melanoma and BCC . . . . .	100
3.6 Excess B cell counts predicts regressing cancer. Melanoma and BCC show different patterns for likelihood of progression. . . . .	105
3.7 Bifurcation in cancer-induced death rate . . . . .	119
3.8 Bifurcation for melanoma varying both $k$ and $d_e$ with $m = 0.5$ . . . . .	121
3.9 Bifurcation for BCC varying both $k$ and $d_e$ with $m = 0.5$ . . . . .	122

3.10	Bifurcation for melanoma varying both $k$ and $d_e$ with $m = 2$ . . . . .	123
3.11	Bifurcation for BCC varying both $k$ and $d_e$ with $m = 2$ . . . . .	124
3.12	Sensitivity of melanoma cancer burden to all parameters . . . . .	126
3.13	Sensitivity of BCC cancer burden to all parameters . . . . .	127
3.14	Comparison of melanoma and BCC sensitivity to immunotherapy . . . . .	128
4.1	Comparison of successful and unsuccessful convergence of HMC . . . . .	134
4.2	Examples of parameter relationships . . . . .	136

# LIST OF TABLES

	Page
2.1 Description of key model parameters. . . . .	33
2.2 Parameter names and descriptions . . . . .	53
2.3 Parameter values . . . . .	56
2.4 Gotzman EMT vs. GO Pos Acute Inflamm Ant . . . . .	64
2.5 GO Pos EMT vs. GO Leuk Act . . . . .	65
2.6 Hollern EMT Breast vs. GO Leuk Act . . . . .	65
2.7 GO Cardiac EMT vs. GO Neg Acute Inf . . . . .	66
2.8 GO Reg EMT vs. GO Neg Acute Inf . . . . .	66
2.9 GO Reg EMT vs. GO Neg Inf . . . . .	67
2.10 GO Cardiac EMT vs. Zhou Inf FIMA Up . . . . .	67
2.11 GO Reg EMT Endo vs. GO Mac Inf Prot 1 Alpha . . . . .	68
2.12 Alonso Met EMT Down vs. GO Neg Reg Inf . . . . .	68
2.13 Jechlinger EMT Down vs. GO Pos Reg Cyto Prod . . . . .	69
2.14 Alonso Met EMT Up vs. Fulcher Inf Resp Lectin LPS Down . . . . .	69
2.15 GO Cardiac EMT vs. GO Reg Inf Resp Wound . . . . .	70
2.16 GO Card EMT vs. Wunder Inf Resp Chol Up . . . . .	70
2.17 Gotzman EMT vs. GO Pos Acute Inflamm Ant . . . . .	71
2.18 GO Reg EMT Endo vs. GO Mac Inf Prot 1 Alpha . . . . .	71
2.19 GO Cardiac EMT vs. GO Reg Inf Resp Wound . . . . .	71
2.20 Summary of reported findings . . . . .	72
3.1 Mathematical model parameters . . . . .	112
3.2 Defining relations of non-dimensionalized ODE . . . . .	113
3.3 Noise parameters . . . . .	114

# ACKNOWLEDGMENTS

I would like to thank all those who have helped me on my journey here.

To all the educators in my life, thank you for your collective years of guidance and development. To my professors at CSULB, thank you for your careful instruction and the personal attention you were able to give. Special thanks to John Brevik for advising me through the process of writing my undergraduate thesis and cementing my desire to pursue a PhD. To my professors at UCI, thank you for sharing your expertise and shaping the mathematician I am today.

Thanks to three people in particular are needed: Qing Nie, Adam MacLean, and Peijie Zhou. You all have been instrumental in aiding in the transition to an applied mathematician and directing my research towards useful and achievable goals. Qing, as my advisor, you believed in me from the beginning and never gave up on me. Without your persistent guidance and encouragement, I would not be reaching this milestone. Adam, as my guide in completing my first project, you always supported me and saw my potential. I owe you lunch for taking on the role of project manager (and so much more). Peijie, as the Adam for my second project, your combination of friendliness and collegiality got me further in these six months than I thought possible. In such a short time, you exposed me to new tools and deeper perspectives in mathematical biology that I will carry forward into future research.

I would also like to thank Matt Karikomi and Emmanuel Dollinger for their collaborative efforts on these projects. Matt, your passion for and dedication to your work pushed our project so much further than I ever could have taken it. Emmanuel, your vision and determination are responsible for my opportunity to contribute on our project.

Thank you, also, to all those in Qing's lab for sharing your own great ideas and also providing me feedback on all mine. You all possess a contagious passion for math and its applications.

There have also been countless others who have shaped me over the years into the person I am today. To name a few: the President's Scholars program at CSULB, Don Allen and the Long Beach Navs, the Dans and CSULB's Ultimate teams, all who supported me on my China travels, the Elys all who were with me on those travels, the Mathletes, Dennis Eichhorn, Alessandra Pantano, staff of the math department, Mariners, our awesome lifegroup, Danny Mann and the PFFs, my officemates, all my other peers in the math department, and many more. Not enough can be said about all of you.

I also have my family to thank for most of who I am and what I have been able to accomplish. Thank you Mom, Dad, and Michael for everything you all have done and continue to do. Good luck, Michael, at NC State! Thank you, Caitlin, for five and a half years of walking this journey with me. You have seen all me this whole way and loved me through it all. Thank you to my in-laws for lovingly accepting me into your family and being a place of rest and refuge all these years.

Caden, you have been an amazing part of life this past year. You are a joy and wonder and

I hope to one day share all the fun parts of this with you.

# VITA

## Daniel Bergman

### Education

- **University of California, Irvine: 2014 - present**
  - Mathematics - Master of Science (Ph.D. expected 2020)  
Advisor: Dr. Qing Nie  
Ongoing research in mathematical biology
- **California State University, Long Beach: 2008 - 2012**
  - Mathematics - Bachelor of Science
  - Physics - Minor  
Honors thesis: *Analytic Isomorphisms of Curve Singularities*.  
Graduated Summa Cum Laude. Graduated from the Honors Program.

**Current position** *Graduate Student*, Mathematics Department, University of California at Irvine

### Areas of specialization

Mathematics, Applied Mathematics, Mathematical Biology, Pedagogy

### Teaching Experience

- UCI Mathematics Department, Teaching Assistant: 2014 - present
  - Courses Taught:
    - \* The Calculus Series: 2A, 2B, 2D, 2E
    - \* Linear Algebra: 3A, H3A, 121A
    - \* Differential Equations: 3D
    - \* Intro to Abstract Math: 13
    - \* Mathematical Modeling in Biology: 113A, 113B
    - \* Abstract Algebra: 120B
- Taught virtually all lower division courses with class sizes from 20-60.
- Taught a broad spectrum of upper division courses with class sizes from 10-50.
- Led biweekly discussion sections.

- Designed worksheets for in-class active learning.
- Implemented think-pair-share model of active learning.
- Created and graded weekly quizzes.
- Held office hours to guide students in one-on-one interactions.
- Utilized EEE/Canvas Learning Management System to create course websites.
- Attended over 90 hours of department-specific pedagogical training seminars.
- Received consistently high instructor evaluations ranging up to 4.00.
- Promoted an inclusive environment for a diverse student body.

## Publications & talks

- **Bergman, D.**, MacLean, A., Karikomi, M., Nie, Q. (2019). *Modeling the competing effects of the immune system and EMT on epithelial cancers*. Submitted.
- **Bergman, D.**, Dollinger, E., Nie, Q. *Single-cell-data-informed mathematical model reveals diagnostic targets for melanoma immunotherapy*. In preparation.
- *Multiscale modeling of the competing roles of the immune system in EMT-Mediated cancer*. Computational and Mathematical Population Dynamics 5. May 2019. Invited Talk.
- *Increasing students' conceptual understanding of linear algebra through active discussion sessions*. Joint Mathematical Meetings. January 2020. Talk.

## Leadership Roles

- *Executive Committee Member*, SIAM at UCI: 2018 - present  
Guiding new leaders through our first transition of leadership.
- *Founding President*, SIAM at UCI: 2017 - 2018  
Crafted the club constitution, established the club, and set the vision for the future.
- *Team Captain*, Mathletes: 2017 - 2018  
Organized several math department intramural sports teams by recruiting players, holding practices, handling administrative responsibilities, and leading on the field.
- *Coordinator*, UCI Mathematics Graduate Student Colloquium: 2017 - 2019  
Continued the long-standing practices of the MGSC and also increased focus on attracting first and second year students to the colloquia. Trained the next coordinators.

- *Coordinator*, UCI Mathematics Department First Year Mentor Program: 2017 - 2018  
Formalized the mentor program by establishing regular events, creating an accountability system, and successfully advocating for a yearly budget from the math department.
- *President and Captain*, CSULB Men's Ultimate Team: 2011 - 2012  
As president, handled big-picture ideas while delegating day-to-day details to other officers.  
As captain, planned practices, developed team strategy, oversaw player development, and maintained on-field and off-field morale.
- *President*, CSULB Society of Physics Students: 2009 - 2010  
Rebuilt the club. Organized club activities. Recruited and trained future leaders.

## Service

- *Mentor*, High School Student Researcher: 2019  
Mentored high school researcher. Manuscript in preparation.
- *Instructor*, Teaching Assistant Professional Development Program: 2019  
Prepared and conducted a two day workshop for incoming TAs preparing them for all facets of teaching at UCI as a graduate student.
- *Peer Mentor*, UCI Mathematics Department: 2016 - 2019  
Guided first year math graduate students through the transition to graduate school.
- *Tutor*, Math Community Educational Outreach: 2015 - 2017  
Tutored math to middle school students from underprivileged communities for two hours once per week.
- *ESL Tutor*, South Coast Literacy Council: 2018 - present  
Planned and taught English lessons to a small class two hours per week.

## Awards, Recognition, and Research Funding

- Graduate Student Research Summer Fellowship: 2016-2019
- Outstanding Teaching Assistant, UCI Mathematics Department: 2016 - 2017
- Outstanding Contributions to the Department, UCI Mathematics Department: 2015 - 2016
- Graduated Summa Cum Laude, CSULB: 2012
- Graduated with Honors, CSULB: 2012
- Presidential Scholarship, CSULB: 2008-2012

# ABSTRACT OF THE DISSERTATION

Mathematical modeling of cancer-immune interactions: agent-based and continuous modeling reveal novel, non-monotonic patterns

By

Daniel Bergman

Doctor of Philosophy in Mathematics

University of California, Irvine, 2020

Chancellor's Professor Qing Nie, Chair

The study of cancer-immune dynamics is broad. There are myriad instances of these dynamics and much heterogeneity among the two. To explore these relationships in their fullness, mathematical modeling is used to go further faster than can be done by experiments alone. In this work, two models of cancer-immune dynamics are explored and their applications to clinical settings are predicted. The mathematical techniques are also explicated. The first model looks at epithelial-to-mesenchymal transition in epithelial cancers and the effects on progression and invasive disease. We find and validate evidence that key parameters in this process control the time to invasion in non-monotonic ways. The second model studies the newly discovered relevance of B cells to immune checkpoint therapy in two skin cancers: melanoma and BCC. We find an explanation for the difference in response rates between the two cancers as well as a means to assess the sustained effects of immunotherapy by a single sample of cells. Finally, we utilize a nascent tool for parameter inference and show how it could be applied to an ODE model and show the types of results such a tool can produce.

# Chapter 1

## Introduction

The human body is composed of an incredibly diverse collection of cells undergoing an even greater number of processes. This complexity presents one of the greatest frontiers of science to explore. Occupying a central position on this frontier is the phenomenon of cancer. Simply put, cancer is a disease of uncontrolled growth. Better put, cancer is cells growing unconstrained from some of the normal feedback processes active in healthy cells. As the cancer cells grow and divide, they can grow beyond what the body can sustain and lead to death. According to the CDC, cancer is the second most common cause of death in the United States in 2017 with nearly 600,000 deaths attributed to the disease that calendar year[57].

Moreover, cancer has proven difficult to understand. By its very nature, cancer manifests in different people in different ways. In fact, within a single tissue type, there can be multiple forms of cancer with varying sources and varying disease trajectories. It is a disease that has confounded scientists for millennia and defied being cured.

In response, the scientific community has left no stone unturned in the pursuit of treatments. The history of this struggle is rich and enthralling (see Siddhartha Mukherjee's excellent

treatment of this subject [98]). We have developed many means of fighting cancer: surgery, radiotherapy, chemotherapy, and immunotherapy to name a few. Ultimately, the goal is to augment these existing therapies and possibly add new ones.

Here, our more modest goal is to look at the nascent field of immunotherapy. The immune system was long thought to be incapable of defending the body against cancer, but in the last decade, new therapies have emerged that target the immune system and done so to great success. However, much remains uncertain. What cancers are most amenable to such treatment? What cells and cellular processes can be most efficiently and safely targeted? Why do these therapies work for some patients, but not others?

We have sought to offer some answers to some of these questions and lay them out in the following pages. Specifically, we use mathematical modeling to experiment on *in silico* patients in an effort to better understand oncoimmunology. The remainder of this chapter is devoted to filling in the background details on the cancers we have studied (Section 1.1), the components of the immune system we explored as relevant to those cancers (Section 1.2), and the mathematical tools we have deployed in our approach (Section 1.3). In Chapter 2, we look at how epithelial-to-mesenchymal transition (EMT) can change how a cancer will respond to the immune system and observe that there is a “sweet spot” for mesenchymal parameters that lead to better patient outcomes. In Chapter 3, we look at the emerging role of B cells in immune checkpoint blockade (ICB) therapy and observe how these interactions inform treatment and predict why ICB could have different responses between patients and between types of skin cancer. In Chapter 4, we present some preliminary work using Bayesian methods for parameter inference with some results for an ODE model.

## 1.1 The cancers we have studied

The first group of cancers we study are epithelial cancers. These cancers have the potential for an epithelial-to-mesenchymal transition (EMT) program to be initiated, and recent work has shed light on the potential for this program to have implications in cancer, including its interaction with immune cells[102].

The second group of cancers we study are skin cancers. Melanoma and basal cell carcinoma are two skin cancers that in the last decade have shown to have some level of responsiveness to immune checkpoint blockade therapy[87, 120]. It is, however, unknown exactly why only some patients respond to immunotherapy and not others. It is also unclear how the differences between these two particular skin cancers will manifest in their responses to immunotherapy.

### 1.1.1 Epithelial cancers

Epithelial cancers are cancers that originate in epithelial tissue. Epithelial cells are characterized –in part – by their cell-cell adhesion and thus tightly regulated structure. This cell state is malleable, however, as epithelial cells can lose this adhesive property and gain migratory properties, among others, through EMT to become mesenchymal cells. Typically, this pathway is not engaged in adult tissues. However, EMT has been studied as a potential mechanism for cancer cells to both escape immune surveillance and undergo metastasis. We were thus prompted to study EMT and, in the course of our analysis, two cancers in particular: bladder (BLCA) and uterine (UCEC) cancer.

## EMT

Epithelial-to-mesenchymal transition (EMT) describes a reversible process by which cells displaying an epithelial phenotype transition into cells with a mesenchymal phenotype. Epithelial cells are – in part – defined by tight cell-cell adhesion. Mesenchymal cells exhibit less adhesion, greater ranges of motility, and may possess stem-like properties [102], although controversy regarding ‘stemness’ and EMT remains [101, 132]. Recent work has shown that – rather than being a binary process – at least two stable intermediate EMT states exist [59, 65]. Ongoing investigations into the plasticity and stability of EMT overlap with discussions elsewhere, e.g. of discrete vs. continuous processes during cell differentiation [96]. Intermediate states have emerged as a central mechanism by which cell fates (and the noise inherent within them) can be controlled [92, 117, 140].

Two features of the mesenchymal phenotype are of particular relevance in the context of cancer-immune interactions. i) mesenchymal tumor cells proliferate less than epithelial cells, we refer to this as mesenchymal growth arrest (MGA), and can be considered related to (in the sense of quiescence) the “stemness” phenotype of the mesenchymal tumor cells [158]. ii) mesenchymal cells are less susceptible to immune clearance [143]. As a cell is targeted by cytotoxic immune cells for clearance, a physical connection between the two cells must be established. This immunological synapse – mediated in part by T-cell receptors bound to antigens and the major histocompatibility complex on the target cell – is down-regulated in mesenchymal cells, thus inhibiting formation of the synapse [143]. We refer to this phenotype as mesenchymal immune evasion (MIE).

In addition to the prominent role it plays in metastasis, EMT has more recently been shown to also regulate other aspects of tumor progression [102, 110]. TGF- $\beta$ , a master regulator of EMT [84], is at once implicated heavily in tumor-mediated immune responses, since T regulatory cells release TGF- $\beta$  upon arriving at the tumor site[143]. In hepatocellular

carcinoma, for example, there is direct evidence linking Treg-secreted TGF- $\beta$  with EMT [136]. Thus, even by considering only the TGF- $\beta$  pathway, we find compelling evidence that these three core components (the tumor, the immune system, and EMT) all interact. It therefore strikes us as a priority to develop models to understand how the interactions between each of these three components affect cancer incidence and progression.

### 1.1.2 Skin cancers

As mentioned, much work recently has focused on using immunotherapy to treat two particular skin cancers: melanoma and BCC. Melanoma has proven amenable to this therapy modality whereas BCC is still in the early stages of understanding this question. What remains unclear in this realm, is why some patients respond to therapy while others do not. We also have recent single-cell analysis of melanoma and BCC patients both before and after immunotherapy that implicates previously unexplored immune populations and pathways as predictive of the success or failure of the therapy. This discovery coincided with a trio of separately published studies all observing the same phenomenon. We thus studied melanoma and BCC.

#### Melanoma

Melanoma is a particularly dangerous skin cancer[95]. It is also a highly immunogenic skin cancer which has led to its position at the center of cancer immunotherapy[109]. Though recent results in treating melanoma are encouraging, the prognosis for the disease remains poor.

Melanoma begins with melanocytes in the epidermis mutating and growing in that layer. Eventually, it can manage to break through the basement membrane and begin to metastasize

to distal locations.

## **BCC**

Basal cell carcinoma (BCC) is another skin cancer that has shown some responsiveness to immunotherapy, though testing is still in the early stages[87]. By contrast with melanoma, however, BCC is much more common and much less deadly. It is also far less immunogenic with natural killer cells and B cell in particular being absent from the tumor site[14, 55].

BCC originates in the basal cell layer of the epidermis. Similar to melanoma, it spreads from there down into the dermis before possibly invading into other sites.

## **1.2 The immune components we studied**

The effects of the immune system on a tumor can be broadly summarized into two branches. The pro-inflammatory branch of the immune system – such as natural killer cells (NKs), cytotoxic T cells (CTLs), and B cells – exert negative regulation of the tumor by eliminating cancer cells. The cancer, in turn, often exerts reciprocal effects on these cells causing them to lose efficacy, deactivate, or transition to an immunosuppressive role[39].

The anti-inflammatory branch of the immune system – such as T regulatory cells (Tregs) and tumor-associated macrophages (TAMs) – inhibits the effective functioning of the pro-inflammatory branch [123]. This can be carried out by suppressing activation, proliferation, or efficacy and by up-regulating cell death and deactivation. The cancer often acts to attract such anti-inflammatory cells and enhance their functions to create a more pro-tumor microenvironment.

### **1.2.1 Pro-inflammatory immune composition**

The pro-inflammatory immune cells are responsible for initiating and maintaining a robust response to a pathogen. Many roles are taken on by such cells, but typically only a subset are of interest in any one experiment and model. We will focus on cells that have been shown to directly target cancer cells, rather than those that recruit or activate such cells. There is also a whole host of cytokines and other chemical signaling factors that are at work which we ignore here in our treatment.

#### **Natural killer cells**

Natural killer cells are part of the body's innate immune system. They are ever-present in tissues, even in the absence of pathogens. This way, they serve as a front-line for the body's defenses. They are capable of recognizing cancer cells to a limited extent and subsequently clearing them. However, this ability is significantly weaker than that of CTLs and in fact has been shown to follow a different functional form as they kill cancer cells[112].

#### **Cytotoxic T lymphocytes**

CTLs are part of the body's adaptive immune system. A long sequence of events starting with the recognition of a cancer antigen eventually results in the activation of CTLs, usually in a lymph node. Once activated, CTLs begin to proliferate and migrate to the site of the insult. Following chemotactic gradients, CTLs move throughout the tumor microenvironment (TME) checking cells for surface markers that will identify the cell as a cancer cell. The efficacy of CTLs is significantly higher than that of NKs and other immune cells, and thus CTLs are often the central focus in immunotherapy.

## Memory B cells

Memory B cells are B cells that have developed an affinity for a specific antibody and await the reintroduction of the pathogen releasing that antibody. Once activated, memory B cells will differentiate into activated B cells that proliferate and release large quantities of antibodies that will attach to cancer cells, leading to the death of the cancer cells.

Recent work has challenged the notion that CTLs are the most important component of an immune response, specifically in the context of immune checkpoint blockade[56, 111, 15]. These studies found that the presence of B cells and the tertiary lymphoid structures that house them better predicts response to immunotherapy than previously studied metrics, often related to CTLs. My collaborator's analysis of several published single-cell data sets revealed that memory B cells, in particular, play a major role in immunotherapy, not only prognostically but also functionally.

### 1.2.2 Anti-inflammatory immune composition

In contrast to the pro-inflammatory branch of the immune system, the anti-inflammatory branch works to end the immune response so as to ensure that the immune system does not begin to indiscriminately attack healthy cells. Again, there are many roles to be played and we will focus on two cell populations that directly work against CTLs and B cells. There is also one cytokine,  $TGF-\beta$ , that is especially important in EMT.

## T regulatory cells

Tregs, like CTLs, are part of the adaptive immune system and so have a similar activation process. Once they enter the TME, they release copious quantities of cytokines, including  $TGF-\beta$ , that are responsible for deactivating CTLs and other pro-inflammatory immune

cells.

### **Tumor-associated macrophages**

Tumor-associated macrophages (TAMs) are a type of macrophage displaying an anti-inflammatory signature. They are often attracted to the TME via hypoxic and apoptotic signaling, both of which exist in large quantities, especially at the center of a tumor. They are sources of many different cytokines that stymie various components of the pro-inflammatory branch. My collaborator's analysis (the same as mentioned above) shows that there is communication between these TAMs and memory B cells with the memory B cells responding by losing efficacy.

### ***TGF- $\beta$***

*TGF- $\beta$*  is an important cytokine in cancer. It is necessary for Treg activation in lymph nodes, it plays a role in deactivating CTLs in the TME, and it is a master regulator of EMT.

## **1.3 Mathematical preliminaries**

To approach these problems, we use mathematical tools. The purpose of such tools is to understand details of the systems involved in ways that might be difficult to do in a laboratory setting. We use the following tools for such purpose.

### 1.3.1 Agent-based modeling

Agent-based modeling (ABM) is an approach by which each cell is represented as an agent and these agents interact according to a set of rules. At each discrete update to the system, every cell chooses from among a set of options for what it will do next, including nothing. The probabilities at which the cells choose from among these options is one of the chief focuses in developing such a model.

In the model studied in Chapter 2, the agents are the cancer cells. We chose to model the immune infiltrate (NKs, CTLs, and Tregs) as continuous variables whose role is to alter the cancer cells' probabilities for doing the different options.

Our model also has another component, the inflammation score, that acts system-wide to influence both the cancer cells and the immune cells.

#### MOAT

To understand the sensitivity of the model to the large set of parameters, we implemented the Morris one-at-a-time (MOAT) algorithm. The algorithm works by first choosing probability distributions for each parameter. Then, from several randomly sampled points in parameter space, the algorithm perturbs one parameter at a time to see how the outcome of the model changes with each change. The absolute change to the studied metric of the system is recorded and the parameters that affect this the most end up with the highest sensitivity.

#### KM curves

A Kaplan-Meier (KM) curve shows the probability of survival until a given time. When data is right-censored, the computation of a KM curve is more complicated. However, we

are computing KM curves for synthetic data and so they are all censored at the same time making the computation of the curve straightforward. If  $\{(t, f(t)) \mid t \geq 0\}$  is the graph of the curve, then  $f(t) = n(t)/N$  where  $n(t)$  is the number still surviving at time  $t$  and  $N = n(0)$ .

### 1.3.2 ODE modeling

Ordinary differential equations (ODEs) are one of the most common types of mathematical modeling. It describes the rate of change of a set of state variables as a function of time and the state variables themselves. Solving an ODE given an initial condition amounts to integrating.

#### Numerical solutions

To numerically solve an ODE, we use MATLAB's *ode45*. It relies on an explicit Runge-Kutta (4,5) formula that uses only the previous time point to compute the next one. We also use the nonnegativity constraint as our state variables represent quantities (or concentrations) of cells that must remain greater than or equal to 0.

#### Equilibria and their stability

At any point in state space at which the rates of change of all the state variables are simultaneously 0 is an equilibrium. If the system is ever at such a point, it will not evolve away from that point.

If in addition, small changes to the state variables result in the system returning to an equilibrium, then the equilibrium is called stable. This can be assessed by computing the Jacobian of the ODE function and analyzing the eigenvalues. If they all have negative real

part, then the equilibrium is stable. If any have a positive real part, then the equilibrium is unstable.

## **Bifurcation analysis**

As parameters in an ODE model are varied, the equilibria of the system can undergo large changes. Whenever a change in parameter values results in the stability of an equilibrium to change, this is called a bifurcation. Sometimes, this could mean that an equilibrium disappears entirely, as is the case with a saddle bifurcation.

### **1.3.3 SDE modeling**

To model the noise inherent in biological systems, we can include noise in ODE models and thus turn them into stochastic differential equations (SDEs). In such a model, the stability of an equilibrium does not guarantee the system will remain nearby. In fact, it is guaranteed that the system will evolve far away from an equilibrium given sufficient time.

## **EM algorithm**

To sample trajectories from the SDE, we use the classic Euler-Maruyama method. This relies on discretizing time and sampling from the noise distribution accounting for the magnitude of this time step. We implement the algorithm in our context with a nonnegativity constraint that is described in Section 3.6.5.

## Energy landscape

To understand which parts of state space the system is likely to be in given sufficient time to evolve from starting conditions, we sampled many trajectories and binned their locations at each update. We can thus compute the probability that the state is in any one given bin. For visualization purposes, we often marginalize over all but two state variables. Additionally, we take the negative logarithm of these probabilities to arrive at the energy landscape. The minima of this landscape correspond to the most likely states for the system to be in. Thus, they are similar to points of low potential energy from physics and hence the name.

## Transition paths

In the study of SDEs, we would like to know how a system might transition between stable equilibria, that is, answering the question: Which paths connecting two stable equilibria are most likely to be taken when the system transitions from one to the other? The appropriate framework in which to answer this question is that of large deviations theory, and that of Friedlin-Wentzell[41], in particular. The central object in this theory is an action functional that can be understood loosely to measure how much “work” the system has to do to move along a particular path. We would intuitively expect such a functional to be small for paths following the drift term in an SDE and large for paths that were moving against the drift term.

This action functional acts on the space of continuous functions from  $[0, T] \rightarrow \mathbb{R}^n$  where our state space can be embedded in  $\mathbb{R}^n$ . For our purposes, we are considering  $\mathbb{R}^3$ . Note, this is actually a family of functionals parameterized by the positive number  $T$ , the time at which

the path arrives at the endpoint. This action can be written in the following way[58]:

$$S_T(\psi) = \begin{cases} \int_0^T L(\psi, \dot{\psi}) dt & \text{if } \psi \in C(0, T) \text{ is absolutely continuous and the integral converges} \\ +\infty & \text{otherwise} \end{cases} \quad (1.1)$$

where the Lagrangian  $L(x, y)$  is given by

$$L(x, y) = \sup_{\theta \in \mathbb{R}^n} (\langle y, \theta \rangle - H(x, \theta)) \quad (1.2)$$

Here,  $\langle \cdot, \cdot \rangle$  is the standard Euclidean inner product in  $\mathbb{R}^n$  and  $H(x, \theta)$  is the Hamiltonian whose specific form depends on the dynamical system at hand. For our purposes, we are studying an SDE of the form

$$dX^\varepsilon(t) = b(X^\varepsilon(t)) dt + \sqrt{\varepsilon} \sigma(X^\varepsilon(t)) dW_t \quad (1.3)$$

And so the Hamiltonian in our context is given by

$$H(x, \theta) = \langle b(x), \theta \rangle + \frac{1}{2} \langle \theta, a(x) \theta \rangle \quad (1.4)$$

where  $a(x)$  is the diffusion tensor,  $\sigma \sigma^T$ . This allows for the simplification of the Lagrangian above to simply

$$L(x, y) = \langle y - b(x), a^{-1}(x)(y - b(x)) \rangle \quad (1.5)$$

Large deviations theory gives an estimate for the probability that the trajectory  $X^\varepsilon(t)$ ,

$t \in [0, T]$ ,  $T < \infty$ , of the SDE lies in a small neighborhood around a given path  $\psi \in C(0, T)$ . The estimate is given by

$$\mathbb{P}_x \left\{ \sup_{0 \leq t \leq T} |X^\varepsilon(t) - \psi(t)| \leq \delta \right\} \approx \exp(-\varepsilon^{-1} S_T(\psi)) \quad (1.6)$$

for  $\delta$  and  $\varepsilon$  sufficiently small and where  $\mathbb{P}_x$  denotes the probability conditional on  $X^\varepsilon(0) = x$  and where the path  $\psi$  is assumed to satisfy  $\psi(0) = x$ . This estimate can be made precise as done by others[41, 138, 146], however, it is sufficient for our purposes to consider the following result for any Borel subset,  $B \subset \mathbb{R}^n$ :

$$\mathbb{P}_x \{X^\varepsilon(T) \in B\} \asymp \exp\left(-\varepsilon^{-1} \inf_{\psi} S_T(\psi)\right) \quad (1.7)$$

where  $f(\varepsilon) \asymp g(\varepsilon)$  if and only if  $\log(f(\varepsilon))/\log(g(\varepsilon)) \rightarrow 1$  as  $\varepsilon \rightarrow 0$ , and the infimum is taken over all paths  $\psi$  satisfying  $\psi(0) = x$  and  $\psi(T) \in B$ .

In particular, we can take  $B = y$  where  $x$  and  $y$  are both stable equilibria of the deterministic ODE,  $dX/dt = b(X)$ . Then the minimizer,  $\psi$ , of the above equation is the path of maximum likelihood connecting two stable states in  $T$  time units, in other words, the most likely path by which the system will transition subject to the constraint that the transition happens within  $T$  time units.

One could argue that in our particular application of this theory, such a limitation imposed on us is acceptable. After all, humans are not living to indefinite ages so a reasonably large  $T$  would be sufficient to capture the most likely transition path a human cancer would take. In fact, minimizing over larger and larger  $T$  could very well lead to the minimizing path changing drastically from what we would find on the scale of human life. However, the reality is that our models will be incomplete pictures of the complicated biological processes underway. Not only will we by necessity leave out key state variables and their interactions,

but we will also misquantify the effects of noise and other perturbations to the system. Taken together, these could change the calculus determining how readily we would be likely to see a transition take place and we may find that what should only have become likely beyond a human's life expectancy is actually likely during it. Thus, the quasi-potential, with its ability to ignore time limitations, gives us the appropriate tool to study transitions in the cancer context.

**Quasi-Potential** With the functional  $S_T$  in hand, we can now turn our attention to the quasi-potential given by

$$V(x, y) = \inf_{T>0} \inf_{\psi \in \bar{C}_x^y(0, T)} S_T(\psi) \tag{1.8}$$

where  $\bar{C}_x^y(0, T)$  is the space of all absolutely continuous functions from  $[0, T] \rightarrow \mathbb{R}^n$  starting at  $x$  and ending at  $y$ . This definition allows us to eschew time-interval constraints on our functional and instead solely consider the task of moving from  $x$  to  $y$ . In fact, an alternative definition[41] can make this clearer:

$$V(x, y) = \lim_{T \rightarrow \infty} \lim_{\delta \rightarrow 0} \lim_{\varepsilon \rightarrow 0} (-\varepsilon \log \mathbb{P}_x \{ \tau_{\delta, y}(X^\varepsilon) \leq T \}) \tag{1.9}$$

where  $\tau_{\delta, y}(X^\varepsilon)$  is the first time at which  $X^\varepsilon$  enters the ball of radius  $\delta$  centered at  $y$ , or

$$\tau_{\delta, y}(X^\varepsilon) \equiv \inf \{ t > 0 | X^\varepsilon(t) \in B_\delta(y) \} \tag{1.10}$$

In the case we are considering of an underlying ODE with two stable equilibria, we can write down asymptotic expressions for the transition rates between these two equilibria in

the zero-noise limit:

$$k_{1,2} \asymp \exp(-\varepsilon^{-1}V(x, y)), \quad k_{2,1} \asymp \exp(-\varepsilon^{-1}V(y, x)) \quad (1.11)$$

So, with a solution to compute the quasi-potential, we can then estimate transition rates between our two stable equilibria. This can be useful for assessing how long we expect a given cancer to sit in a low equilibrium before transitioning to a higher, more aggressive state. That is, we can quantify how likely a tumor is to progress on a given timescale.

**Moving towards the gMAM** To begin solving this question, the first thing to do is reformulate the quasi-potential in purely geometric terms. That is, since we only care about the path through space (not through time), we want to express the quasi-potential without reference to a finite time interval.

In our particular context of an SDE, we can reformulate the equation for the quasi-potential as

$$V(x, y) = \frac{1}{2} \inf_{T>0} \inf_{\psi \in \tilde{C}_x^y(0, T)} \int_0^T \left| \dot{\psi}(t) - b(\psi(t)) \right|_{a(\psi)}^2 dt \quad (1.12)$$

where our norm  $|\cdot|_{a(\psi)}$  is the associated norm from the inner product  $\langle x, y \rangle_{a(\psi)} = x^T a^{-1}(\psi) y$ .

As done in[58], we can first reformulate this as

$$V(x, y) = 2 \inf_{T, \psi} \int_0^T |\dot{\psi}|_{a(\psi)} |b(\psi(t))|_{a(\psi)} \sin^2 \frac{1}{2} \eta(t) dt \quad (1.13)$$

where  $\eta$  is the angle between  $\dot{\psi}$  and  $b(\psi(t))$  in the metric induced by  $\langle \cdot, \cdot \rangle_{a(\psi)}$ . Now, this is in fact independent of the particular parameterization in time of  $\psi$  as seen by this equivalent

expression:

$$V(x, y) = 2 \inf_{\gamma} \int_0^1 |b(\gamma(s))|_{a(\gamma(s))} \sin^2 \frac{1}{2} \eta(s) ds \quad (1.14)$$

where  $\gamma$  is taken to be any curve satisfying  $\gamma(0) = x$  and  $\gamma(1) = y$ . Now, it remains to solve this minimization problem.

**gMAM for SDEs** In Heymann's original treatment of gMAM, they arrive at this equation through proving the following equivalent representation of the quasi-potential:

$$V(x, y) = \inf_{\varphi \in \tilde{C}_x^y(0,1)} \hat{S}(\varphi) \quad (1.15)$$

where  $\hat{S}(\varphi)$  is given by

$$\hat{S}(\varphi) = \int_0^1 \frac{L(\varphi, \lambda \dot{\varphi})}{\lambda} d\alpha, \quad \lambda = \lambda(\varphi, \dot{\varphi}) \quad (1.16)$$

and  $\lambda$  is defined implicitly by the following pair of equations:

$$H(\varphi, \hat{\theta}) = 0, \quad H_{\theta}(\varphi, \hat{\theta}) = \lambda \dot{\varphi}, \quad \lambda \geq 0 \quad (1.17)$$

A nice result of looking at this formulation is in understanding the meaning of  $\lambda(\varphi, \dot{\varphi})$ . It in fact gives the optimal speed for moving in the direction of  $\dot{\varphi}$  when starting at  $\varphi$ . That is,  $\lambda$  can be used to recover the time to traverse a path once we have the time-independent solution to our minimal action problem. Unfortunately,  $\lambda$  is known to vanish at zeros of the drift function and this leads to the fact that the time to arrive at or leave an equilibrium point is infinite. Thus, in our current context of transitioning between equilibria, we can at

most calculate the time of maximum likelihood for transitioning along a subset of our path.

Our numerical implementation of the gMAM is given in Section 3.6.6.

### 1.3.4 Parameter inference

To build models that comport with the real world as observed through experiments and the data generated, we can use the tools of parameter inference. The first step in such a process is to build some objective function that tells you how closely a model reproduces the observed data. The simplest next step is to optimize this function over the parameters of the model (or some subset thereof). While effective in its own right, this assumes that there is no possibility for biological variation between individuals or, in our case, tumor sites. Thus, we instead prefer to create a probability distribution over the space of chosen parameters that provides a fuller picture of the possibility space. This approach subsumes the first as we can understand that output as simply a Dirac delta function.

To achieve this goal, we turn to Bayesian statistics to infer parameter values. In particular, we make use of Markov chain Monte Carlo (MCMC) techniques to explore parameter space and find our parameter distribution.

#### Bayesian inference

Bayesian statistics is built off of Bayes' Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{1.18}$$

where the probability of an event  $A$  given  $B$  is expressed in terms of the reverse conditional and the absolute probabilities of the two events. If we let  $A$  be the event that the parameters

are a particular value—we will say  $\theta = \theta^*$ —and let  $B$  be the event that the model generates particular data—we will say  $y = y^*$ —then Bayes' Theorem looks like the following:

$$P(\theta = \theta^* | y = y^*) = \frac{P(y = y^* | \theta = \theta^*) \cdot P(\theta = \theta^*)}{P(y = y^*)} \quad (1.19)$$

Equation 1.19 expresses the main goal here: we can compute the probability of a particular set of parameter values ( $\theta^*$ ) by computing the probability that these parameters generate the observed data and the probability of these parameters being true. The denominator can be ignored because it simply scales all the probabilities equally and so it must be chosen so that the cumulative probability is 1. The left-hand-side quantity is called the posterior distribution.

The first factor,  $P(y = y^* | \theta = \theta^*)$ , is called the likelihood function and expresses the likelihood of the observed data being generated from the given set of parameters. This requires assigning not a point value for the output of the model, but rather a distribution in state space. In our setting, we will be looking at an ODE and a subset of its trajectory through state space corresponding to the time points of the observed data. At each of these points along the trajectory, we will assign a normal distribution centered at this point and assume these are all independent so we simply take the product of these.

The second factor,  $P(\theta = \theta^*)$ , is called the prior distribution. This is chosen based on whatever prior assumptions can be made about the parameters. Often the prior distribution is a uniform or a normal distribution. The more observations to compare against, the less the prior distribution influences the posterior distribution.

## Markov chain Monte Carlo

To actually go about computing the posterior distribution, we turn to numerical methods. There are many existing methods for doing this, most of which fall under the umbrella term Markov chain Monte Carlo (MCMC). The Markov chain aspect comes from the path the algorithm takes through parameter space. If the algorithm begins by visiting a given sequence of parameter vectors,  $(\theta^1, \theta^2, \dots, \theta^t)$ , then the next parameter vector,  $\theta^{t+1}$ , only depends on  $\theta^t$ . At each step, the algorithm updates the transition probabilities so that it converges towards the posterior distribution. The algorithm computes many such “chains” of parameter vectors each with randomized initial starting values and takes their composite as the final estimate for the posterior distribution. Often, a fixed number of iterations are dropped from the beginning of each chain.

The particular version of this algorithm used by the software package we used is called Hamiltonian Monte Carlo (HMC). This introduces a momentum component to the exploration of parameter space. This improves the efficiency of the algorithm and helps ensure full exploration of some otherwise difficult to reach regions of parameter space. See Gelman’s tome for further details[46].

# Chapter 2

## Modeling the competing effects of the immune system and EMT on epithelial cancers

This chapter is a pre-print. We build a mathematical model for understanding the effects of epithelial-to-mesenchymal transition (EMT) on invasive cancer. We also build a statistical framework and pipeline to assess how cancers may or may not show similarities to the predictions made by the model.

### 2.1 Summary

During progression from carcinoma *in situ* to an invasive tumor, the immune system is engaged in complex sets of interactions with various tumor cells. Tumor cell plasticity also alters disease trajectories via epithelial-to-mesenchymal transition (EMT). Several of the same pathways that regulate EMT are involved in tumor-immune interactions, yet little

is known about the mechanisms and consequences of crosstalk between these regulatory processes. Here we introduce a multiscale evolutionary model to describe tumor-immune-EMT interactions and their impact on epithelial cancer progression from *in situ* to invasive disease. Through *in silico* analyses of large patient cohorts, we find controllable regions that maximize invasion-free survival. We identify that delaying tumor progression depends crucially on properties of the mesenchymal tumor cell phenotype: its growth rate and its immune-evasiveness. Through analysis of EMT-inflammation-associated data from The Cancer Genome Atlas, we find that association with EMT significantly worsens invasion-free survival probabilities in support of our model, and we predict new genes influencing outcomes in bladder and uterine cancer, including FGF pathway members. These results offer novel means to delay disease progression by regulating properties of EMT through specific gene interactions, and demonstrate the importance of studying cancer-immune interactions in light of EMT.

## 2.2 Introduction

The majority of deaths from cancer are due to metastasis of the disease [32]. It is thus of critical importance to understand better the progression from *in situ* to invasive disease. Underlying this progression are genetic and epigenetic events, including mutations in pathways critical to the success of the cancer cell (driver mutations) [124]. These pathways include cell proliferation, apoptosis, and immunogenicity.

Cancer and the immune system interact in myriad ways. The immune system modulates the tumor microenvironment (TME), since immune signals that affect the tumor can be amplified or repressed through feedback in response to local inflammatory signals. This complex cell signaling occurs alongside the targeting (and potential eradication) of the tumor by immune cells [31].

The effects of the immune system on a tumor can be broadly summarized into two branches. The cytotoxic branch of the immune system, such as natural killer cells (NKs) and cytotoxic T cells (CTLs), seek out and lyse tumor cells. Upon carrying out their effector functions, these cytotoxic cells lose efficacy or deactivate [39]. The regulatory branch of the immune system (Tregs, and other factors), inhibits the effective functioning of the cytotoxic branch [123]. Inflammation can increase the probability of cancer incidence and progression, with some of the most pronounced effects seen for tumors originating in gastrointestinal and pancreatic tissues [6, 60]. Recent work has shown, contrary to the typical effects of inflammation on cancer, that under certain conditions inflammation may not be oncogenic but rather onco-protective [54].

Immunotherapies are beginning to realize their potential, with significant impacts on patient health and survival [108, 119], and may even provide a cure for certain hematopoietic cancers via anti-CD19 CAR-T cells [159]. The presentation of antigens on tumor cells is recognized by innate immune cells that are transported to lymph nodes where T cells (and other components) can be activated [129]. The tumor also engages in processes that can indirectly modify the TME, for example by releasing transforming growth factor beta ( $TGF-\beta$ ), which can shift the TME towards a tumor-supportive environment by enhancing immunosuppression via activation of Tregs [129].

Epithelial-to-mesenchymal transition (EMT) describes a reversible process by which cells displaying an epithelial phenotype transition into cells with a mesenchymal phenotype. Epithelial cells are – in part – defined by tight cell-cell adhesion. Mesenchymal cells exhibit less adhesion, greater ranges of motility, and may possess stem-like properties [102], although controversy regarding ‘stemness’ and EMT remains [101, 131]. Recent work has shown that – rather than being a binary process – at least two stable intermediate EMT states exist [59, 65]. Ongoing investigations into the plasticity and stability of EMT overlap with discussions elsewhere, e.g. of discrete vs. continuous processes during cell differentiation [96].

Intermediate states have emerged as a central mechanism by which cell fates (and the noise inherent within them) can be controlled [92, 117, 140].

Two features of the mesenchymal phenotype are of particular relevance in the context of cancer-immune interactions. i) mesenchymal tumor cells proliferate less than epithelial cells, we refer to this as mesenchymal growth arrest (MGA), and can be considered related to (in the sense of quiescence) the “stemness” phenotype of the mesenchymal tumor cells [158]. ii) mesenchymal cells are less susceptible to immune clearance [143]. As a cell is targeted by cytotoxic immune cells for clearance, a physical connection between the two cells must be established. This immunological synapse – mediated in part by T-cell receptors bound to antigens and the major histocompatibility complex on the target cell – is down-regulated in mesenchymal cells, thus inhibiting formation of the synapse [143]. We refer to this phenotype as mesenchymal immune evasion (MIE).

In addition to the prominent role it plays in metastasis, EMT has more recently been shown to also regulate other aspects of tumor progression [102, 110] and tumor dormancy [115]. *TGF- $\beta$* , a master regulator of EMT [84], is at once implicated heavily in tumor-mediated immune responses, since Tregs release *TGF- $\beta$*  upon arriving at the tumor site [143]. In hepatocellular carcinoma, for example, there is direct evidence linking Treg-secreted *TGF- $\beta$*  with EMT [136]. Thus, even by considering only the *TGF- $\beta$*  pathway, we find compelling evidence that these three core components (the tumor, the immune system, and EMT) all interact. It therefore strikes us as a priority to develop models to understand how the interactions between each of these three components affect cancer incidence and progression.

Mathematical oncology, that is, mathematical models of cancer incidence, progression, and treatment, has become a well-developed field; many models have offered insight into the cellular interactions underlying cancer and its interplay with the immune system, including older [4, 112, 135] and more recent works [3, 8, 13, 19, 42, 44, 50, 51, 67, 77, 91, 106, 130, 154]. These studies have increased our understanding of how tumors grow in the presence of various

immune components, and how treatment regimes can be designed to maximize the efficacy of cytotoxicity while minimizing risks to the patient. However, to our knowledge no models have addressed how the effects of EMT alter interactions between the immune system and cancer, and the subsequent implications for treatment.

Here we develop a model with the goal of studying interactions between the tumor, the immune system and EMT. We seek to describe a set of crucial molecular and cellular interactions in epithelial tumor cells, including effects due to DNA damage and mutation, to investigate the probability that *in situ* tumors will progress and, if so, when. A recent model of cancer-immune interactions [54] described the effects of the TME on the risk of cancer, and we build on the core cell cycle component of this model, adding significant new interactions to the immune component of the model (which was previously modeled by a single interaction), as well as adding the effects of EMT. In doing so, we shift the focus of the previous model from cancer initiation to cancer progression. We do this to reflect the fact that cancer progression hinges on escape from the immune system and the fact that EMT has a more well-defined role during progression and metastasis. We seek to understand whether this more complex immune module will change our understanding of inflammatory effects on the tumor, and how the epithelial-mesenchymal axis influences these.

In the next section we develop the model, explaining the intuition behind each of its components. We go on to analyze its behavior: global “*one-at-a-time*” sensitivity analysis identifies parameters that are crucial for progression. We study these in more depth, focusing on the competing effects of EMT and of the immune system on progression, and discover that EMT intricately regulates progression: under certain regimes a careful balance of EMT- and immune-driven processes can significantly prolong invasion-free survival. To test these predictions, we analyze data from the Cancer Genome Atlas (TCGA) using a new pipeline, and find strong evidence for the synergistic effects of inflammation and EMT, predicted through co-expression effects, for patients with bladder and uterine cancers.

## Quick Guide to Equations and Assumptions

To become invasive, An *in situ* tumor relies on mutations to alter cellular signaling pathways that enable cancer progression. The immune system simultaneously responds to the tumor upon recognition of neoantigens, and shapes the TME through dynamic inflammatory and regulatory signals.

To capture these dynamics, we developed a non-spatial agent-based model. Tumor cells are modeled individually as agents and immune cell populations are described homogeneously by differential equations. We consider two tumor cell types: epithelial tumor cells (ETCs) and mesenchymal tumor cells (MTCs). Time is treated discretely in 18 hour steps; approximating the time of one cell cycle. During each cell cycle, tumor cells can either undergo division, apoptosis, immune clearance, or arrest in  $G_0$ . The likelihood that a cell will proliferate depends on the tumor size (competition for resources) and on cell-intrinsic factors. The likelihood that a cell will undergo apoptosis is constant but varies between cell types (ETC and MTC). The likelihood of immune clearance depends on the number and type of mutated cells in the tumor, cytotoxicity, regulatory cells, and cell-intrinsic factors.

As tumor cells proliferate, DNA damage can occur, and over time they become increasingly likely to acquire pathway mutations that change their propensities for proliferation or cell death. Natural killer (NK) cells identify and clear tumor cells, a process which results in neoantigens priming and activating T cells in local lymph nodes. T cells can subsequently infiltrate the TME. At the tumor site, cytotoxic T cells (CTLs) lyse tumor cells, and T regulatory cells (Tregs) suppress cytotoxic activity. The following equation determines probability (per cell cycle) that a tumor cell will be lysed by a CTL:

$$\rho_{CTL} = \delta_{MUT} \frac{N_{CTL}}{N_C/K_1 + N_{CTL}} \frac{E_{CTL}}{1 + N_{Treg}/K_2} (1 - \delta_{IE}\Delta_{IE})(1 - \zeta\Delta_{MIE})$$

Here,  $\delta_{MUT}$  is 1 or 0 depending on whether or not the cell has a mutation. The second

term is a hill function modeled after [30] and the third term describes onco-protective effects of Tregs. The second-to-last term describes the increased immune-evasiveness that can occur following mutation in the immune evasion pathway, and the last term quantifies the additional immune evasiveness of associated with MTCs. The inflammatory state of the TME thus impacts (through multiple factors) immune recruitment and cytotoxicity at the tumor site.

EMT impacts tumor cells through their proliferation and potential to evade the immune system. MTCs have reduced proliferation and increased immune evasiveness.  $TGF-\beta$ , an activator of EMT, is produced both by Tregs and tumor cells, thus connecting tumor-immune interactions with EMT, and as a result plays an important role in shaping tumor outcomes. To determine whether a cell undergoes EMT or MET, the  $TGF-\beta$  is randomly divided among the cells so that associated with each cell  $i$  is a value  $\tau_i$ . This value is given by the following equation:

$$\tau_i = \frac{\tau_{\max}}{N_C} \frac{\tau/K_3}{1 + \tau/K_3} + X_i, \quad X_i \sim N(0, \sigma^2)$$

Here,  $\tau$  is the concentration of  $TGF-\beta$  in the TME and  $\tau_{\max}$  represents a limit on the amount of  $TGF-\beta$  that can be absorbed by all cells; a Gaussian noise term is added.

For each cell,  $\tau_i$  is summed with the current EMT score for that cell and if the result is above a threshold value, the cell undergoes EMT, otherwise MET. This summation expresses the assumption that the epithelial and mesenchymal phenotypes are stable and cells will move along the EMT spectrum towards the equilibrium they identify as under typical conditions.

Our measured outcome is the time at which the cancer becomes invasive, determined through the proportion of tumor cells harboring mutations in pathways that permit escape, relative to the total tumor cell population.

## 2.3 Methods

Here we briefly describe the core components of the model. Full details and equations are provided in the Supplementary Data. We develop an agent-based model to describe the relationships between cancer, the immune system, and EMT, building on the cell-cycle and tissue-cell components described in [54]. The agents in the model are the cells that have already formed an *in situ* tumor yet lack key pathway mutations to become invasive. In the process of the simulation, these cells can acquire mutations altering any of three key pathways (Fig. 2.1A).

We model immune cells as continuous variables, i.e. we assume that the tumor microenvironment is well-mixed with regards to the infiltrating immune cells. The cytokine  $TGF-\beta$  is also assumed to be well-mixed in the tumor microenvironment. Tumor cells can take on either epithelial or mesenchymal phenotypes in a plastic manner: these phenotypes depend on both the TME and cell-intrinsic factors. While the EMT score is continuous, a threshold determines if a given cell is labeled as epithelial or mesenchymal (Fig. 2.1).

### 2.3.1 Tumor evolution

Associated with each tumor cell are two essential features: their mutational signature and their EMT score. We consider three idealized pathways that can be mutated: proliferation, when altered this increases the probability of the cell proliferating within each cell cycle; apoptosis, when altered this decreases the probability of a cell undergoing apoptosis; and immune evasion, when altered this decreases the probability that a mutated cell will be cleared by immune components.

### 2.3.2 Immune population dynamics

The immune system is modeled by three immune cell types: NKs, CTLs, and Tregs. The NKs and CTLs act on the system by recognizing invasive cells and clearing them. Upon clearance, they are deactivated and removed from the immune population. Tregs suppress the function of NKs and CTLs (reduce tumor cell clearance), and in addition, release  $TGF-\beta$  which further shapes the TME by pushing tissues cells more towards a mesenchymal phenotype.

### 2.3.3 Periodic cycling inflammation states

Inflammation is modeled as a cycling scheme between low and high inflammatory states, with varying on/off durations and intensities. For the purpose of simulation we consider the default state to be low inflammation, and update the immune activity parameters whenever a switch to the high state occurs. In Table 2.3 we give full details of parameter settings during low and high inflammation.

### 2.3.4 Epithelial-to-mesenchymal transition

Each tissue cell has an EMT score between 0 and 1 with is set according to the concentration of  $TGF-\beta$  in the TME. Above a threshold, the cell acquires the phenotype of a mesenchymal tumor cell (MTC); otherwise, it is an epithelial tumor cell (ETC). For the purpose of simulation, ETCs are considered to be in the base state, and MTCs will have a subset of their parameters updated. In modeling EMT this way, we are assuming that the same factor,  $TGF-\beta$ , drives EMT both at initiation and through progression of cancer.

Cells that have undergone EMT (i.e. MTCs) experience a reduction in proliferation, referred

to as mesenchymal growth arrest (MGA), and a decrease in the likelihood that they will be cleared by immune cells (NKs or CTLs), referred to as mesenchymal immune evasion (MIE). Both these parameters lie within the range  $[0, 1]$ , thus we can sufficiently sample from their joint parameter space to explore it in depth with the need for informative priors to constrain their values.

### 2.3.5 Model simulation

**Initial conditions.** Simulations are initialized with  $N_0$  *in situ* tumor cells, determined by the choice of parameter values. A number of warmup cycles are run so that the model reaches steady state. During warmup, no mutations occur, and the only immune cells present are NKs. After warmup, mutations are permitted. Cells that do not mutate undergo an increase in their probability of mutation in a later cell cycle.

**Tumor cell fate.** During each cell cycle, the fate of each cell is assigned: proliferation, apoptosis, immune clearance, and rest in  $G_0$ , according to the model rules. The probability of proliferation is affected by mutations to the proliferation pathway (increased) and my cells in a mesenchymal state (decreased). Probabilities of immune clearance are affected by the number of mutations harbored: cells with more mutations are assumed to be more immunogenic and have a higher probability of being cleared by the immune system, unless the cell has a mutation in the immune evasion pathway. Cells in a mesenchymal state can exhibit greater capacity to evade immune clearance.

**Completing the cell cycle.** Once all tumor cells have been updated and fates chosen accordingly, non-tumor model components are updated. Immune cell populations are updated in two steps. First, immune cell exhaustion is calculated based on the number of tumor cells cleared, e.g. clearance of one tumor cell by an NK cell results in the NK cell population

decreasing by one. Second, all immune cells (NKs, CTLs, Tregs) are updated according to a system of coupled ordinary differential equations that govern their population dynamics. CTL and Treg recruitment rates are dependent on the number of tumor cells; in addition  $TGF-\beta$  enhances the recruitment rate of Tregs.

At the end of each cell cycle, new mutations can occur in cells that have undergone division, according to cell-specific probabilities that increase if no mutation occurs and are reset to 0 in the event of a mutation. Finally, the concentration of  $TGF-\beta$  and the EMT score for each cell are updated. Tregs and (to a lesser extent) invasive tumor cells are the sources of  $TGF-\beta$ ; the total concentration per cell cycle is divided randomly among tumor cells. EMT is then assessed, depending on the EMT score of the cell and the local concentration of  $TGF-\beta$ .

**Mutational burden and progression to invasive disease.** At the end of each cell cycle, the proportion of tumor cells that are invasive is calculated based on their mutational burden, and if it is above a certain threshold, the tumor is declared to have progressed to an invasive state and the simulation ends. The time to invasion is calculated as the time from the start of the simulation, minus the warmup period, until the invasive state is reached. Simulations run until either this occurs or until the maximum number of cell cycles has been reached.

### 2.3.6 Parameter estimation and sensitivity analysis

To study parameter sensitivity, we implemented Morris one-step-at-a-time global sensitivity analysis. Parameters are varied one at a time from a set of sampled “base” points and the resulting simulations recorded [97, 139]. For each run we simulated 1000 patients, and initialized the Morris sampling with 30 points in parameter space (at least 10 are recommended in [139]). Parameter sampling a choice of prior parameter distributions. For many

Name	Description
$p$	proliferation rate of tumor cells
$d_C$	death rate of tumor cells
$\Delta_{\text{MIE}}$	mesenchymal immune evasion
$\Delta_{\text{MGA}}$	mesenchymal growth arrest
$\Delta_A$	decrease in apoptosis rate in cells with apoptosis pathway mutation
$\Delta_{\text{IE}}$	increase in immune evasion in cells with immune pathway mutation
$\Delta_P$	increase in proliferation in cells with proliferation pathway mutation
$K_0$	EC50 for feedback of tumor cells on proliferation
$K_1$	EC50 for NK cells to identify a mutated cell
$K_2$	EC50 for Treg inhibition of cytotoxic functions
$K_3$	EC50 for relative internal concentration of $TGF-\beta$
$K_4$	EC50 for activation of Tregs by $TGF-\beta$
$E_{\text{NK}}$	rate at which NKs clearing mutated cells
$E_{\text{CTL}}$	rate at which CTLs clearing mutated cells
$\sigma_{\text{NK}}$	NK source rate
$\sigma_{\text{CTL}}$	CTL source rate per cleared mutated cell
$\sigma_{\text{Treg}}$	Treg source rate per cleared mutated cell
$d_{\text{NK}}$	NK death rate
$d_{\text{CTL}}$	CTL death rate
$d_{\text{Treg}}$	Treg death rate
$k_{\text{EMT}}$	EMT rate
$\sigma$	standard deviation of noise in $TGF-\beta$ input signal
$\tau_{\text{max}}$	max amount of $TGF-\beta$ input signal
$\tau_{\text{MUT}}$	rate of $TGF-\beta$ production by mutant cells
$\tau_{\text{Treg}}$	rate of $TGF-\beta$ production by Treg

Table 2.1: **Description of key model parameters.** Note that some are not constant as they can be affected by the inflammation state of the system.

parameters, such as for the immune population dynamics, measurements or estimates were available from literature [30]. For parameters such as MIE and MGA related to the mesenchymal phenotype, little prior information was available, thus these were sampled across all possible values in  $[0, 1]$ . Tumor size in the model was scaled from cell numbers on the order of  $10^9$  cells [30] to the order of  $10^2$ , and parameter values were scaled accordingly. Where parameter estimates existed, the prior for parameter  $\theta_i$  is given as  $\theta_i \sim N(m_e, 2m_e)$ , where  $m_e$  is the previous estimate and we take twice this value as the variance to obtain a range of samples that does not rely too heavily on previous work. The Morris algorithm computes the sensitivity,  $\mu^*$ , as the average of the absolute change of the output, which in our model is the area under the survival curve (Fig. 2.2).

### 2.3.7 Analysis of patient survival data from TCGA

We obtained primary tumor bulk mRNA sequencing and censored survival data for individuals monitored by cancer type from the Cancer Genome Atlas (TCGA) [142, 153], accessed through the Genomic Data Commons portal [53]. We developed methods to study: i) how the synergistic effects of EMT + inflammation compare to the effects of each of these individually; and ii) the importance of mesenchymal proliferation rates in determining cancer prognosis (Fig. 2.11), which allow us to test predictions from the agent-based model. See Section 2.6.4 for more information on my collaborator's methods.

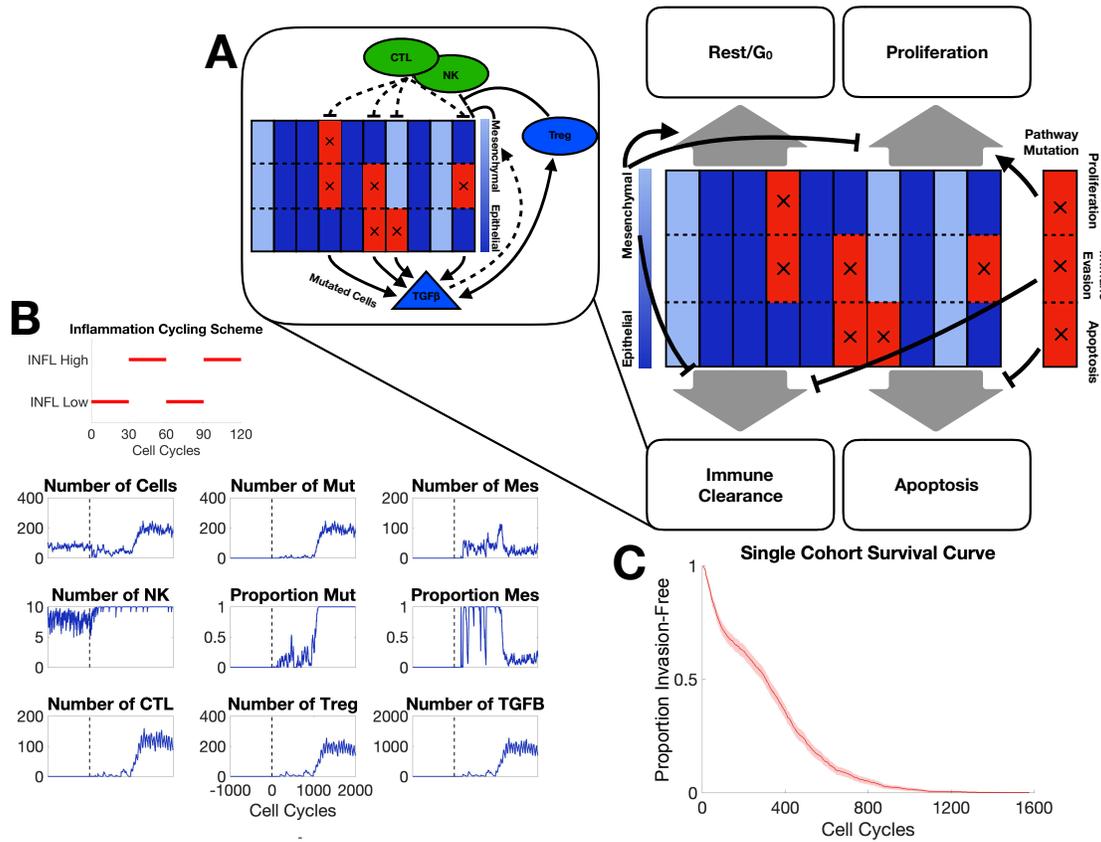


Figure 2.1: **Model cartoon, sample trajectory, and sample survival curve.** A) Schematic depiction of agent-based model components; each of the 10 columns represents one tumor cell divided into three compartments representing the state of three pathways with tumorigenic potential; blue/red denote baseline/altered pathway activity. Black arrows depict cell fate regulation in each cell cycle. Inset depicts major interactions between the immune system and tumor cells. B) A representative simulation of one patient. The parameter values used can be found in Table 2.3. The inflammation cycling scheme (red) is shown above the patient dynamics. The vertical dashed line denotes the end of the warmup period. Mut: malignant cells; Mes: mesenchymal cells. C) Survival curve for one cohort of patients for parameter values given in Table 2.3.

## 2.4 Results

### 2.4.1 A multiscale agent-based model of EMT-immune-tumor cell interactions to study tumor progression

We begin by investigating general features of the model to establish baseline conditions and assess the impact of different model components on the key measured outcomes: the probability of progression, and the time to invasion. During the cell cycle, cell fate is determined by rules that are influenced by EMT and immune interactions (Fig. 2.1A), e.g. if a cell undergoes EMT, its probability of proliferation is reduced; if it gains a mutation in the apoptosis pathway, its probability of apoptosis is reduced. Meanwhile, NK cells and CTLs attempt to clear malignant tumor cells, and deactivate upon successful tumor cell clearance; Tregs inhibit this cytotoxic activity (Fig. 2.1A Inset).

The inflammation cycling scheme for a typical *in silico* patient consists of alternating high and low regimes with corresponding effects on the cell populations (Fig. 2.1B). For this patient, after warmup, mutations are observed at a rate low enough that they are cleared by cytotoxic cells for about 700 cell cycles, after which the mutated and thus invasive cell population begins to grow, leading to large recruitment of CTLs and Tregs and a peak in the concentration of  $TGF-\beta$ . After 841 cell cycles, the proportion of invasive cells reaches 50%: the threshold defining progression, thus this patient has a time to invasion of 841 cell cycles, or 631 days. Beyond this timepoint, we see a rapid increase in the number of invasive cells until it comprises 100% of the tumor population. Interesting EMT dynamics are also observed, the proportion of MTCs peaks shortly after the tumor becomes invasive, subsequently the majority of cells transition back to an epithelial state. We observe that while the NK population varies little over the simulation, CTLs and Tregs both undergo large expansions. CTLs and Tregs also appear to oscillate, however note that this is a direct

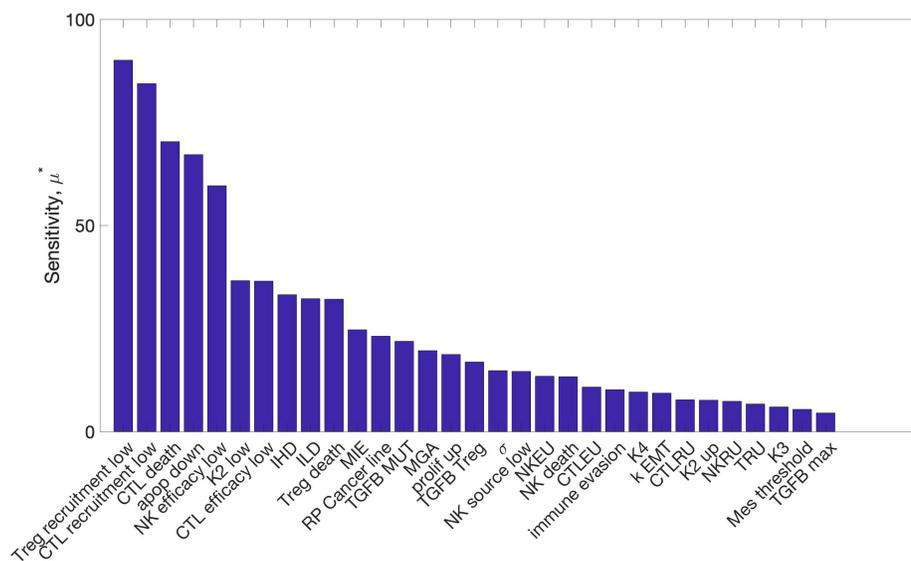


Figure 2.2: **Global sensitivity analysis of model parameters.** The sensitivity ( $\mu^*$ ) denotes the average absolute change in the time to invasion over the range of variation of the parameter.

result of the inflammation state, and is not immune cell-intrinsic.

In order to quantify patient dynamics and invasion-free survival as a population level, we simulate large cohorts of patients similar to the single patient shown in Fig. 2.1B. For a cohort of 500 patients, we simulate survival curves and see that a large number progress quickly to form invasive tumors, whereas a few lie in the tail of the distribution after the mutagenic event that a large number of tumors quickly progress while others takes some time before progressing Fig. 2.1C. By approximately 1200 cell cycles (2.5 years), all tumors have become invasive..

## 2.4.2 Identification of key model parameters via global sensitivity analysis

Exploring the parameter spaces of systems biology models *adequately* is – in general – a hard problem. Fitting parameters via (Bayesian) parameter inference is advisable wherever

possible [68]. Here, despite a wealth of data on tumor growth dynamics, a lack of sufficient molecular measurements (i.e. immune cell dynamics) precludes inference of the full model. In addition, while inference schemes for agent-based models are developing [43, 152], simulation times remain a hurdle [76]. Parameters for some components of the model studied previously can be constrained [54]. However, even here, new biological processes in the current system could push the model into new behavioral regimes. Thus to sample and characterize the parameter space of the model we use sensitivity analysis.

The results of Morris one-step-at-a-time sensitivity analysis on the 31 model parameters (Fig. 2.2) find a subset of parameters with much higher levels of sensitivity than others. The two most influential by this analysis are the recruitment rates of Tregs and CTLs in the low inflammation state. The parameters influencing EMT are also identified as influencing model outcomes. Since one goal of our analysis is to assess the specific effects of EMT on immune-cancer dynamics, parameters MIE and MGA are of particular interest. In addition, inflammation parameters controlling the periodic high/low inflammation states are of interest because they strongly influence model outcomes and are capable of being targeted by therapeutic treatments. For immune cell dynamics, the secretion of  $TGF-\beta$  by Tregs is found to be sensitive and thus will also be studied further below.

### **2.4.3 Mesenchymal properties dramatically alter invasion-free survival times**

Mesenchymal tumor cells (MTCs) are characterized by changes in two parameters: mesenchymal immune evasion (MIE) and mesenchymal growth arrest (MGA). Here we assess the effects of each, alongside the effects of  $TGF-\beta$  through its production by Tregs. As MIE increases, the invasion-free survival decreases (Fig. 2.3A) for all sets of parameters studied: as the subpopulation of invasive cells becomes more resistant to immune clearance, the tumor

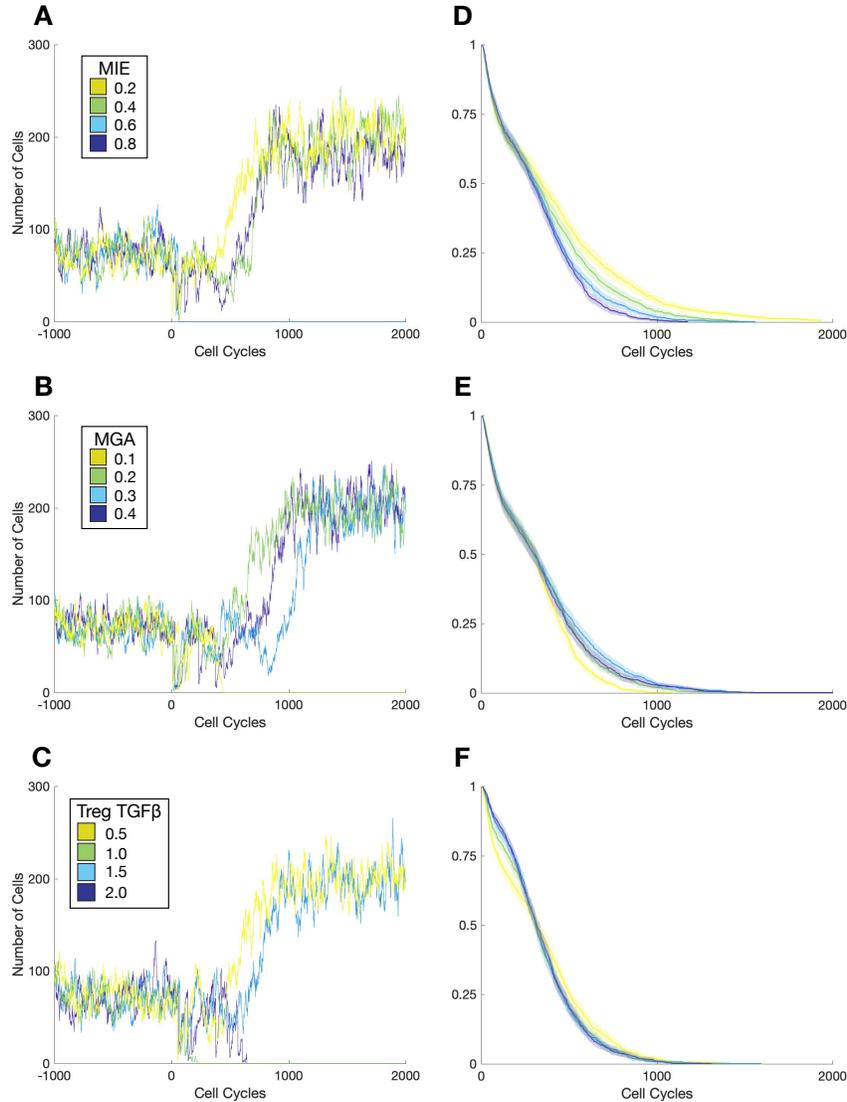


Figure 2.3: **Effects of mesenchymal tumor cell properties on the time to invasion.** Trajectories of one patient per cohort including warmup and 2000 cell cycles for (A) mesenchymal immune evasion (MIE); (B) mesenchymal growth arrest (MGA); (C) Production of  $TGF-\beta$  by Tregs. (D) Survival curve corresponding to changes in the parameter MIE (A) for a patient cohort of 1000. Shaded region represents the 95% confidence interval over the cohort. (E) Survival curve corresponding to changes in MGA. (F) Survival curve corresponding to changes in Treg production of  $TGF-\beta$ .

as a whole grows more resilient and thus can grow faster (Fig. 2.3D).

The relationship between MGA and invasion-free survival times displays a very different trend, and is non-monotonic with a local maximal value. For small values of MGA, increasing MGA results in increasing the invasion-free survival (Fig. 2.3B, E). However for large values of MGA, invasion-free survival times decrease. This is explored further below.

$TGF-\beta$  varies according to its production by tumor cells and its production by Tregs. We assess the effects of varying the production of  $TGF-\beta$  by Tregs on invasion-free survival (Fig. 2.3C, F), and find that at lower production rates of  $TGF-\beta$ , the survival curve initially declines faster whereas higher production rates result in a steeper drop off in survival later. While lower values of  $TGF-\beta$  production lead to a steeper initial decline, these differences vanish for higher values of  $TGF-\beta$ . The steeper initial decline may be due to the rapid clearance of tumor cells by adaptive immune cells before Tregs have had sufficient time to modulate the TME (which they do through secretion of  $TGF-\beta$ ).

#### **2.4.4 A key EMT regime maximizes cancer-free survival time under chronic inflammation**

To investigate how competing interactions within the inflammatory tumor microenvironment affect EMT, we explored the effects of varying inflammation on invasion-free survival. Patient cohorts were simulated under different inflammation regimes: permanently low inflammation; permanently high inflammation; or variable (periodic high/low) inflammation. Compared to the other inflammation states, permanently high inflammation results in outcomes that vary more subtly with changes in the mesenchymal parameters (Fig. 2.4). When the inflammation state is either permanently or temporarily low, surprising trends emerge. In both these cases, invasion-free survival time is negatively correlated with MIE, and a local maximum for the invasion-free survival time is found with respect to MGA (close to  $\Delta_{MGA} = 0.2$ ).

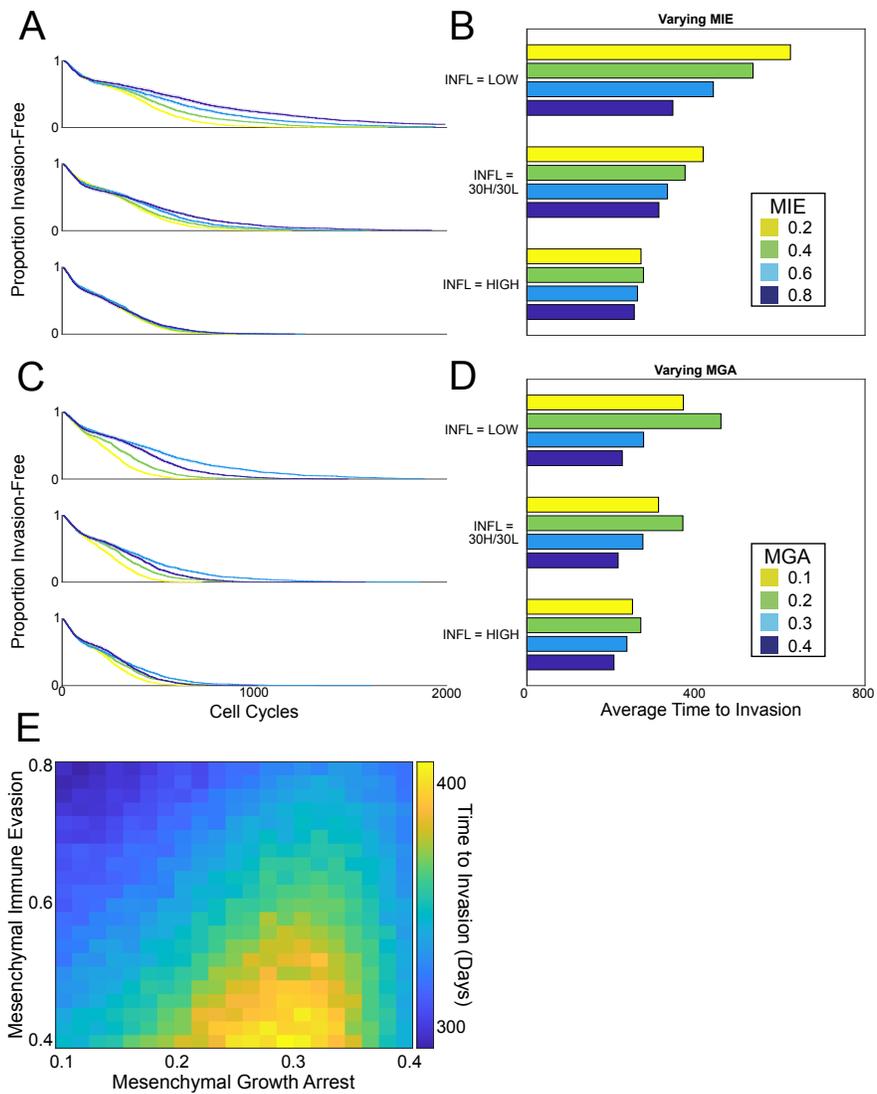


Figure 2.4: **Effects of inflammation on the time to invasion under different cycling schemes.** A-B) As MIE varies, survival curves (each of 200 patients) and corresponding bar plots to summarize the mean Time to Invasion for each cohort are shown. C-D) As MGA varies, survival curves and corresponding bar plots to summarize the mean Time to Invasion for each cohort are shown. E) Summary of the effects of MIE and MGA on invasion-free survival.

These differences in the mean invasion-free survival lead to striking variation in outcomes: tumors can be contained *in situ* for up to twice as long as they would otherwise be simply by varying the rates of mesenchymal growth arrest. These predictions point to intriguing therapeutic outcomes: a patient suffering intermittent high inflammatory attacks will benefit directly from EMT-directed therapies, however patients for whom a relatively high inflammation state is observed chronically will not obtain this benefit.

When MIE is varied under different inflammation cycling schemes, for all the periodic inflammation schemes studied, increasing MIE will decrease the invasion-free survival (i.e. worsen cancer progression and prognosis) (Fig. 2.4B). In the case of continuously high inflammation, the effects of MIE are minimal. Thus, under any inflammation regime with periods of low inflammation, as we might intuitively assume, any reduction in mesenchymal immune evasion will lead to improvements in patient outcomes. To summarize the mesenchymal properties of immune evasion and growth arrest, we plot the joint density of these parameters against the time to invasion (Fig. 2.4E): where we see that for a given value of mesenchymal immune evasion, there is a value of mesenchymal growth arrest that maximizes the time to invasion.

#### **2.4.5 TCGA data analysis supports predictions highlighting importance of mesenchymal cell proliferation in determining outcome**

Given the model prediction that mesenchymal proliferation rates exert essential control on invasion-free survival times (Fig. 2.4E), we performed analysis of 14 cancer types from TCGA to assess the importance of mesenchymal proliferation-associated genes using clinical outcomes. We assessed the effects of mesenchymal proliferation in cancer gene expression data (Fig. 2.11). Using both overall survival and disease-free interval as endpoints, we identified three tumor types where EMT-inflammation associated genes predicted clear differences

between patient groups. See Section 2.6.7 for more of my collaborator’s results.

## 2.5 Discussion

Despite the importance of studying interactions between cancer and the immune system, as well as studying the effects of EMT on cancer, there has not previously, to the best of our knowledge, been a model developed that combines all three of these components. Here we studied cancer, the immune system, and EMT, during the progression from an *in situ* tumor to invasive disease. We saw this as a particularly pressing need given the shared factors influencing all these components, including  $TGF-\beta$  and Wnt signaling. We used an individual cell-based model framework to describe the multiscale processes leading to invasive disease, and we compared model predictions with a novel TCGA analysis framework to predict the effects mesenchymal phenotype-associated genes.

We found that the model recapitulated invasion-free survival dynamics. Using global parameter sensitivity analysis, we identified parameters exerting key control over model behavior. Focusing on these led us to identify that increasing mesenchymal immune evasion and increasing Treg  $TGF-\beta$  production both lead to shorter invasion-free survival times. However, varying the level of inflammation led to paradoxical effects with regards to mesenchymal growth arrest: under regimes with periods of low inflammation, an optimal level of mesenchymal growth arrest can improve outcomes and maximize the invasion-free survival. To capture the essential characteristics of the model, we summarized *in silico* patient studies with a single parameter: the invasion-free survival time. There are, of course, many trajectories that result in progression to invasion. Further analysis of the transient cell dynamics in tumors *in situ* and during progression is needed to gain greater insight into the EMT-associated dynamics of cancer.

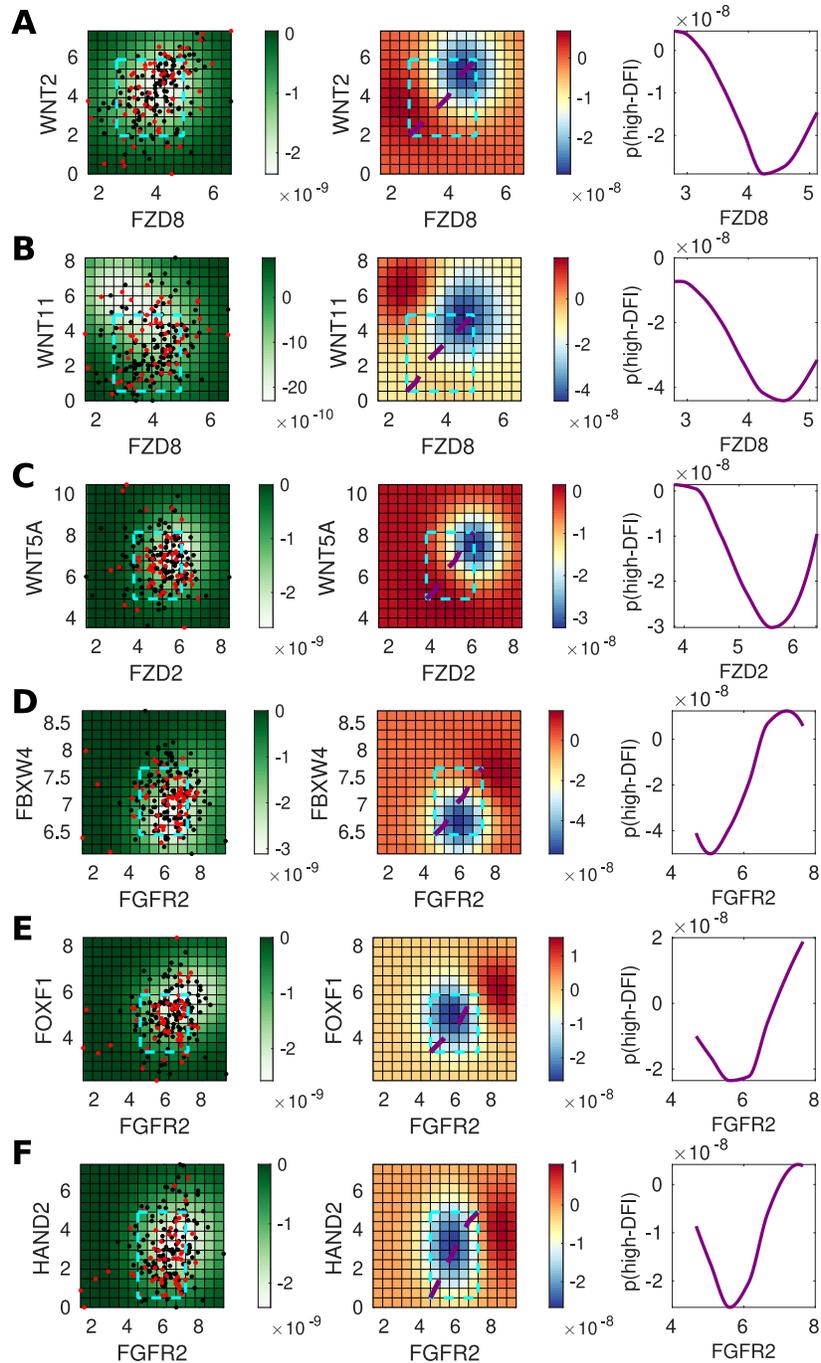


Figure 2.5: **Genes predictive of invasiveness in BLCA.** A) For gene pair *WNT2* and *FZD8*, the left panel shows the posterior variance on log-log expression plot of the predicted probability overlaid with patient samples (red = low-DFI, black = high-DFI), 90% confidence interval box drawn for standardized expression values (cyan); middle panel: posterior log probability of high-DFI over the same region as left, where the diagonal line (purple) shows the co-expression trend (diagonal line through the 90% CI of standardized expression values); right panel: posterior log probability of high-DFI plotted against expression of *FZD8*, values simulated along the diagonal (purple) corresponding to the middle panel. B) As above for *WNT11* and *FZD8*. C) As above for *WNT5A* and *FZD2*. D) As above for *FBXW4* and *FGFR2*. E) As above for *FOXF1* and *FGFR2*. F) As above for *HAND2* and *FGFR2*.

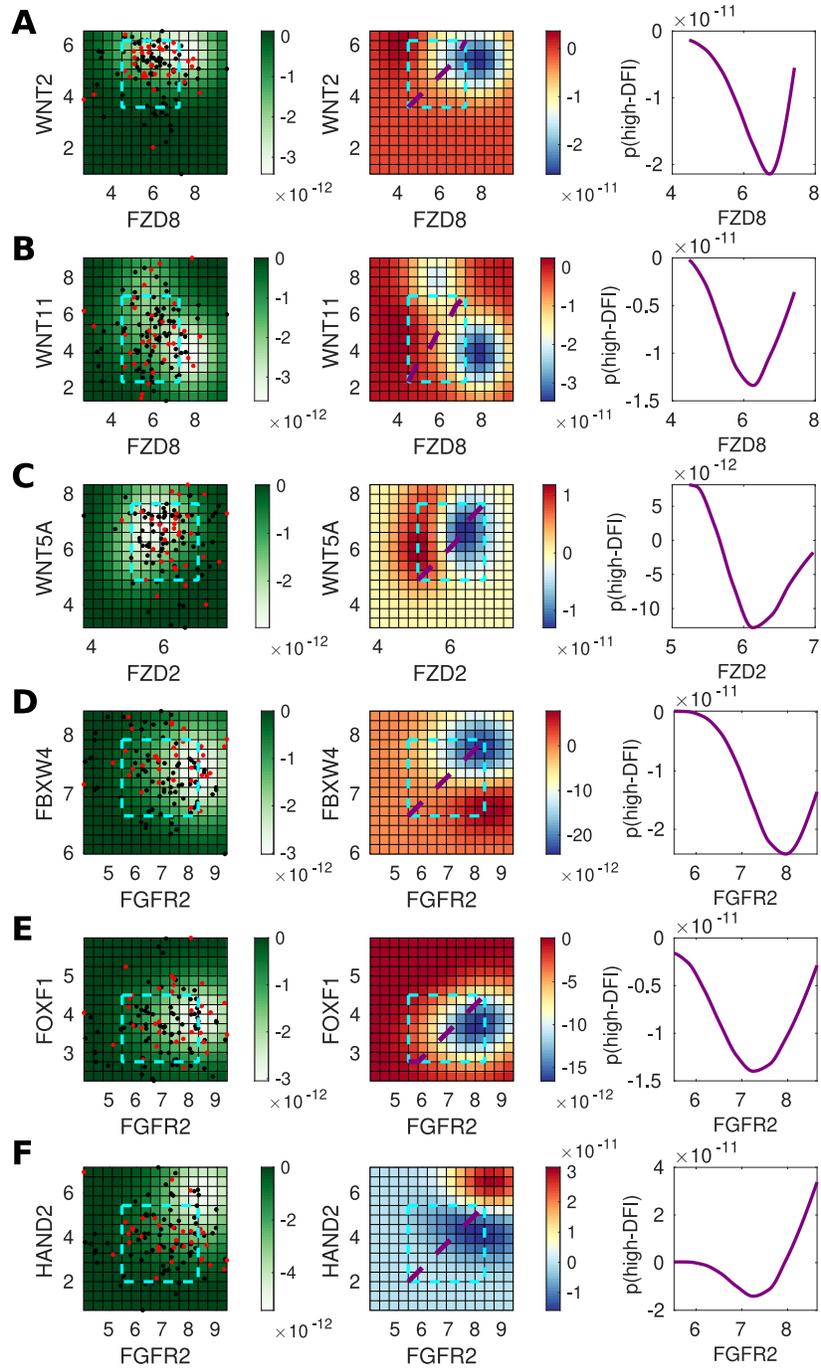


Figure 2.6: **Genes predictive of invasiveness in UCEC.** A) For gene pair *WNT2* and *FZD8*, the left panel shows the posterior variance on log-log expression plot of the predicted probability overlaid with patient samples (red = low-DFI, black = high-DFI), 90% confidence interval box drawn for standardized expression values (cyan); middle panel: posterior log probability of high-DFI over the same region as left, where the diagonal line (purple) shows the co-expression trend (diagonal line through the 90% CI of standardized expression values); right panel: posterior log probability of high-DFI plotted against expression of *FZD8*, values simulated along the diagonal (purple) corresponding to the middle panel. B) As above for *WNT11* and *FZD8*. C) As above for *WNT5A* and *FZD2*. D) As above for *FBXW4* and *FGFR2*. E) As above for *FOXF1* and *FGFR2*. F) As above for *HAND2* and *FGFR2*.

We tested the model prediction that mesenchymal phenotypes play key roles in tumor invasiveness through a novel TCGA data analysis framework. In support of this prediction, we found that mesenchymal-associated genes controlled outcomes and predicted differences between high vs. low disease-free interval patient groups. This analysis yielded predictions of the effects of single genes or gene pairs, many of which corresponded to known effects, including the effects of both canonical and non-canonical Wnt signaling on tumor progression. Our modeling also predicted opposing roles for FGF signaling in bladder and uterine cancers: where *FGFR2* exerts a tumor-suppressor effect in bladder cancer yet a tumorigenic effect in uterine cancer. Evidence for these opposing roles already exists in the literature, but notably, through our modeling, we also predict entirely novel interactions between *FGFR2* and other transcription factors (*FBXW4*, *FOXF1*, and *HAND2*) that act to enhance or suppress the effects of *FGFR2* alone, and could offer significant novel therapeutic strategies.

In future work, further development of the inflammation module is important given the large and at times paradoxical roles that the inflammatory state exerts on tumor cells and invasion-free survival. Currently, inflammation is modeled as independently cycling between high and low schemes, yet several known factors contribute to the inflammatory state. For example, model extensions could assume that the level of inflammation depends on the number of and the degree of mutations that tumor cells harbor. The competing effects that *TGF- $\beta$*  exerts on the tumor and its microenvironment also warrant further investigation. We found that – below a certain threshold – reduction of *TGF- $\beta$*  increases the time to invasion, i.e. reducing *TGF- $\beta$*  in the TME benefits survival. Experimental work in support of this result includes a study of *TGF- $\beta$*  tumor suppression in pancreatic cancer through the promotion of EMT [27]. The *TGF- $\beta$*  pathway is however implicated in numerous other cellular signaling processes besides EMT; changing *TGF- $\beta$*  concentration even in a local environment could have large off-target effects. Indeed, it has been shown that *TGF- $\beta$*  promotes invasion and heterogeneity while suppressing cell proliferation in squamous cell carcinoma [105]. To account for this complex signaling, future work should incorporate the effects of signaling factors downstream

of  $TGF-\beta$  on the cancer dynamics. It is also important to note that in this model, EMT is initiated entirely by  $TGF-\beta$ . While  $TGF-\beta$  does play a large role in cancer EMT, it is by no means the only factor at play; in reality, cells must contend with and respond to a milieu of EMT-associated signals.

Tumor heterogeneity often helps the tumor to evade immune effects and complicates our approaches to treatment. Rigorous study of the consequences of the increased heterogeneity that follows disease incidence (i.e. decanalization [47]) is too-often sidelined, despite mounting evidence in support of its prominent role in cancer evolution [23, 24, 116]. Despite these challenges, great progress in predicting disease complexity continues to be made. As we are rapidly approaching a new generation of immunotherapies, it is these very complexities that we must better understand in order to control or eradicate the disease.

## 2.6 Supplementary Information

### 2.6.1 Model Description

#### Tissue cell fate

During each cell cycle, every cell randomly is assigned a cell fate from the following options:

- proliferation
- apoptosis
- immune clearance (by NKs or CTLs)
- rest in  $G_0$

For each cell, a weight is chosen for each option and these are normalized to probabilities which then are used to randomly determine what each cell does during the cell cycle.

**Proliferation** There are four factors that contribute to the weight of a cell to proliferate. The first is a base proliferation rate that all cells have,  $p$ . Second, if the cell has a mutation in the proliferation pathway ( $\delta_P = 1$ ), then the weight for proliferation is proportionally increased by  $\Delta_P$ . Third, if the cell is mesenchymal ( $\zeta = 1$ ), then the weight for proliferation is proportionally decreased by  $\Delta_{\text{MGA}}$ , which stands for mesenchymal growth arrest. This lost proliferation for mesenchymal cells will later be used to increase their chance of resting. Fourth, there is a negative feedback of the cells on their own proliferation which is quantified by a Hill factor as a function of the tissue cell population,  $N_C$ , with EC50 term  $K_0$ . In total, the weight for proliferation is given by

$$\rho_P = p(1 + \delta_P \Delta_P)(1 - \zeta \Delta_{\text{MGA}}) \frac{K_0}{K_0 + N_C} \quad (2.1)$$

**Apoptosis** There are two factors that contribute to a cell's weight for undergoing apoptosis. There is a basal apoptosis rate that all cells experience,  $d_C$  for death. Second, if the cell has a mutation in the apoptosis pathway ( $\delta_A = 1$ ), then the weight for undergoing apoptosis is proportionally decreased by  $\Delta_A$ . In total, the weight for apoptosis is given by

$$\rho_A = d_C(1 - \delta_A \Delta_A) \quad (2.2)$$

**Immune Clearance** For both NK clearance and CTL clearance, the weights are built with the same factors but have different parameter values for NK and CTLs. First of all,

the cell needs to be malignant ( $\delta_{\text{MUT}} = 1$ ). Second, there is a Hill factor that captures the probability of an immune cell finding and interacting with the given tissue cell with EC50 term  $K_1$ . Third, NKs and CTLs have their own efficacy parameters,  $E_{\text{NK}}$  and  $E_{\text{CTL}}$ , which can be understood as the rate of immune clearance given an immune cell has found the mutated cell. Fourth, there is a decreasing Hill factor based on the number of Treg cells present with EC50 term  $K_2$ . Finally, there are two factors that proportionally decrease the weight of immune clearance depending on if the cell has an immune evasion mutation ( $\delta_{\text{IE}} = 1$ ) or if it is mesenchymal ( $\zeta = 1$ ) with respective decreases  $\Delta_{\text{IE}}$  and  $\Delta_{\text{MIE}}$ . In total, the weight of NK clearance is given by

$$\rho_{\text{NK}} = \delta_{\text{MUT}} \frac{N_{\text{NK}}}{N_C/K_1 + N_{\text{NK}}} \frac{E_{\text{NK}}}{1 + N_{\text{Treg}}/K_2} (1 - \delta_{\text{IE}}\Delta_{\text{IE}})(1 - \zeta\Delta_{\text{MIE}}) \quad (2.3)$$

A similar formula holds for CTLs with only the number of CTLs and their efficacy being different from the above equation.

**Rest in  $G_0$**  The weight associated with rest is taken as 1 except in the case of mesenchymal cells. Recall that mesenchymal cells had their proliferation rate decreased by  $1 - \zeta\Delta_{\text{MGA}}$  (see Eq. 2.1). The biological assumption here is that mesenchymal cells instead of proliferating will instead rest, so this lost proliferation weight is added to the resting weight. Hence, the weight of rest is given by

$$\rho_R = 1 + \zeta p(1 + \delta_P\Delta_P)\Delta_{\text{MGA}} \frac{K_0}{K_0 + N_C} \quad (2.4)$$

Again, the reason for adding that term is due to the understanding that overall mesenchymal

cells proliferate less as individual cells rest longer in the  $G_0$  phase.

**Completing the Cell Cycle** After the cell fates are determined and the results reflected in the system, there are a few things that happen before the system moves on to a new cell cycle. First, the NK and CTL populations are reduced by the number of mutated cells they cleared. This represents the fact that individual immune cells lose efficacy as they carry out their effector functions. Second, all proliferating cells have a cell-specific probability of undergoing a driver mutation in one of the three pathways. If they do, one is randomly chosen among the three pathways and the pathway in that cell becomes altered. If the cell does not undergo a mutation, then its probability of mutation during subsequent cell cycles increases.

Finally, the EMT values for each cell is updated. This depends on the cells current EMT score and how much  $TGF\text{-}\beta$  is currently in the system. The amount of  $TGF\text{-}\beta$  absorbed by all cells is given by an increasing Hill function in terms of the  $TGF\text{-}\beta$  in the TME. The saturation effect is to limit the amount of  $TGF\text{-}\beta$  a cell can absorb in a given time interval. This quantity is then divided up randomly among the  $N_C$  living cells via a normally distributed noise term to determine how much exogenous  $TGF\text{-}\beta$  each cell receives during this cell cycle. Should this value,  $\tau_i$  in Eq. 2.5, be negative, we interpret this as the cell losing  $TGF\text{-}\beta$  to the TME and thus being more likely to undergo MET.

$$\tau_i = \frac{\tau_{\max}}{N_C} \frac{\tau/K_3}{1 + \tau/K_3} + X_i, \quad X_i \sim N(0, \sigma^2) \quad (2.5)$$

We then combine  $\tau_i$  with the current EMT score of the cell, as a proxy for the endogenous  $TGF\text{-}\beta$ . Finally, if this quantity is large enough, the EMT score of the cell increases towards 1; otherwise, it decreases towards 0. Each cell then is relabeled as either epithelial

or mesenchymal depending on its new EMT score and whether it is below or above the mesenchymal threshold. Thus, there are two main factors that determine if a cell will end a cell cycle as mesenchymal: concentration of  $TGF-\beta$  in the system and the current EMT score of the cell.

Next, the amount of  $TGF-\beta$  for the next cell cycle is determined by the number of mutated cells,  $N_{\text{MUT}}$ , and the number of Treg cells,  $N_{\text{Treg}}$ , each one producing a fixed amount of  $TGF-\beta$ . It is given by

$$\tau = \tau_{\text{MUT}}N_{\text{MUT}} + \tau_{\text{Treg}}N_{\text{Treg}} \quad (2.6)$$

Finally, the immune populations are updated. For the NKs, they obey the following differential equation:

$$N'_{\text{NK}} = \sigma_{\text{NK}} - d_{\text{NK}}N_{\text{NK}} \quad (2.7)$$

which is discretized to

$$N_{\text{NK}}(k+1) = \left( N_{\text{NK}}(k) - \frac{\sigma_{\text{NK}}}{d_{\text{NK}}} \right) \exp(-d_{\text{NK}}\Delta t) + \frac{\sigma_{\text{NK}}}{d_{\text{NK}}} \quad (2.8)$$

For CTLs and Tregs, they rely on malignant cells being cleared before they can be activated. Let  $N_{\text{MUT}}^*(k)$  represent the number of malignant cells cleared by the immune system during cell cycle  $k$ . In addition, Treg recruitment is upregulated by  $TGF-\beta$ , which will be incorpo-

rated via a Hill function with EC50 term  $K_4$ . We choose the following differential equations to govern the CTL and Treg populations:

$$\begin{aligned}
 N'_{\text{CTL}} &= \sigma_{\text{CTL}} N_{\text{MUT}}^* - d_{\text{CTL}} N_{\text{CTL}} \\
 N'_{\text{Treg}} &= \sigma_{\text{Treg}} N_{\text{MUT}}^* \frac{\tau}{1 + \tau/K_4} - d_{\text{Treg}} N_{\text{Treg}}
 \end{aligned}
 \tag{2.9}$$

Discretized, these are:

$$\begin{aligned}
 N_{\text{CTL}}(k+1) &= (N_{\text{CTL}}(k) - \sigma_{\text{CTL}} N_{\text{MUT}}^*(k)/d_{\text{CTL}}) \exp(-d_{\text{CTL}} \Delta t) + \sigma_{\text{CTL}} N_{\text{MUT}}^*(k)/d_{\text{CTL}} \\
 N_{\text{Treg}}(k+1) &= \left( N_{\text{Treg}}(k) - \frac{\sigma_{\text{Treg}} N_{\text{MUT}}^*(k)}{d_{\text{Treg}}} \frac{\tau(k)}{1 + \tau(k)/K_4} \right) \exp(-d_{\text{Treg}} \Delta t) \\
 &\quad + \frac{\sigma_{\text{Treg}} N_{\text{MUT}}^*(k)}{d_{\text{Treg}}} \frac{\tau(k)}{1 + \tau(k)/K_4}
 \end{aligned}
 \tag{2.10}$$

## 2.6.2 Definition of parameters specifying the model

Name	Description
$p$	proliferation rate of tissue cells
$d_C$	death rate of tissue cells
$\Delta_{\text{MIE}}$	mesenchymal immune evasion
$\Delta_{\text{MGA}}$	mesenchymal growth arrest
$\Delta_A$	mutant cells decreased apoptosis
$\Delta_{\text{IE}}$	mutant cells increased immune evasion
$\Delta_P$	mutant cells increased proliferation
$K_0$	EC50 term for negative feedback of tissue cells on own proliferation
$K_1$	EC50 term for probability of NK cell finding mutant cell
$K_2$	EC50 term for Treg inhibition of cytotoxic functions
$K_3$	EC50 term for how much $TGF\text{-}\beta$ each cell has
$K_4$	EC50 term for $TGF\text{-}\beta$ activation of Tregs
$E_{\text{NK}}$	rate of NKs clearing mutants
$E_{\text{CTL}}$	rate of CTLs clearing mutants
$\sigma_{\text{NK}}$	NK source rate
$\sigma_{\text{CTL}}$	CTL source rate per cleared mutant cell
$\sigma_{\text{Treg}}$	Treg source rate per cleared mutant cell
$d_{\text{NK}}$	NK death rate
$d_{\text{CTL}}$	CTL death rate
$d_{\text{Treg}}$	Treg death rate
$k_{\text{EMT}}$	EMT/MET rate
$\sigma$	standard deviation of noise in $TGF\text{-}\beta$ each cell receives
$\tau_{\text{max}}$	max amount of $TGF\text{-}\beta$ any cell can receive
$\tau_{\text{MUT}}$	rate of $TGF\text{-}\beta$ production by mutant cells
$\tau_{\text{Treg}}$	rate of $TGF\text{-}\beta$ production by Treg

Table 2.2: **The model parameter names and descriptions.** Note that many of these values are affected by the inflammation state of the system.

### 2.6.3 Parameter values used for simulation

Name	Description	INFL Low Value	INFL High Value
$p$	weight of proliferation for tissue cells	0.28	
$d_C$	weight of apoptosis for tissue cells	0.14	

$\Delta_{\text{MIE}}$	MIE	0.6	
$\Delta_{\text{MGA}}$	MGA	0.2	
$\Delta_A$	proportional decrease to weight of apoptosis for cells with mutated apoptosis pathway	0.3	
$\Delta_{\text{IE}}$	proportional increase to weight of immune evasion for cells with mutated immune evasion pathway	0.48	
$\Delta_P$	proportional increase to weight of proliferation for cells with mutated proliferation pathway	0.36	
$K_0$	EC50 term for negative feedback of tissue cells on own proliferation	80 cells	
$K_1$	EC50 term for probability of NK cell finding mutant cell	8 cells	
$K_2$	EC50 term for Treg inhibition of cytotoxic functions	5 cells / volume	0.025 cells / volume
$K_3$	EC50 term for cumulative absorption of $TGF-\beta$	200 amount / volume	

$K_4$	EC50 term for $TGF-\beta$ activation of Tregs	50 amount / volume	
$E_{NK}$	weight of NKs clearing mutants	10	30
$E_{CTL}$	weight of CTLs clearing mutants	200	600
$\sigma_{NK}$	NK source rate	1.3 cells / cycle	
$\sigma_{CTL}$	CTL source rate per cleared mutant cell	100 cells / (cleared mutants $\times$ cycles)	
$\sigma_{Treg}$	Treg source rate per cleared mutant cell	200 cells / (cleared mutants $\times$ concentration of $TGF-\beta$ $\times$ cycles)	
$d_{NK}$	NK death rate	0.13 / cycle	
$d_{CTL}$	CTL death rate	0.0260 / cycle	
$d_{Treg}$	Treg death rate	0.0260 / cycle	
$k_{EMT}$	EMT/MET rate	0.01 / concentration of $TGF-\beta$	
$\sigma$	standard deviation of noise in $TGF-\beta$ each cell receives	6 concentration of $TGF-\beta$	
$\tau_{max}$	max amount of $TGF-\beta$ any cell can receive	500 concentration of $TGF-\beta$	

$\tau_{\text{MUT}}$	rate of $TGF-\beta$ production by mutant cells	0.05 concentration of $TGF-\beta$ / cell / cycle	
$\tau_{\text{Treg}}$	rate of $TGF-\beta$ production by Treg	0.5 concentration of $TGF-\beta$ / cell / cycle	
	RP Cancer Line	0.5	
	INFL High Duration	30 cycles	
	INFL Low Duration	30 cycles	
	Mes Threshold	0.7	
	maximum initial mutation damage after warmup	0.01	
	increase in probability to mutate for non-mutating proliferating cells	0.0001	

Table 2.3: **The model parameter names, descriptions, and values during both low and high inflammation.** Parameters with only one value do not change with the inflammatory state.

## 2.6.4 TCGA Analysis

### Overview

As stated in the main text, we developed methods to study: i) how the synergistic effects of EMT + inflammation compare to the effects of each of these individually; and ii) the importance of mesenchymal proliferation rates in determining cancer prognosis (Fig. 2.11),

which allow us to test predictions from the agent-based model.

For (i), we identified cases where synergistic effects due to the combination of EMT and inflammation pathways have greater influence on overall survival than the effects of either of these factors alone. For each cancer type, we obtained from MSigDB [83] gene sets that contain EMT or Inflammation-related genes and, for each gene set, we tested whether EMT, inflammation, or the combination of these two effects best predicts overall survival. We selected for those gene sets that exhibit strong synergistic effects as identified by a Cox proportional-hazard (CPH) model. For tumor types where synergistic effects were most evident, we asked whether unsupervised clustering of patients, based on a low-dimensional representation of the combination gene set, could predict statistically significant differences in overall survival via the Kaplan-Meier (KM) model. For tumor types where the both the CPH and KM analysis were consistent, we conducted further analysis (ii) of the role played by proliferation on tumor invasiveness.

For (ii), we utilized the TCGA outcome “disease-free interval” (DFI) as it best resembles the invasion-free survival metric used in modeling (in many cases, disease may be undetected until it becomes invasive). We noticed that these times were highly bi-modal, suggesting that the GRN-based regulation of invasion could be learned via binary classification. We determined the patient DFI class by fitting these invasion times to a two-component Gaussian mixture model, which assigns each patient to either high-DFI or low-DFI. We then used Gaussian process classification to learn the regulatory structure of a group of mesenchymal proliferation genes based on their ability to predict DFI class. Specifically, we clustered the genes based on the rank order statistics of their respective maximum-a-posteriori (MAP) factor-analysis distances. Finally, we used simulations of the learned model to further examine the co-regulation of these genes, highlighting the interaction between immunity, tumor progression and invasiveness in the context of treatment-response. Full details of the methods used for this analysis can be found in the Supplementary Text.

## Clinical endpoints

The Cancer Genome Atlas (TCGA) [20] provides multiple clinical endpoints, including overall survival (OS) and disease-free interval (DFI) [88]. In order to investigate the link between EMT, inflammation and invasive phenotype, we corroborate our model with clinical data from TCGA in a two pronged approach:

- **Overall Survival:** Test whether EMT and inflammation can jointly separate clinical cohorts based on the OS endpoint for a selection of cancer sub-types (Section 2.6.4).
- **Disease-Free Interval:** Identify pathway genes which regulate the proliferation/tumor-invasiveness axis in the context of a synergistic EMT/Inflam effect (Section 2.6.4).

Our analysis of tumor invasiveness due to mesenchymal growth arrest, in the context of EMT and inflammation takes place in two steps (Fig. 2.11): First, identify relevant cancers by A) defining sets of genes which represent some union of inflammation and EMT pathways while simultaneously having a quantitatively greater effect on overall survival than the component pathways acting alone, and B) simulating the dosage effect of proliferation markers on tumor invasiveness, for cancer types where a synergistic EMT/Inflam pathway was identified in (A). Following guidelines published in [88], we investigated the 14 TCGA tumor types recommended for both OS and DFI analysis.

## Overall survival

**Cox-PH Model** While immunological interactions and EMT are known to be related [52], there is uncertainty regarding both the individual pathways which govern this dependence and the extent to which the interaction between inflammation and EMT is synergistic. Our approach identifies pathways by gene set (among all pairwise combinations of EMT and

inflammation gene sets available from MSigDB [82]) for which the synergistic relationship between EMT and inflammation has a greater effect on OS than either process individually. For each combination of gene sets, we created three (one-dimensional) UMAP projections [93] of the data, one each from A) the EMT genes, B) the Inflammation genes, and C) the concatenation of (A) and (B). This yielded a three-dimensional projection of the data, on which we build a Cox proportional hazard model (CoxPH). This approach resembles a PCA-based approach introduced in [157] for SNP-based predictions. We identified several gene set combinations (combos) for which the global statistical significance (by likelihood ratio test) of the corresponding model was high ( $p_{LR} \leq 5e-2$ ), as were all three predictors, but for which the hazard ratios for the concatenation embedding were at least 5% greater in magnitude than either EMT or inflammation alone. Prior to CoxPH analysis, the proportional hazard assumption was tested and only tumor-type/combos were retained whose Schoenfeld residual was equal to 0 [69]. Our screen identified 13 tumor-type/combos across 8 tumor types: urothelial bladder carcinoma (BLCA, Table 2.4), Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC, Table 2.5), colon adenocarcinoma (COAD, Table 2.6), esophageal carcinoma (ESCA, Table 2.7), cervical kidney renal papillary cell carcinoma (KIRP, Tables 2.8 and 2.9), liver hepatocellular carcinoma (LIHC, Tables 2.10 and 2.11), pancreatic adenocarcinoma (PAAD, Tables 2.12 and 2.13), and uterine corpus endometrial carcinoma (UCEC, Tables 2.14,2.15 and 2.16). See supplementary file "OS\_test\_results.csv" for full results (all tumor-type/combos).

In order to provide further confirmation of a relationship between survival in these four cancers and the synergistic activation of relevant pathways, we tested the separation (adjusted  $p_{\log \text{rank}} \leq 0.05$ , [69]) of KM models fitted to subgroups defined unsupervised hierarchical density-based clustering [16, 37] (DBSCAN) of the UMAP-embedded combined gene set. We guided the unsupervised clustering by scaling down the minimum neighborhood size (starting with 30 patients) until the number of clusters was at least two. In addition to assigning cluster labels, DBSCAN determines outliers based on the the neighborhood struc-

ture of the graph [16, 37]. In our KM models and in subsequent analysis of DFI prediction, these outliers were discarded in order to ensure that groups of patients were maximally homogeneous with respect to EMT/inflammation. A single gene set combo met these criteria for the following cancers: BLCA (Table 2.17, Fig. 2.12), LIHC (Table 2.18, Fig. 2.13), and UCEC (Table 2.19, Fig. 2.14).

This analysis robustly identified BLCA, LIHC, and UCEC as cancers for which synergistic interaction between EMT and inflammation is the primary driver of patient survival. Our approach has several advantages. First, we utilized the MSigDB resource [82] in order to optimize the search space over relevant pathways. This allows the large volume of prior knowledge encoded in this database to guide exploratory data analysis that would otherwise be impossible or impractical at the transcriptomic scale [157]. Second, our use of dimensionality-reduction provides the following two-fold advantage: clear interpretation of the synergistic response between EMT and INFLAM and the compression of the parameter space to only three predictors, which means that sensitive prediction of survival can be carried out on the limited number of primary tumor samples in our data set. In the sequel, we address the role of mesenchymal proliferation pathways in the invasiveness of these tumors, by utilizing the disease-free interval (DFI) endpoint [88], rather than the OS endpoint.

## **Disease-Free Interval**

**DFI is a Clinical Analogue of Tumor Time to Invasion** Our agent-based simulations cover the incremental progression from in-situ to invasive disease from a homogeneous initial point, whereas the data in TCGA address how cancer may progress following treatment, thus comparisons between model and data should be made carefully. Nonetheless recent clinical and experimental evidence suggests that core cellular tumor dynamics are at play both during the tumor progression addressed by the model, and post-treatment progression described in data from TCGA. Of particular note, the plasticity of tumor cells allows them to

evade treatment by undergoing post-treatment processes resembling the de-novo appearance of cancer [127].

**Ontology-Based Investigation of Proliferation Pathways** As stated above, we predict that for certain EMT and inflammatory environments, the time to invasion is maximized by a specific proliferative regime, where the proliferative potential of a transformed tumor cell is being held in check by mesenchymal growth arrest programs. Therefore we investigate the timing of invasion as a function of proliferation by searching for proliferative regimes where the Disease Free Interval (DFI) is maximized for patients in remission after treatment. In contrast to the search-based strategy above, we used the Gene Ontology (GO) resource [5, 21] to select an appropriate pathway for this analysis. GO is designed to provide a semantic index of genes, allowing gene lists to be retrieved interactively by simply browsing its hierarchy. We selected GO:0010463 (Mesenchymal Cell Proliferation) for our analysis of proliferation-dependent DFI.

**Binary Classification of DFI Endpoints** Binarization of survival endpoints has previously been explored [9, 18, 29, 61, 70, 79, 85]. In contrast to previous approaches which utilize a pre-determined time threshold for the response (e.g. early and late relapse), we utilized an imputed high/low risk classification scheme based on a two-component Gaussian mixture model, which implicitly deals with cancer-specific thresholds. This approach was motivated by the observation that in all three cases, the DFI exhibited multiple modes with cancer-specific thresholds: BLCA (Fig. 2.15), LIHC (Fig. 2.17), UCEC (Fig. 2.16). Under this scheme, a tumor with a short DFI represents highly invasive disease for which the time-to-invasion is short.

### Summary of the Model

The log counts from TCGA bulk mRNA sequencing for 52 genes are used to predict the

computed DFI-class for each patient. The list of genes includes the 42 human genes from GO:0010463 "mesenchymal cell proliferation" (which omits all but LRP5 among known receptors for WNT2/11/5A) augmented with missing receptors for those Wnts: FZD2, FZD4, FZD6, FZD8, ROR1, ROR2, RYK, and LRP6. The model encodes the response  $y$  as either +1 or -1 (Eq. 2.11) for high-DFI and low-DFI respectively and is fitted via a generalized form of Bayesian logistic regression [118] using the expression levels of these genes as predictors. Since the likelihood (Eq. 2.11) is non-Gaussian, the posterior (Eq. 2.13) becomes analytically intractable, so expectation propagation (EP) is used to approximate it during inference and hyperparameter optimization [118]. Inference and hyperparameter optimization were performed using the gpstuff toolbox for MATLAB [145].

Our model simultaneously considers the (gene-expression of) multiple distinct biological pathways using the product of squared-exponential kernels over the predictor genes (Eq. 2.14). Within the context of the GP classifier, this kernel specifies the covariance of the joint distribution over any subset of the input data. The constant term  $\sigma_0$ , magnitude  $\sigma^2$ , and gene-wise length scale  $\lambda_d$  are given priors (Eqs. 2.15, 2.16, 2.17, and 2.18) which facilitate the discovery of their MAP values by EP. This basic structure assumes little prior knowledge

about the predictors, while offering good out of sample prediction accuracy.

$$p(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^n \frac{1}{1 + \exp(y_i f_i)} \quad (2.11)$$

$$p(\mathbf{f}|\mathbf{X}, \theta) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}) \quad (2.12)$$

$$p(\mathbf{f}|\mathbf{X}, \mathbf{y}, \theta) = \frac{p(\mathbf{f}|\mathbf{X}, \theta)}{p(\mathbf{X}|\theta)} \prod_{i=1}^n p(y_i|f_i) \quad (2.13)$$

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \sigma_0 + \sigma^2 \exp \left[ -\frac{1}{2} \sum_{d=1}^D \left( \frac{\mathbf{x} - \mathbf{x}'}{\lambda_d} \right)^2 \right] \quad (2.14)$$

$\mathbf{x}$  and  $\mathbf{x}'$  any two patients

$$\log(\sigma_0) \sim \mathcal{N}(0, 0.1) \quad (2.15)$$

$$\log(\sigma^2) \sim \mathcal{N}(1, 0.25) \quad (2.16)$$

$$\log(\lambda_d) \sim \mathcal{N}(1, \Sigma_0) \quad (2.17)$$

$$\Sigma_0 \sim \mathcal{IG}(3, 1) \quad (2.18)$$

This model achieves very high ( $\sim 1$ ) LOO-CV accuracy on the training data, so it is instructive to measure its performance relative to linear SVM on the sub-cohorts (noisy resamplings of the patient data) used for clustering. the average classification performance over all 1000 subcohorts is shown in the following table:

Type	$-\log(p(y))$	Naive	Linear SVM	GP
LIHC	197.800	0.582	0.582	1.0
BLCA	106.053	0.686	0.689	1.0
UCEC	62.321	0.708	0.726	1.0

Above, we list the leave-one-out cross-validated (LOO-CV) classification accuracy in each case for GP, and the 5-fold cross-validation accuracy for linear SVM, computed using

the MATLAB Optimization Toolbox. The LOO-CV approach of [104] utilizing the cavity distribution of the EP likelihood approximation is utilized for tractability. This approach aims to discover the out of sample prediction accuracy for the model while simultaneously using all the data [147]. Compared to linear SVM [40], the GP classifier for all cancers achieves 100% LOO-CV, while SVM achieves only a modest improvement over naive (selecting high-DFI for all patients) for BLCA and UCEC, while failing improve the naive estimate for LIHC. This latter result is consistent with the higher negative log marginal likelihood ( $-\log(p(y))$ ) for LIHC, indicating that the association between our chosen markers and the DFI endpoint is less justified. Therefore, LIHC was excluded from further analysis.

## 2.6.5 Cox-PH Tables

### BLCA

<i>Dependent variable:</i>	
time	
EMT	-0.045** (0.018)
INFLAM	0.025*** (0.009)
BOTH	0.082*** (0.021)
Observations	401
R <sup>2</sup>	0.056
Max. Possible R <sup>2</sup>	0.991
Log Likelihood	-924.075
Wald Test	22.290*** (df = 3)
LR Test	23.300*** (df = 3)
Score (Logrank) Test	22.401*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.4: **Gotzman EMT vs. GO Pos Acute Inflamm Ant**  
 GOTZMANN\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION\_UP  
 vs.  
 GO\_POSITIVE\_REGULATION\_OF\_ACUTE\_INFLAMMATORY\_RESPONSE\_TO\_ANTIGENIC\_STIMULUS

### CESC

<i>Dependent variable:</i>	
time	
EMT	-0.100*** (0.032)
INFLAM	-0.067*** (0.023)
BOTH	0.092** (0.041)
Observations	291
R <sup>2</sup>	0.060
Max. Possible R <sup>2</sup>	0.910
Log Likelihood	-340.512
Wald Test	18.320*** (df = 3)
LR Test	18.083*** (df = 3)
Score (Logrank) Test	18.630*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.5: **GO Pos EMT vs. GO Leuk Act**  
GO\_POSITIVE\_REGULATION\_OF\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
GO\_LEUKOCYTE\_ACTIVATION\_INVOLVED\_IN\_INFLAMMATORY\_RESPONSE

## COAD

<i>Dependent variable:</i>	
time	
EMT	-0.142*** (0.041)
INFLAM	0.067** (0.026)
BOTH	0.126*** (0.048)
Observations	276
R <sup>2</sup>	0.048
Max. Possible R <sup>2</sup>	0.904
Log Likelihood	-316.811
Wald Test	15.000*** (df = 3)
LR Test	13.455*** (df = 3)
Score (Logrank) Test	14.372*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.6: **Hollern EMT Breast vs. GO Leuk Act**  
HOLLERN\_EMT\_BREAST\_TUMOR\_UP  
vs.  
GO\_LEUKOCYTE\_ACTIVATION\_INVOLVED\_IN\_INFLAMMATORY\_RESPONSE

## ESCA

<i>Dependent variable:</i>	
	time
EMT	0.086** (0.038)
INFLAM	-0.113*** (0.042)
BOTH	0.185*** (0.064)
Observations	183
R <sup>2</sup>	0.052
Max. Possible R <sup>2</sup>	0.972
Log Likelihood	-321.566
Wald Test	9.760** (df = 3)
LR Test	9.838** (df = 3)
Score (Logrank) Test	9.802** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.7: GO Cardiac EMT vs. GO Neg Acute Inf

GO\_CARDIAC\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
GO\_NEGATIVE\_REGULATION\_OF\_ACUTE\_INFLAMMATORY\_RESPONSE

## KIRP

<i>Dependent variable:</i>	
	time
EMT	-0.090** (0.041)
INFLAM	-0.040*** (0.015)
BOTH	0.103** (0.040)
Observations	287
R <sup>2</sup>	0.101
Max. Possible R <sup>2</sup>	0.769
Log Likelihood	-194.964
Wald Test	35.480*** (df = 3)
LR Test	30.644*** (df = 3)
Score (Logrank) Test	41.090*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.8: GO Reg EMT vs. GO Neg Acute Inf

GO\_REGULATION\_OF\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
GO\_NEGATIVE\_REGULATION\_OF\_ACUTE\_INFLAMMATORY\_RESPONSE

<i>Dependent variable:</i>	
time	
EMT	-0.129*** (0.033)
INFLAM	-0.068*** (0.026)
BOTH	0.121*** (0.030)
Observations	287
R <sup>2</sup>	0.113
Max. Possible R <sup>2</sup>	0.769
Log Likelihood	-193.046
Wald Test	44.270*** (df = 3)
LR Test	34.480*** (df = 3)
Score (Logrank) Test	52.588*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.9: **GO Reg EMT vs. GO Neg Inf**

GO\_REGULATION\_OF\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
GO\_NEGATIVE\_REGULATION\_OF\_INFLAMMATORY\_RESPONSE

## LIHC

<i>Dependent variable:</i>	
time	
EMT	-0.029** (0.012)
INFLAM	-0.159** (0.065)
BOTH	0.181*** (0.062)
Observations	364
R <sup>2</sup>	0.058
Max. Possible R <sup>2</sup>	0.974
Log Likelihood	-654.198
Wald Test	20.970*** (df = 3)
LR Test	21.748*** (df = 3)
Score (Logrank) Test	21.321*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.10: **GO Cardiac EMT vs. Zhou Inf FIMA Up**

GO\_REGULATION\_OF\_CARDIAC\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
ZHOU\_INFLAMMATORY\_RESPONSE\_FIMA\_UP

<i>Dependent variable:</i>	
time	
EMT	-0.026** (0.012)
INFLAM	-0.027*** (0.010)
BOTH	0.033** (0.017)
Observations	364
R <sup>2</sup>	0.046
Max. Possible R <sup>2</sup>	0.974
Log Likelihood	-656.570
Wald Test	16.730*** (df = 3)
LR Test	17.005*** (df = 3)
Score (Logrank) Test	17.070*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.11: **GO Reg EMT Endo vs. GO Mac Inf Prot 1 Alpha**  
GO\_REGULATION\_OF\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION\_INVOLVED\_IN\_ENDOCARDIAL\_CUSHION\_FORMATION  
vs.  
GO\_MACROPHAGE\_INFLAMMATORY\_PROTEIN\_1\_ALPHA\_PRODUCTION

## PAAD

<i>Dependent variable:</i>	
time	
EMT	-0.036** (0.018)
INFLAM	-0.063** (0.029)
BOTH	0.052*** (0.018)
Observations	177
R <sup>2</sup>	0.061
Max. Possible R <sup>2</sup>	0.991
Log Likelihood	-407.162
Wald Test	10.610** (df = 3)
LR Test	11.187** (df = 3)
Score (Logrank) Test	10.742** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.12: **Alonso Met EMT Down vs. GO Neg Reg Inf**  
ALONSO\_METASTASIS\_EMT\_DN  
vs.  
GO\_NEGATIVE\_REGULATION\_OF\_INFLAMMATORY\_RESPONSE

<i>Dependent variable:</i>	
time	
EMT	-0.099*** (0.033)
INFLAM	-0.043** (0.022)
BOTH	0.070** (0.029)
Observations	177
R <sup>2</sup>	0.099
Max. Possible R <sup>2</sup>	0.991
Log Likelihood	-403.487
Wald Test	17.620*** (df = 3)
LR Test	18.536*** (df = 3)
Score (Logrank) Test	17.776*** (df = 3)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.13: **Jechlinger EMT Down vs. GO Pos Reg Cyto Prod**

JECHLINGER\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION\_DN  
vs.  
GO\_POSITIVE\_REGULATION\_OF\_CYTOKINE\_PRODUCTION\_INVOLVED\_IN\_INFLAMMATORY\_RESPONSE

## UCEC

<i>Dependent variable:</i>	
time	
EMT	-0.158*** (0.045)
INFLAM	-0.140** (0.062)
BOTH	0.216*** (0.075)
Observations	174
R <sup>2</sup>	0.081
Max. Possible R <sup>2</sup>	0.789
Log Likelihood	-127.984
Wald Test	14.660*** (df = 3)
LR Test	14.652*** (df = 3)
Score (Logrank) Test	14.971*** (df = 3)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.14: **Alonso Met EMT Up vs. Fulcher Inf Resp Lectin LPS Down**

ALONSO\_METASTASIS\_EMT\_UP  
vs.  
FULCHER\_INFLAMMATORY\_RESPONSE\_LLECTIN\_VS\_LPS\_DN

<i>Dependent variable:</i>	
time	
EMT	-0.179*** (0.054)
INFLAM	-0.047** (0.019)
BOTH	0.232*** (0.079)
Observations	174
R <sup>2</sup>	0.088
Max. Possible R <sup>2</sup>	0.789
Log Likelihood	-127.269
Wald Test	11.950*** (df = 3)
LR Test	16.082*** (df = 3)
Score (Logrank) Test	12.639*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.15: **GO Cardiac EMT vs. GO Reg Inf Resp Wound**  
GO\_CARDIAC\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
GO\_REGULATION\_OF\_INFLAMMATORY\_RESPONSE\_TO\_WOUNDING

<i>Dependent variable:</i>	
time	
EMT	-0.038** (0.019)
INFLAM	-0.172*** (0.044)
BOTH	0.174*** (0.050)
Observations	174
R <sup>2</sup>	0.094
Max. Possible R <sup>2</sup>	0.789
Log Likelihood	-126.768
Wald Test	21.670*** (df = 3)
LR Test	17.084*** (df = 3)
Score (Logrank) Test	16.551*** (df = 3)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 2.16: **GO Card EMT vs. Wunder Inf Resp Chol Up**  
GO\_CARDIAC\_EPITHELIAL\_TO\_MESENCHYMAL\_TRANSITION  
vs.  
WUNDER\_INFLAMMATORY\_RESPONSE\_AND\_CHOLESTEROL\_UP

## 2.6.6 KM Tables

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Cluster 1	118	35.00	54.28	6.85	11.61
Cluster 2	52	25.00	19.29	1.69	1.99
Cluster 3	38	20.00	15.50	1.31	1.49
Cluster 4	93	54.00	44.93	1.83	2.79

p=0.008

Table 2.17: **Gotzman EMT vs. GO Pos Acute Inflamm Ant**

GOTZMANN.EPITHELIAL.TO.MESENCHYMAL.TRANSITION.UP  
vs.  
GO.POSITIVE.REGULATION.OF.ACUTE.INFLAMMATORY.RESPONSE.TO.ANTIGENIC.STIMULUS

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Cluster 1	160	69.00	49.64	7.55	15.43
Cluster 2	110	32.00	51.36	7.30	15.43

p=5e-8

Table 2.18: **GO Reg EMT Endo vs. GO Mac Inf Prot 1 Alpha**

GO.REGULATION.OF.EPITHELIAL.TO.MESENCHYMAL.TRANSITION.INVOLVED.IN.ENDOCARDIAL.CUSHION.FORMATION  
vs.  
GO.MACROPHAGE.INFLAMMATORY.PROTEIN.1.ALPHA.PRODUCTION

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Cluster 1	88	21.00	14.86	2.54	5.12
Cluster 2	76	9.00	15.14	2.49	5.12

p=0.02

Table 2.19: **GO Cardiac EMT vs. GO Reg Inf Resp Wound**

GO.CARDIAC.EPITHELIAL.TO.MESENCHYMAL.TRANSITION  
vs.  
GO.REGULATION.OF.INFLAMMATORY.RESPONSE.TO.WOUNDING

Gene(s)	Cancer	Effect	Eff.Type	Citation
Wnts	Both	Onc	NA	[71, 12, 7, 100, 94]
FGFR2	BLCA	Supp	NA	[121]
FGFR2	UCEC	Onc	NA	[33]
FBXW4	BLCA	Supp	NA	[90]
HAND2	BLCA	Supp	NA	[148]
HAND2	UCEC	Supp	NA	[33]
FOXF1	BLCA	Supp	NA	New
FGFR2+FBXW4	BLCA	Supp	Coop	New
FGFR2+HAND2	UCEC	Supp	Ant	New

Table 2.20: **Summary of reported findings.** Column Key: Genes = the genes in the specified relationship, Cancer = BLCA or UCEC, Effect = Oncogenic or Suppressor, Effect Type = NA (single gene), Antagonistic or Cooperative, Status = Known or Unknown, Citation = reference utilized in the manuscript to reference a known effect

### 2.6.7 Extended TCGA Results

Given the model prediction that mesenchymal proliferation rates exert essential control on invasion-free survival times (Fig. 2.4E), we performed analysis of 14 cancer types from TCGA to assess the importance of mesenchymal proliferation-associated genes using clinical outcomes. We assessed the effects of mesenchymal proliferation in cancer gene expression data (Fig. 2.11). Using overall survival as an endpoint, and applying strict significance thresholds (Section 2.6.4), we found that three cancer types contained EMT-inflammation associated genes that predicted clear differences between patient groups. The three significant tumor types were bladder (BLCA), uterine (UCEC), and liver cancer (LIHC).

Next we modeled the effects of relevant gene sets on invasion-free survival, using the disease-free interval (DFI) as the endpoint. We clustered patients from each tumor type into two groups (high or low) based on their DFI (Section 2.6.4); these data contain 184 patients for BLCA, 114 patients for UCEC, and 311 patients for LIHC. For this clustering, the predictive accuracies (obtained by leave-one-out cross-validation) were 0.68 (BLCA), 0.69 (UCEC), and 0.621 (LIHC). We note that it is encouraging to obtain this level of accuracy on what is a

challenging task: unsupervised prediction of survival differences, using a set of only 40 genes and with only  $\sim 100$  patients per tumor type. Several of the other 11 cancer types tested also displayed mesenchymal proliferation-associated effects, however these cancers were filtered out at the previous step, as they did not meet the significance thresholds set above.

We used Gaussian process classification to identify relationships between mesenchymal proliferation genes based on their ability to predict invasiveness (high or low DFI). We focus on interactions within the *TGF- $\beta$*  and Wnt pathways, given their important roles in mediating EMT [113], and regulating cancer stem cell identity [103, 137]. *TGF- $\beta$*  and Wnt pathways interact at multiple points, including through the *LEF1/TCF* complex [75], and via dimerization of their respective membrane-bound receptors [100]. The left-hand column of Figs. 2.5 and 2.6 shows the density of patient data, and thus the region (cyan box) to focus predictions on, over a two-dimensional gene expression region. The middle column shows the probability of high DFI over the same region, and the right-hand column depicts a slice through the co-expression plot.

In agreement with the literature, our results show that for both canonical and non-canonical Wnt signaling, higher levels of signaling leads to worse outcomes (Fig. 2.5A-C and Fig. 2.6A-C, summarized in the right-hand column): as the co-expression of the Wnt ligand and its receptor increases, the probability of a high DFI (better outcome) decreases. Studying the Wnt ligand-receptor predictions, though not monotonic in all cases, comparing low vs. high expression levels shows clear differences. This is the case for all three ligand-receptor pairs studied, for both BLCA and UCEC, with the exception in UCEC of *WNT11* and *FZD8* (Fig. 2.6B), where the co-expression effects are less clear. Overall, these predictions agree with expected tumorigenic roles for canonical [71, 12] and non-canonical [7, 100, 94] Wnt signaling in bladder (BLCA) and uterine (UCEC) cancers.

We identified several other gene pairs that can predict differences between high vs. low DFI patient groups (Fig. 2.5D-F and Fig. 2.6D-F). For the gene pairs (*FGFR2*, *FBXW4*)

and (*FGFR2*, *FOXF1*) we see that a “goldilocks” region exists that most benefits survival - displaying striking similarity to the predicted mesenchymal effects in our model (Fig. 2.4E). We see that the predictions made by our modeling of single gene effects are in close agreement with the literature. Based on the agreement seen, we studied several new predictions made regarding the joint effects of co-expression on patient outcomes.

In Fig. 2.5D (middle panel) we show that a strong tumor suppressor effect of *FGFR2* in bladder cancer is predicted by our model. Although FGF signaling plays opposing roles in cancer, and FGFs can be up-regulated in tumors relying on FGF signaling for growth [1], *FGFR2* is implicated as a tumor suppressor in prostate and bladder cancer [78, 121]. We also predict a suppressive role for *FBXW4* (Fig. 2.5D, middle panel): for given *FGFR2* expression, increasing *FBXW4* leads to better outcomes. This agrees with literature suggesting that *FBXW4* is lost or mutated in almost 40% of urinary tract cancers [90]. Co-expression analysis also gives a new prediction: *FGFR2* and *FBXW4* act synergistically in BLCA, such that higher expression levels of both leads to greater outcomes than high expression of either gene alone (2.5D, middle & right panels). In comparison, for UCEC, the role is less clear, although the tumorigenic effect of *FGFR2* in uterine cancer is evident at high levels of *FBXW4* (Fig. 2.6D), which is in line with previous studies reporting mutations that provide constitutive activation of *FGFR2* in a subset of endometrial cancer [33].

We highlight two further notable predictions. First, for BLCA, high *FGFR2* and *FOXF1* co-expression improves patient outcomes (Fig. 2.5E). The tumor suppressor *FOXF1* is a p53 target and it is epigenetically silenced in breast cancer [89, 141], however it was not previously known to have a tumor-suppressive role in BLCA either alone or co-expressed with *FGFR2*. This effect is not seen for UCEC (Fig. 2.6E), where our model predicts that the effects of *FGFR2* are tumorigenic in this region, but not affected by *FOXF1* expression (i.e. no significant differences between high and low co-expression). The second new prediction is that high co-expression of *FGFR2* and *HAND2* improves outcomes in UCEC (Fig.

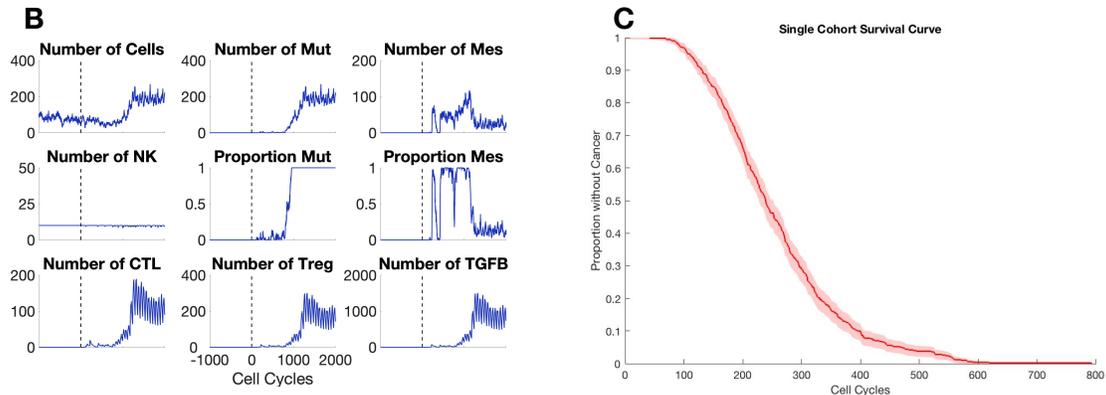


Figure 2.7: **Fig. 2.1BC without targeting cells with pathways mutations.** A. Single patient trajectory without immune cells targeting cells with pathways mutations. Compare to Fig. 2.1B. B. Sample cohort survival curve without immune cells targeting cells with pathways mutations. Compare to Fig. 2.1C.

2.6F); in contrast to the effects seen for the co-expression of *FGFR2* with either *FBXW4* or *FOXF1* (Fig. 2.6D-E), where, in each case, higher *FGFR2* expression led to worse outcomes. *HAND2* antagonizes FGF-dependent epithelial cell proliferation and is a critical regulatory component of both healthy and cancerous endometrial proliferation [66, 81]. For BLCA, we observe a less pronounced although still suppressive effect due to *HAND2* expression (Fig. 2.5F), in line with previous reports [148].

## 2.7 Supplementary Figures

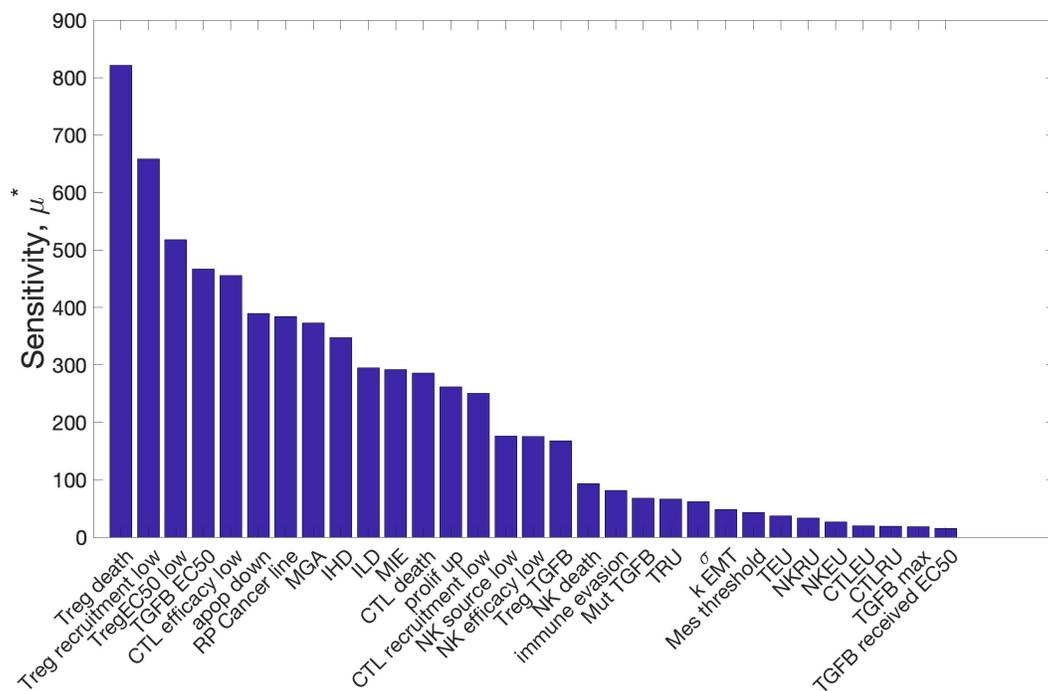


Figure 2.8: Morris-OAT global sensitivity without immune cells targeting cells with pathway mutations. Compare to Fig. 2.2.

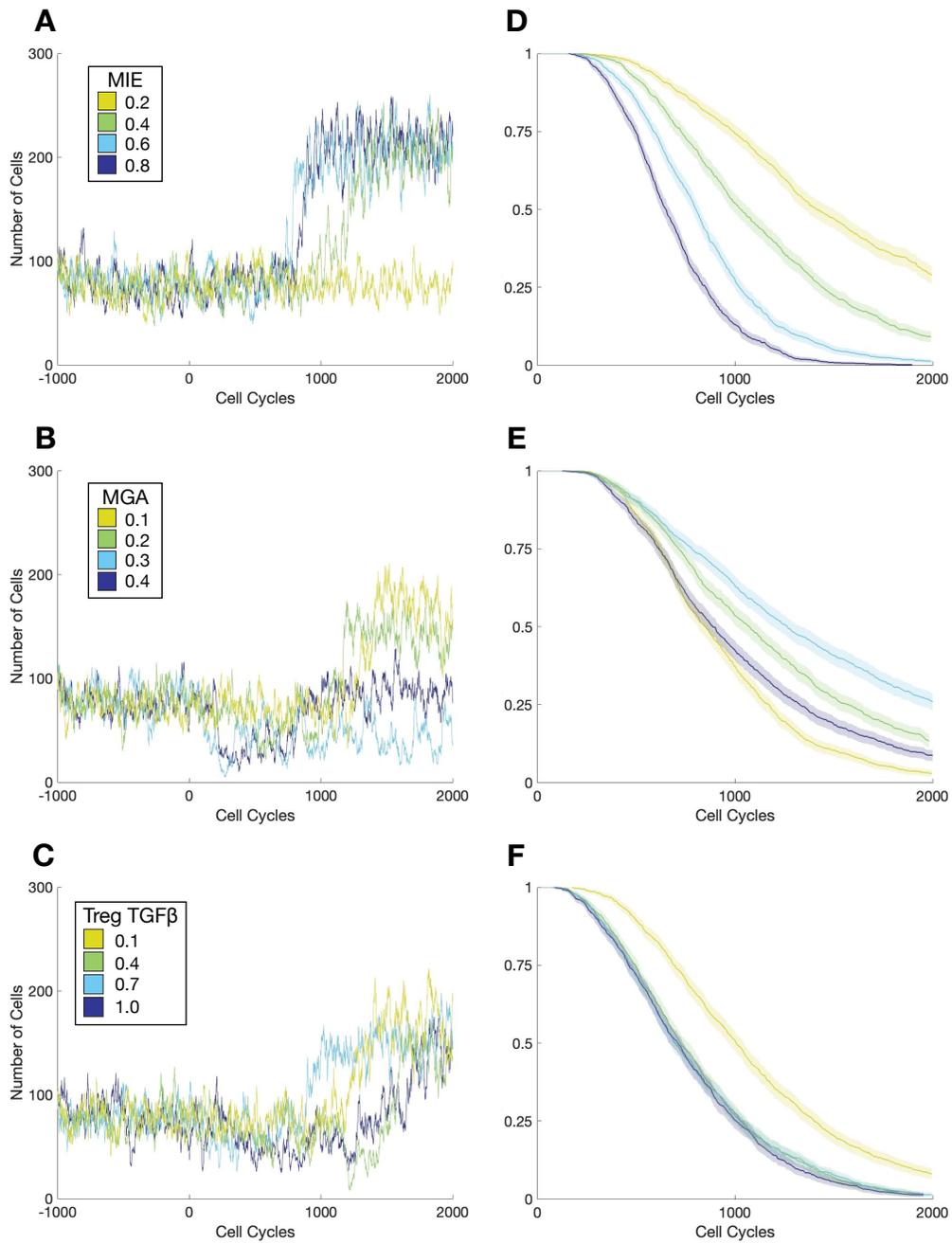


Figure 2.9: **Effects of mesenchymal tumor cell properties on the Time to Invasion without immune cells targeting cells with pathways mutations.** Compare to Fig. 2.3.

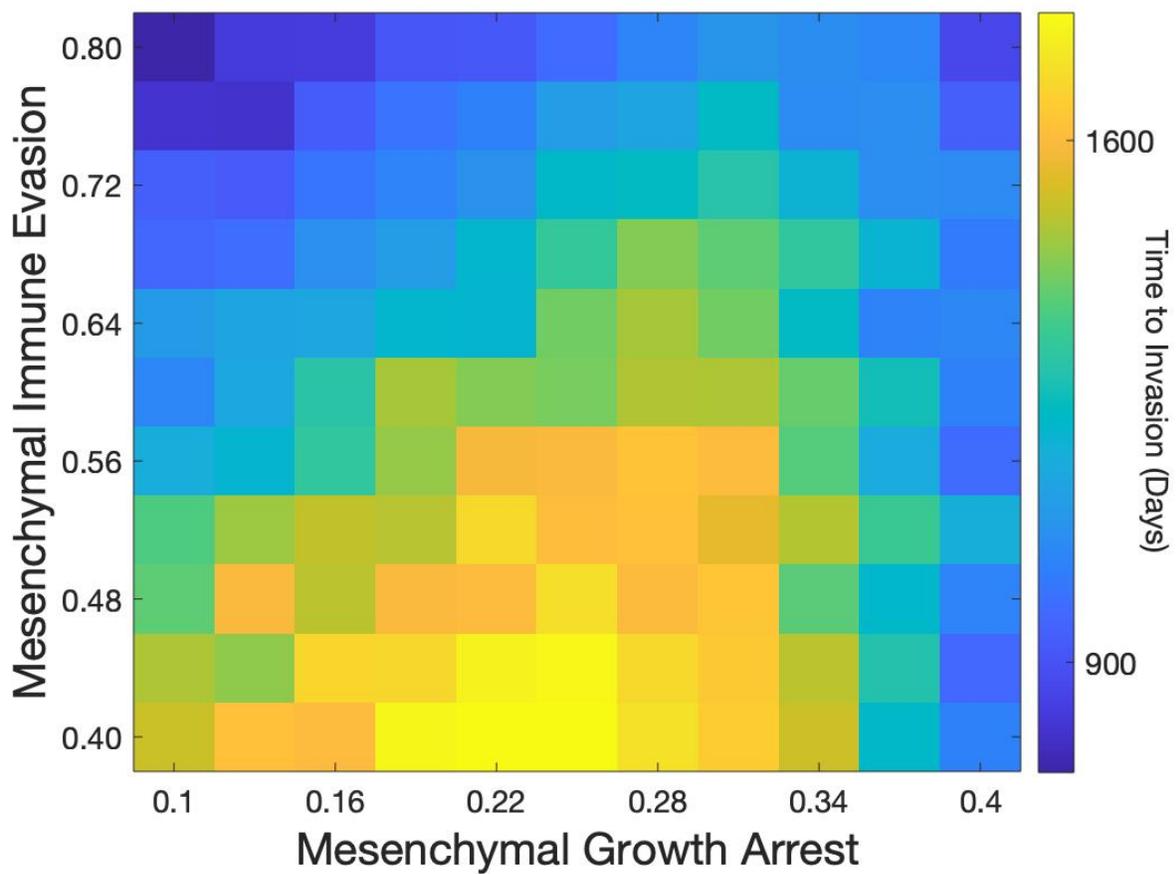


Figure 2.10: Summary of mesenchymal tumor cell properties on the Time to Invasion without immune cells targeting cells with pathways mutations. Compare to Fig. 2.4E.

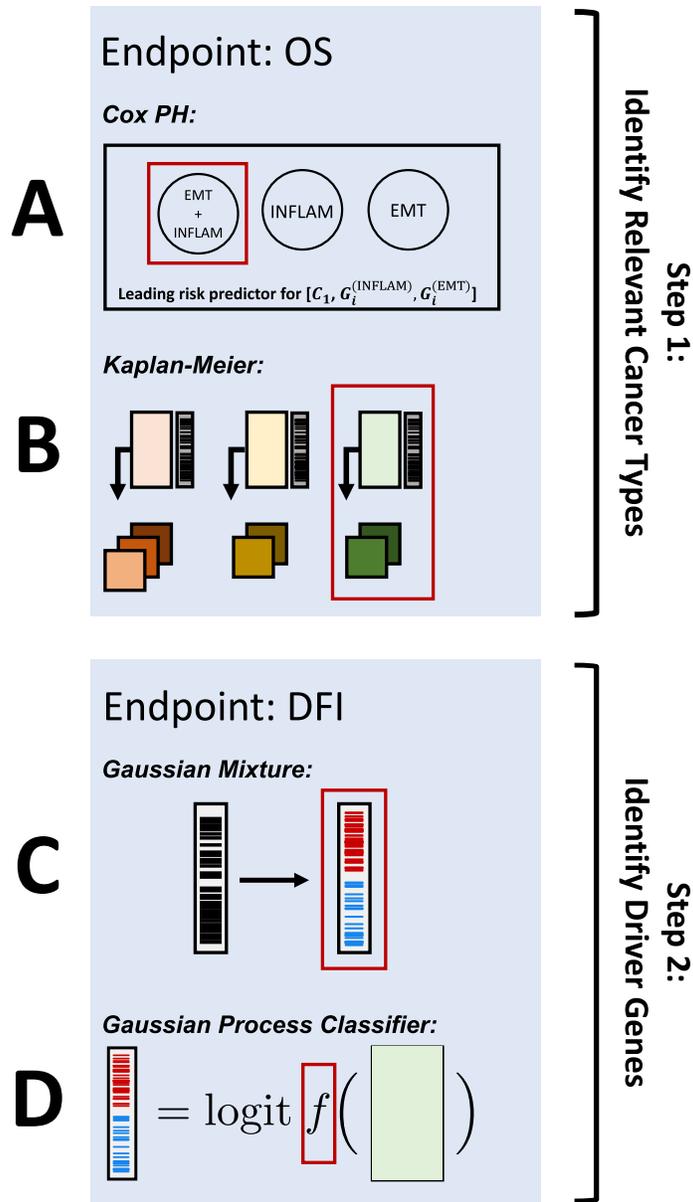


Figure 2.11: **Schematic of overall survival (OS) and disease-free interval (DFI) analysis on TCGA data.** A-B (Step 1): Find cancer types in which EMT and inflammation act synergistically on overall survival. C-D (Step 2): For cancer types identified in Step 1, identify proliferation pathways that regulate the invasiveness of these cancers. A) For each cancer type, identify pairs of inflammation and EMT gene sets where the UMAP projection over their union has a higher (magnitude) hazard ratio than either of its constituents in a three-predictor Cox-PH model with OS response in that cancer type. B) Identify cancer types, for which unsupervised DBSCAN clustering over the 1D UMAP projection of one or more EMT/Inflam union sets yields clusters whose KM survival curves are different. C) Impute the DFI-high/low class based on a two-component Gaussian mixture model of the published disease-free interval time in days. D) Identify relationships between proliferation pathways and tumor invasiveness using a GP classifier trained on the computed DFI-class from (C) and mRNA sequencing for the list of proliferation genes.

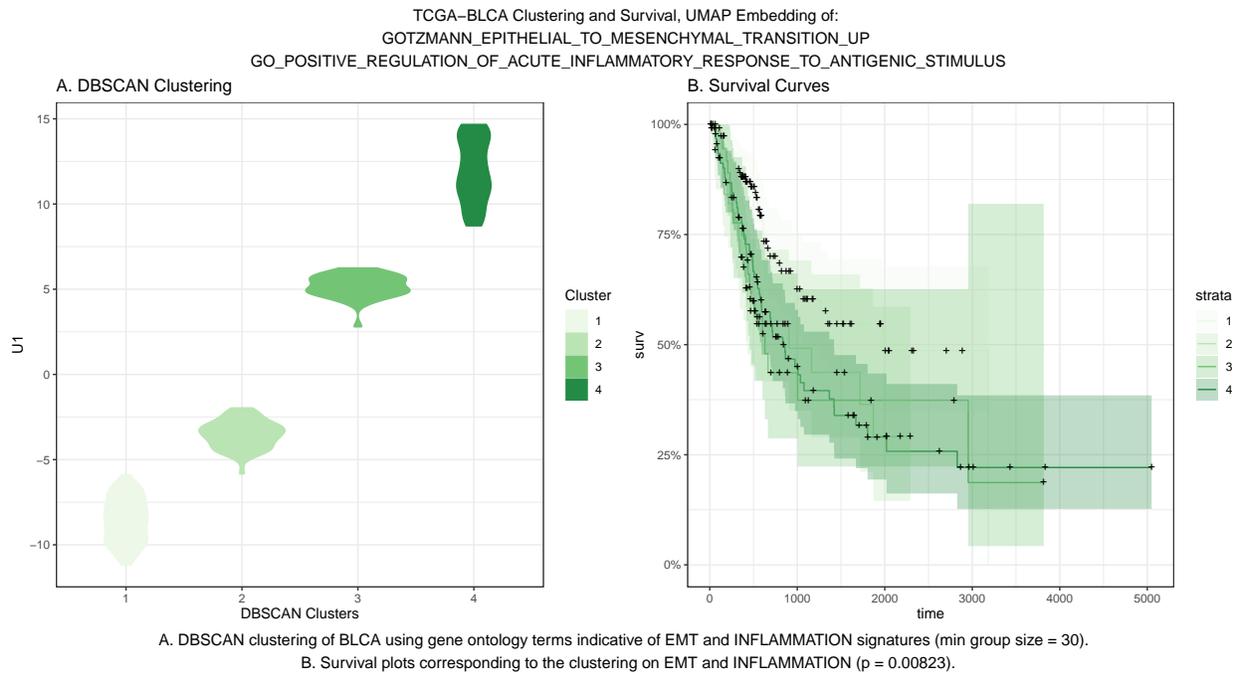


Figure 2.12: DBSCAN clusters for combined EMT+INFLAM embedding of BLCA patients and corresponding Kaplan-Meier (KM) model.

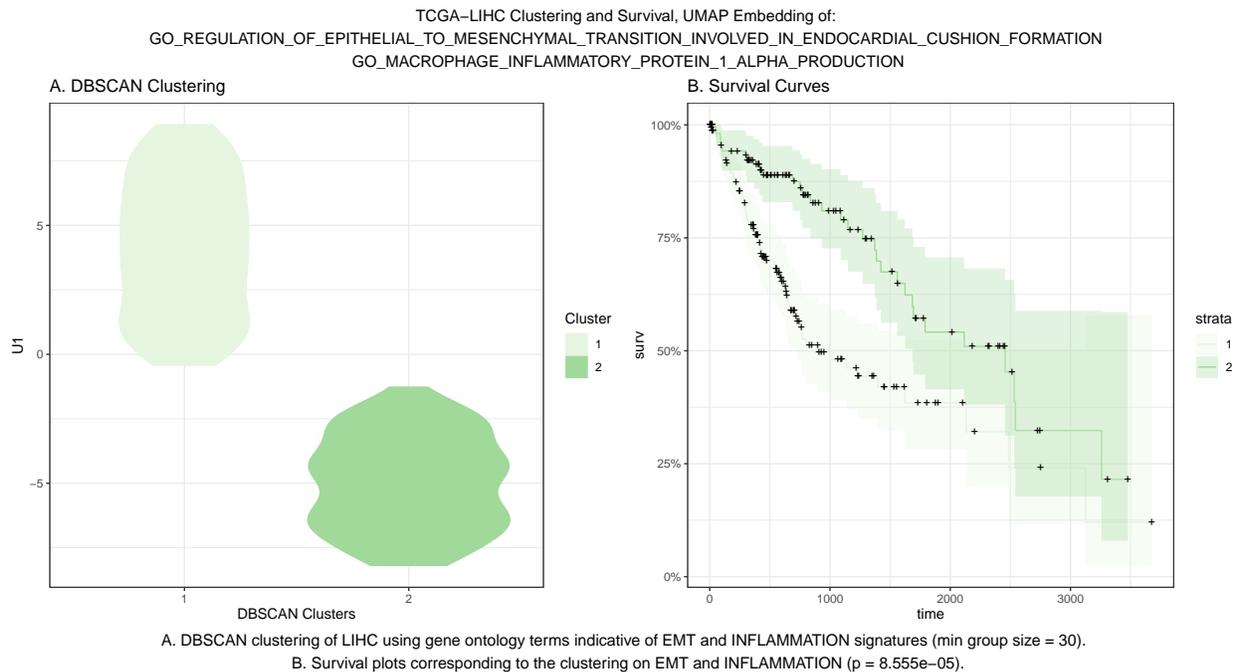


Figure 2.13: DBSCAN clusters for combined EMT+INFLAM embedding of LIHC patients and corresponding Kaplan-Meier (KM) model.

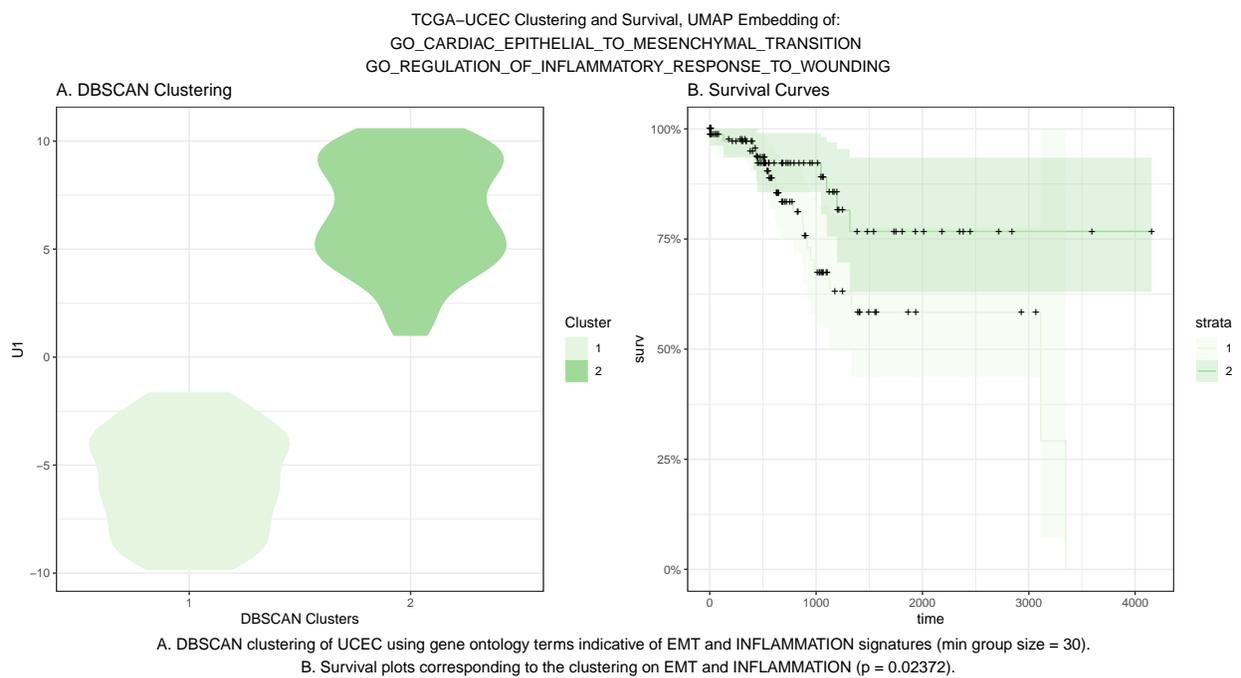


Figure 2.14: **DBSCAN clusters for combined EMT+INFLAM embedding of UCEC patients and corresponding Kaplan-Meier (KM) model.**

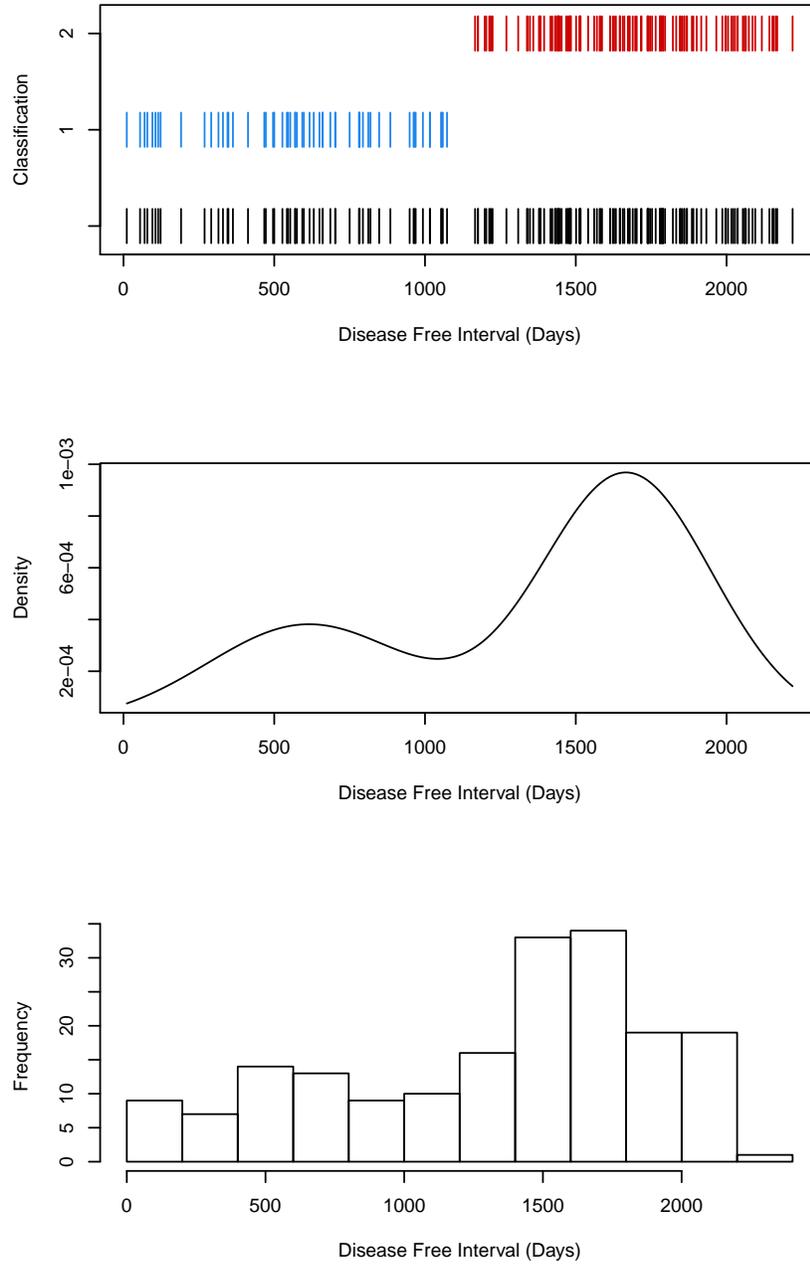


Figure 2.15: **Classification of BLCA patients by DFI.** A) Classification of BLCA patients in short DFI (blue) and long DFI (red). B) 2-component Gaussian mixture density corresponding to the above classification. C) Histogram of DFI for patients.

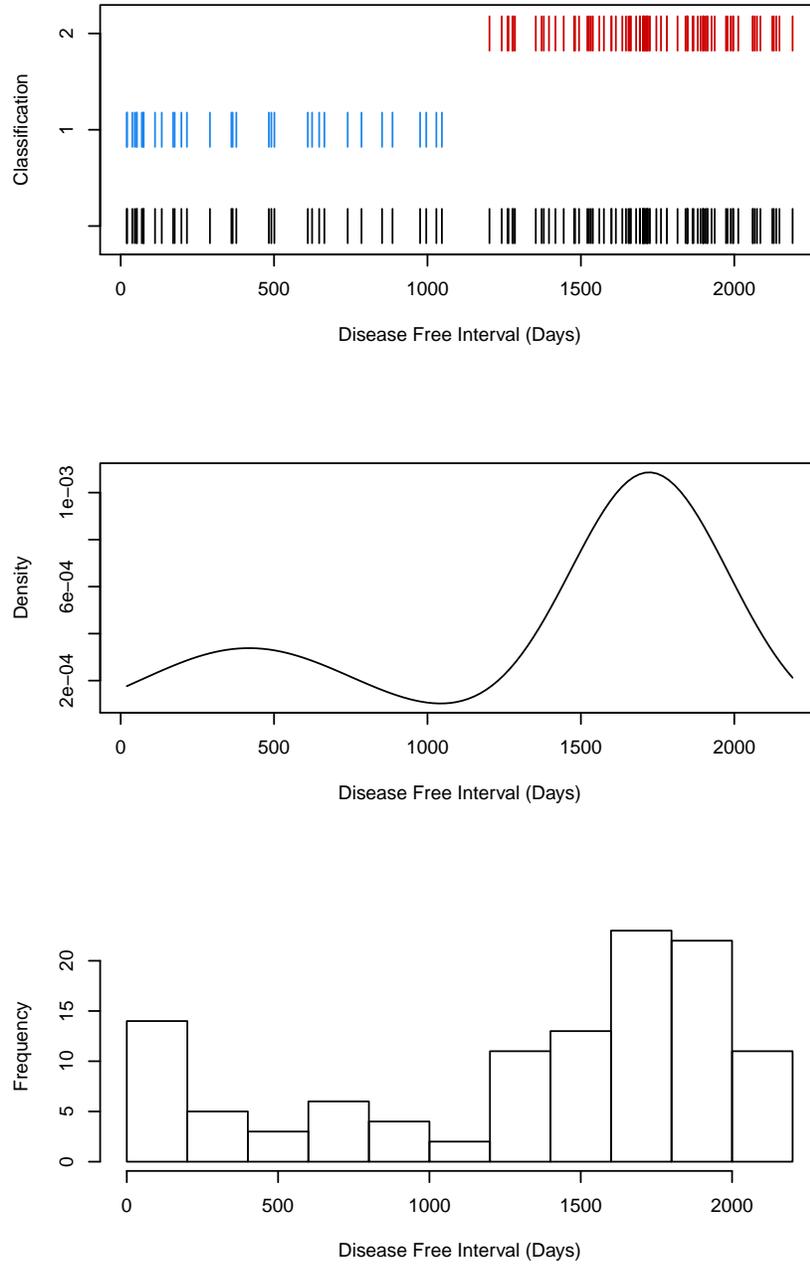


Figure 2.16: **Classification of UCEC patients by DFI.** A) Classification of UCEC patients in short DFI (blue) and long DFI (red). B) 2-component Gaussian mixture density corresponding to the above classification. C) Histogram of DFI for patients.

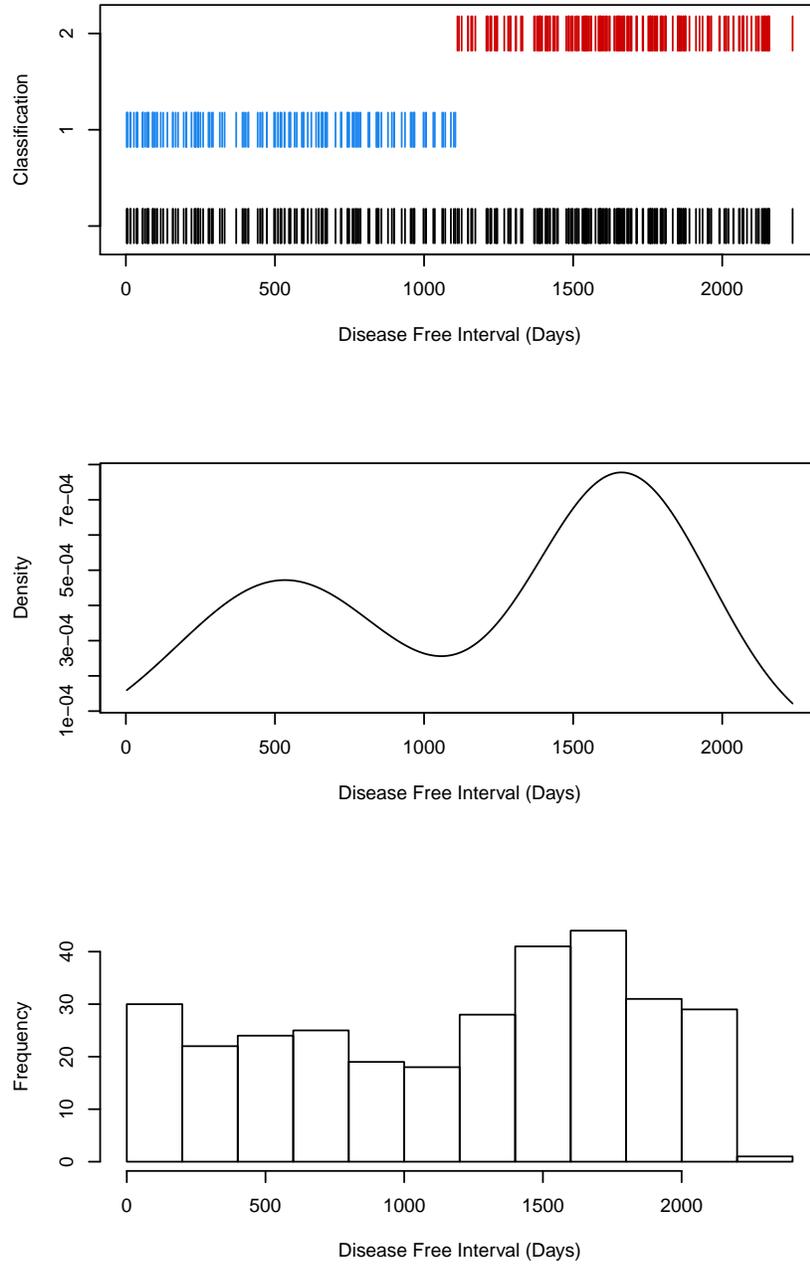


Figure 2.17: **Classification of LIHC patients by DFI.** A) Classification of LIHC patients in short DFI (blue) and long DFI (red). B) 2-component Gaussian mixture density corresponding to the above classification. C) Histogram of DFI for patients.

# Chapter 3

## Modeling interactions among B cells, macrophages, and skin cancers

This chapter is part of a paper that is in preparation. It first reports on analysis of single-cell data performed by my collaborator. This work demonstrates a new area to further explore, which we do using mathematical modeling. We use an ODE and an SDE to analyze the biological system in question.

### 3.1 Summary

The advent of immune checkpoint therapy for metastatic skin cancer has greatly improved patient survival. However, most skin cancer patients are refractory to checkpoint therapy, and the intra-immune signaling driving the response to checkpoint therapy remains uncharacterized. When comparing the immune transcriptome in the tumor microenvironment of melanoma and BCC, we found that the presence of memory B cells and macrophages negatively correlate when stratifying patients by response, with memory B cells more present

in responders. Both macrophages and plasma B cells downregulate memory B cell function in BCC, whereas only macrophages downregulate memory B cells in melanoma. We further explored the relationships between macrophages, B cells and response to checkpoint therapy by developing a stochastic differential equation model which qualitatively agrees with the data analysis. Our model reveals hereto uncharacterized differences between the TME of progressing and regressing tumors that could be used as a diagnostic tool. It also predicts interventions that can be tailored to individual patients to increase response to checkpoint therapy. Finally, it predicts BCC to be more refractory to checkpoint therapy than melanoma.

## 3.2 Introduction

Checkpoint immunotherapy has changed the lives of patients with metastatic cancer[120, 144]. These novel therapeutics can drive durable responses in many metastatic cancers, with most adverse events being grade 1 or 2[133, 86]. Checkpoint therapy’s utility remains limited, however, with most patients either not responding or acquiring resistance to treatment[133].

Current FDA-approved checkpoint inhibitors fall into two categories: CTLA-4 inhibitors and PD-1/PD-L1 inhibitors. CTLA-4 expressed by T regulatory cells (Tregs) outcompete costimulatory molecules on CTLs necessary for their activation and causes anergy. Cancer cells express PD-L1, which binds to PD-1 on CTLs, inhibits their cytotoxicity and promotes anergy and eventual apoptosis (see [120] for a full review of both mechanisms). Inhibition of either pathway leads to durable cancer regression in many cancers with different somatic mutations[120, 133].

Despite cancer checkpoint immunotherapy being hailed as being “arguably one of the most important advances in the history of cancer treatment” [120], our understanding of how these

therapies affect a system as responsive and dynamic as the immune system remains incomplete. Many studies focus on the effect of checkpoint therapy on cytotoxic T lymphocytes (CTLs or CD8+ T cells)[114, 120, 134, 160, 126, 56]. Two major recent studies sequenced the transcriptome of the TME at a single-cell level before and after checkpoint therapy in melanoma[126] and in basal cell carcinoma (BCC)[160]. Both these studies focused on the effect of checkpoint therapy on CTLs, but effects of checkpoint therapy on different immune cell types have been previously observed. Indeed, in a phase I clinical trial for one of the first FDA-approved checkpoint inhibitors (nivolumab, brand name Opdivo), divergent and even opposite effects of nivolumab on T cells and B cells were observed[11]. More recently, B cells have been shown to correlate with response to checkpoint immunotherapy even more strongly than CTL presence[56, 111]; however, this remains contentious, with other studies showing no effect[25].

These results seem to suggest that checkpoint immunotherapy has a broad impact on the immune system, including but not limited to cytotoxic T cells. Other immune cell types may be implicated in response or non-response to treatment. In addition, there has been few if any comparisons of the impact of checkpoint therapy on the immune system in different cancers. Single cell RNA-sequencing technology (scRNA-seq) is particularly well-suited to holistically analyze different immune cell types, due to its ability to capture transcriptomic data from many cell types at once. Dynamical systems modeling can supplement scRNA-seq data by capturing dynamics of cell types and can point the way to new therapeutic interventions.

We therefore independently analyzed the single-cell RNA-seq datasets[160] and[126], to compare and contrast the immune responses of responders and non-responders across cancers. We found that memory B cells are most present in responders and vice versa for macrophages. We characterized their cellular signaling and found that macrophages strongly inhibit memory B cells in melanoma non-responder patients; however, in basal cell carcinoma non-

responders, the inhibitory signaling seems to be coming most strongly from plasma B cells, which are polarized towards a more immune suppressive genotype by cancer-associated fibroblasts.

To fully capture the differences in immune suppression mechanisms between cancers and the dynamics that lead to response/non-response, we built a three-state ordinary differential equations (ODE) model that matches qualitatively with our results. We then added stochasticity to make the model an SDE. The stochastic model predicts that measuring a patient’s tumor burden in conjunction with the number of memory B cells in the tumor microenvironment following a partial response to therapy can determine if the response remains durable. In addition, the model predicts BCC to be more refractory to checkpoint therapy than melanoma. In fact, it predicts that small changes to the system brought about by immunotherapy could have adverse effects on the likelihood for regression.

### **3.3 Results**

#### **3.3.1 BCC and melanoma exhibit similar responses to checkpoint immunotherapy**

We looked at two datasets: one consisting of melanoma patients[126] and one consisting of BCC patients[160]. We started by clustering and identifying the clusters of each dataset (Fig. 3.1AB, see upcoming paper for more details of my collaborator’s methods). As expected, the majority of cells in each dataset are T cells. We then stratified immune cells in each cluster by whether they came from responders or non-responders (Fig. 3.1C; note that the skin cells in the BCC dataset were excluded). Overall, the percents of responders and non-responders in each cluster are similar across cancers. The amount of CD8+ T cells is similar in responders and non-responders in both cancers. Memory B cells seem to be most

concentrated in responders, and macrophages are the most concentrated in non-responders. To verify these results, we compared the percentage of macrophages, memory B cells and Tregs, per patient separated by responders and non-responders (Fig. 3.1D-F). We found that macrophages represent a higher percentage of cells per patient in non-responders, and that the opposite is true for memory B cells. Overall, the distribution of immune cells in the TME of responders and non-responders to immunotherapy is similar in both cancers, despite the differences in cancers, immunogenicity and sequencing technologies.

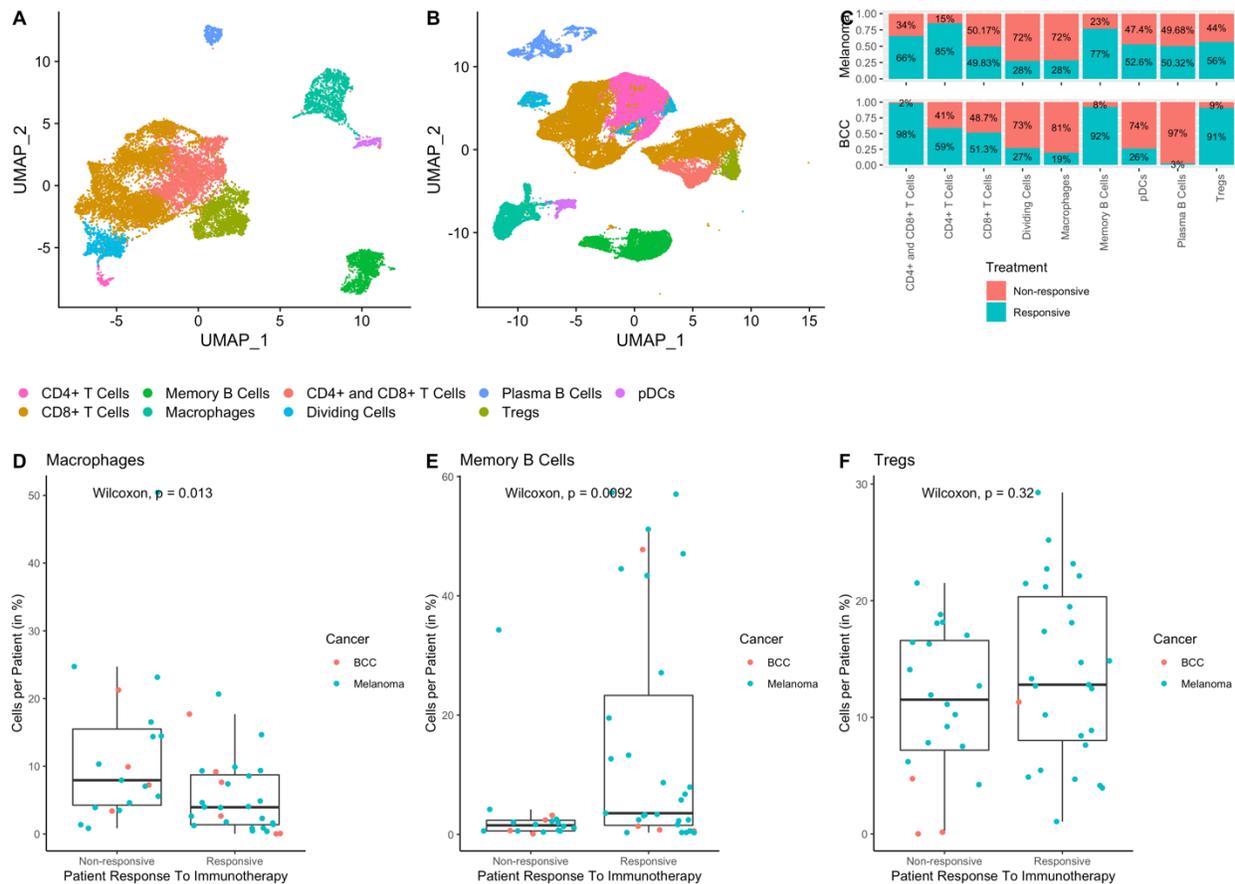
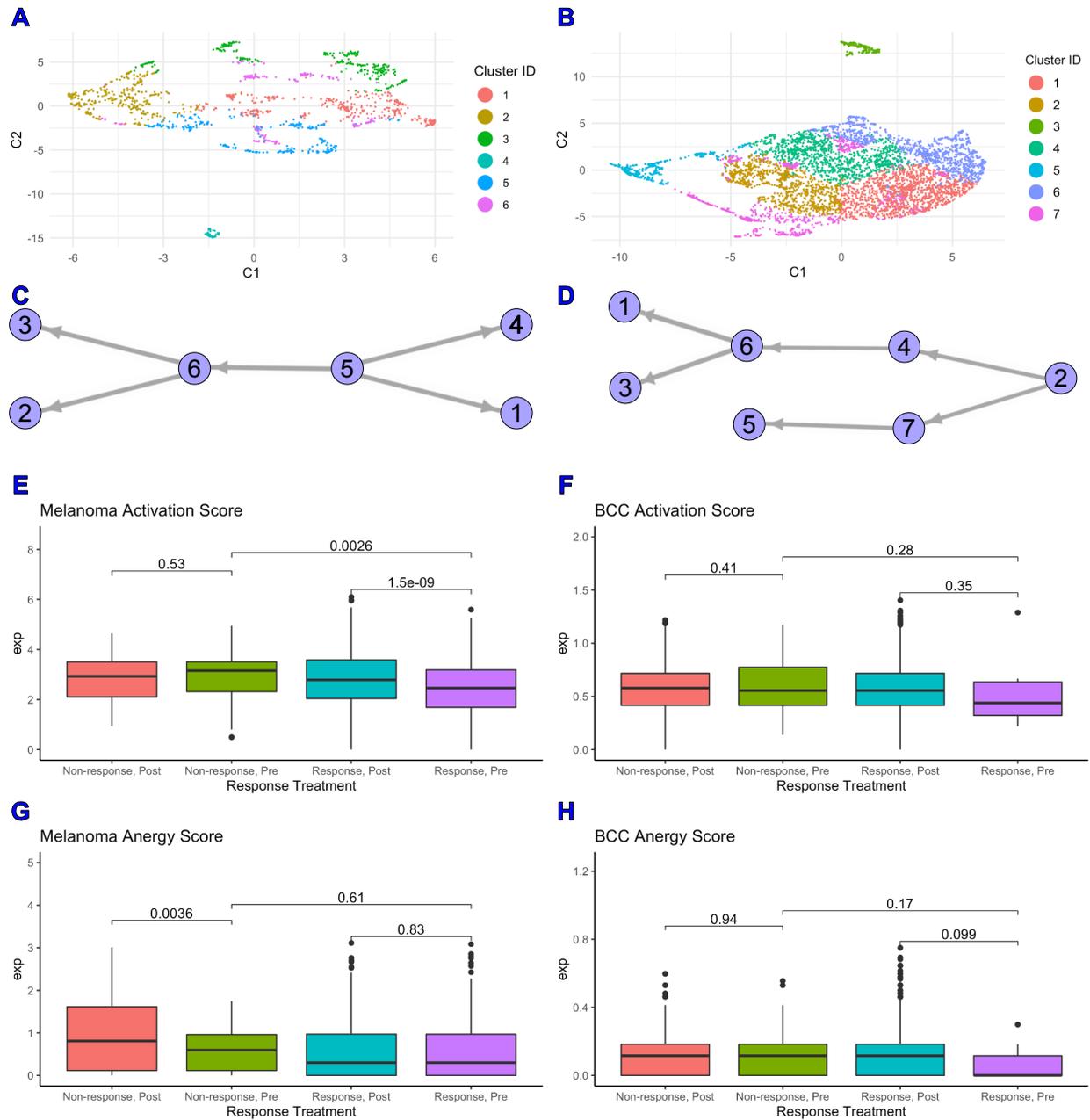


Figure 3.1: **Melanoma and BCC have similar responses to immunotherapy.** A-B): Dimensionality reduction of melanoma (A) and BCC (B). C) Distribution of cells from responders and non-responders, grouped by cluster. (D), (E) and (F): Percentage of macrophages (D), memory B cells (E) and Tregs (F) per patient, grouped by responders and non-responders.

### **3.3.2 Memory B cells are more anergic in non-responders pre-treatment in both cancers**

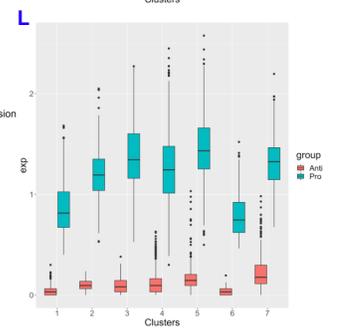
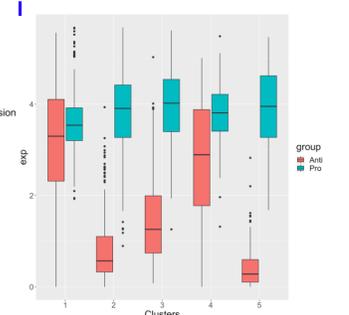
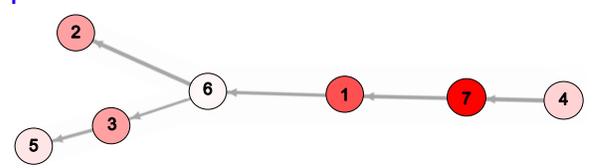
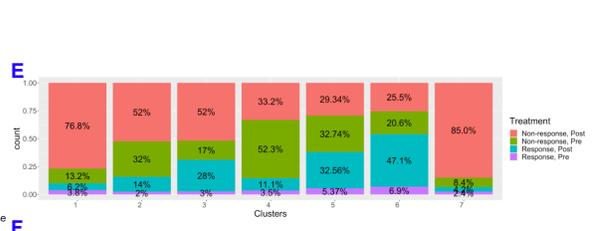
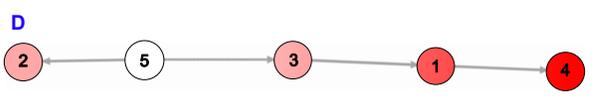
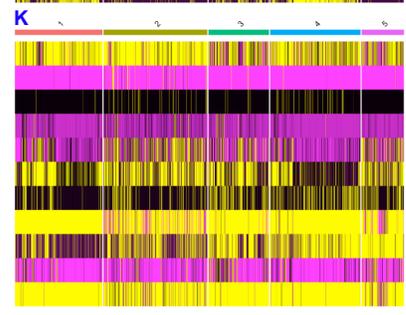
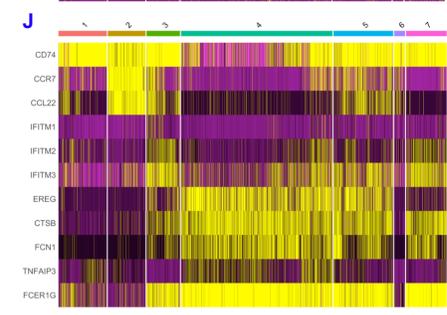
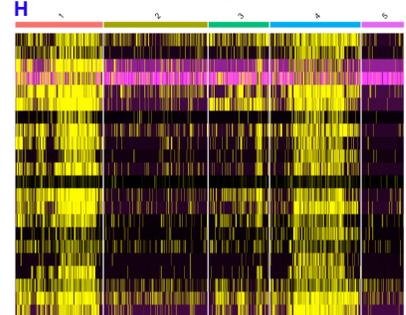
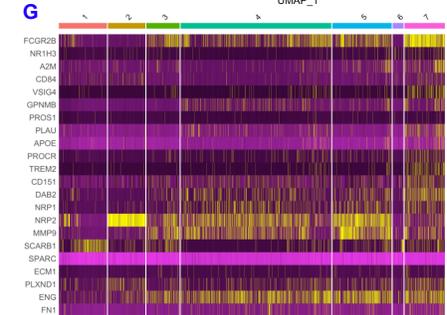
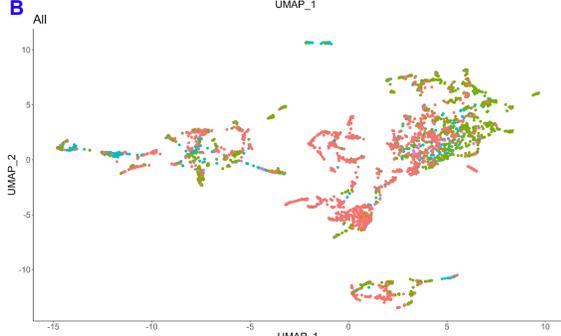
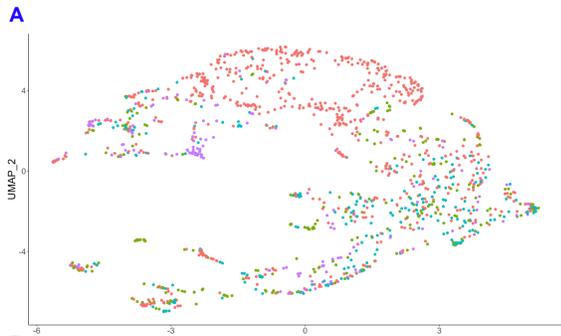
Next we sub-clustered the memory B cells in both datasets (Fig. 3.2AB) and analyzed their lineage via a similarity matrix-based optimization method for inferring clustering and lineage (see upcoming paper for more detail of my collaborator's methods)[150]. We developed two different scores for memory B cell activation and anergy. In the melanoma dataset, the activation score in responders pre-treatment is lower than in responders post-treatment, and the activation score is higher in non-responders pre-treatment than in responders pre-treatment (Fig. 3.2E). There is no difference in the anergy scores in the responders pre- and post-treatment, nor in the non-responders pre-treatment and the responders pre-treatment, but the memory B cells in non-responders post-treatment are more anergic than those in pre-treatment (Fig. 3.2G). Overall, in melanoma the memory B cells are more activated in pre-treatment non-responders and become more anergic in non-response post-treatment. The memory B cells in BCC seem to have similar trends in activation and anergy, with the interesting exception that memory B cells in responders post-treatment seem to be more anergic than in responders pre-treatment (Fig. 3.2H), which could be explained by differences in immunogenicity between melanoma and BCC (see Discussion).



**Figure 3.2: Memory B cells are more activated in non-responders pre response.** A-B) Dimensionality reduction of the memory B cells subsets of melanoma (A) and BCC (B). C-D) Pseudotime ordering of melanoma (C) and BCC (D). E-F) Activation scores of memory B cells in melanoma (E) and BCC (F). G-H) Anergy scores of memory B cells in melanoma (G) and BCC (H).

### 3.3.3 Macrophages in BCC have more of an M1 genotype, regardless of responder status

It is well-understood that macrophages can have an important role in immune suppression in cancer (see e.g. [107]), to the extent that presence of macrophages can correlate with poor prognosis and drugs are being developed to inhibit their suppressive ability[122]. In conjunction with our results indicating that macrophages are more prevalent in non-responders (Fig. 3.1C), this indicated that we should explore the macrophage subsets in both datasets. We found that the lineage for macrophages in melanoma correlates well with the increase in percent of non-responders post-treatment (Fig. 3.3C); however, in BCC the macrophage lineage does not seem to follow the same pattern (Fig. 3.3F). Seeing as macrophage plasticity is well-characterized but remains contentious[48], we built an expression profile of macrophages using genes that are labeled as either “pro-inflammatory” or “anti-inflammatory” in gene ontology[73, 21]. In the melanoma dataset, we found that the anti-inflammatory gene expression correlates well with the percentage of macrophages found in non-responders post-treatment and the macrophage lineage (Fig. 3.3BI). However, the macrophages in BCC have very low expression of anti-inflammation genes, regardless of percent non-responder post-treatment (Fig. 3.3EL). Both cancers have more macrophages in non-responders, but the macrophages have different inflammatory signatures between the two cancers and seem to be linked to different processes overall.



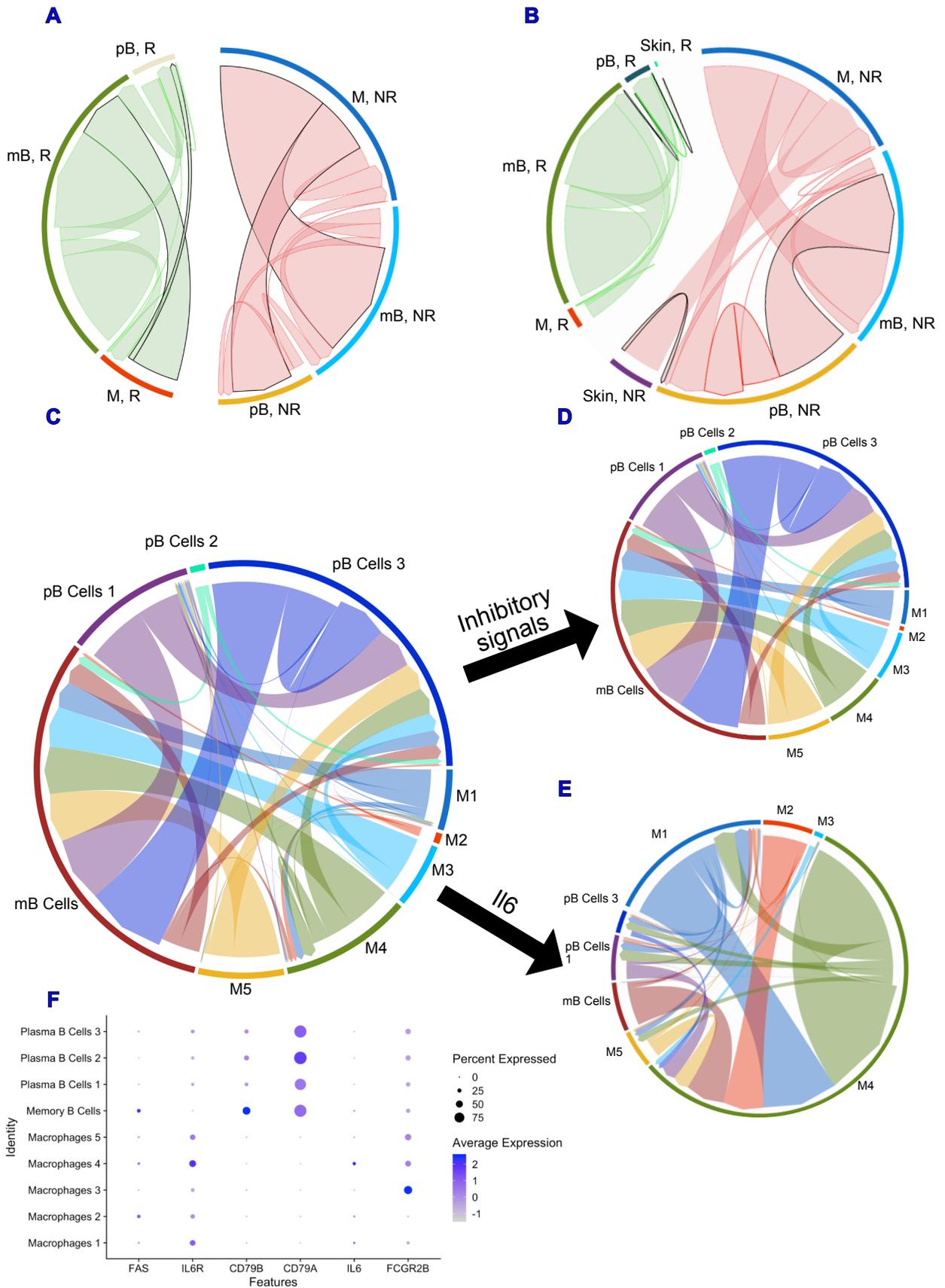
---

Figure 3.3 (*preceding page*): **Macrophages in BCC have more of a pro-inflammatory genotype, regardless of responder status.** A-B) Dimensionality reduction of the macrophage subsets of melanoma (A) and BCC (B). C, E) Percentage of Responders/Non-responders in Pre/Post treatment per macrophage cluster in melanoma (C) and BCC (E). D, F) Pseudotime of macrophage clusters in melanoma (D) and BCC (F). Each pseudotime node is qualitatively colored by percent of non-responders post treatment. The melanoma pseudotime correlates well with the percent of non-responders post-treatment, whereas the BCC pseudotime does not. G-H) Heatmaps of anti-inflammation genes in BCC macrophages (G) and melanoma macrophages (H). The expression of anti-inflammatory genes in melanoma correlate well with the fraction of non-responders post-treatment of each cluster, whereas in BCC there seems to be little correlation. J-K) pro-inflammation genes in BCC macrophages (J) and melanoma macrophages (K). I, L) Average expression of anti- and pro-inflammatory genes by cluster of macrophages in melanoma (I) and BCC (L).

### 3.3.4 Plasma B cells are suppressing memory B cells only in BCC

To understand the profile of intra-immune signaling, we calculated cluster-cluster cell signaling of memory B cells, macrophages, and plasma B cells. We included the plasma B cells in the analysis because one of the starkest differences between the two datasets is the preponderance of plasma B cells in non-responders in BCC (Fig. 3.1C). In the melanoma dataset, we found that macrophages inhibit a greater percentage of both memory B cells and plasma B cells (Fig. 3.4A). There seems to be mutual inhibition of memory B cells and plasma B cells, but this inhibition is much lower in probability relative to the macrophage inhibition, especially in non-responders. The inhibitory signaling found in the BCC dataset is comparable to that of melanoma dataset, with the notable exception that the plasma B cell inhibition seems to be much higher relative to the macrophage inhibition (Fig. 3.4B). This difference correlates with the high percentage of plasma B cells in non-responders in BCC (Fig. 3.1C). Also notable is the signaling between the skin cluster and the plasma B cells via the IL6 pathway; this signaling is stronger in non-responders, indicating that the skin cells in non-responders could be inducing Breg activation, which in turn induces plasma B cells to inhibit memory B cells (Fig. 3.4B).

To further explore the role of plasma B cells in inhibiting memory B cells in BCC, we subclustered the plasma and memory B cells and macrophages from non-responders and ran a signaling analysis via the same pathways (Fig. 3.4C). We found that there are two cluster of plasma B cells that are strongly inhibiting both other plasma B cell clusters and memory B cells (Fig. 3.4D). There seems to be a particular cluster of plasma B cells in BCC non-responder patients that is not only inhibiting memory B cells but also signaling to macrophages, potentially inducing upregulation of inhibitory signals from macrophages. Both melanoma and BCC seem to have macrophages inducing memory B cell anergy in non-responders, with the addition of a suppressive plasma B cell subset in BCC, which seems to be polarized towards a Breg genotype by skin cells.



---

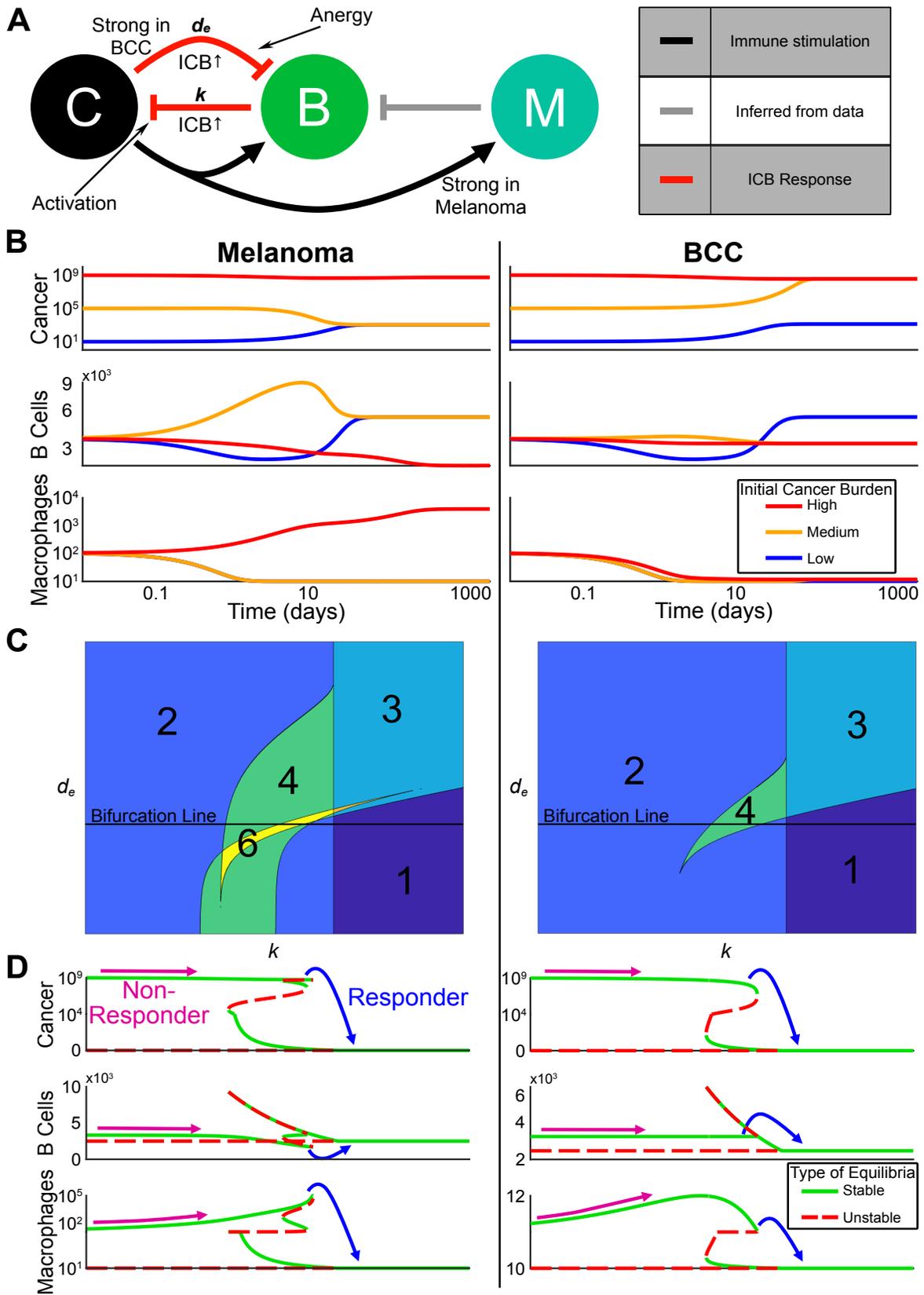
Figure 3.4 (*preceding page*): **Suppression of memory B cells by plasma B cells is only implicated in resistance to therapy in BCC.** A-B) Cell signaling of memory, plasma B cells, macrophages and skin cells in melanoma (A) and BCC (B). The plasma B cells are inhibiting the memory B cells much more strongly in BCC than in melanoma, relative to the suppression of macrophages. The signaling of the macrophages to the memory B cells and the plasma B cells (in melanoma) and the signaling of the plasma B cells to the memory B cells and the signaling from the chosen skin cluster to the plasma B cells (in BCC) is highlighted. C) Signaling of both suppressive and activating cell-cell signals from clusters defined by SoptSC of memory and plasma B cells and macrophages from BCC non-responders. D) Suppressive signaling in the clusters defined in (C). (E) Signaling of IL6 in the clusters defined in (C). F) Dotplot of ligand/receptor/target triads used to generate the signaling plot (C).

### 3.3.5 A dynamical model on interactions among memory B cells, macrophages, and skin tumors

To further explore the role of memory B cells in response to checkpoint treatment and to better understand the dynamics of the immune system during treatment, we developed a three-state ODE model based on the bioinformatic clustering, lineage and signaling analyses. We chose cancer, B cells and inhibitory macrophages (macrophages) as our state variables (Fig. 3.5A). The cancer undergoes logistic growth and has four possible steady states: none, low ( $\sim 10^3$  cells), high ( $\sim 10^8$  cells), and very high ( $\sim 10^9$  cells). The B cells kill cancer cells, and macrophages inhibit B cell proliferation. The parameters of the dynamical model were selected based on our bioinformatic analyses and previous literature.

We made two assumptions from the signaling analysis that differentiate the divergent refractory mechanisms of each cancer. First, consistent with macrophages in BCC having less of an anti-inflammatory phenotype (Fig. 3.3L), we assumed that the cancer-mediated up-regulation of (inhibitory) macrophage proliferation is weaker in BCC relative to melanoma. Second, consistent with plasma B cells suppressing memory B cells only in BCC (Fig. 3.4B), we increased the negative regulation of B cells by BCC cancer cells relative to that in melanoma.

To understand how checkpoint therapy would impact the dynamical model, we compared the anergy and activation scores of memory B cells before and after treatment in single-cell data (Fig. 3.2). We observed that in post-treatment, there was a strong B cell activation signature (presumably leading to a faster killing rate of cancer cells) and a simultaneous increase of the anergy signature in B cells. We therefore assumed that checkpoint therapy can either increase the killing rate, increase the cancer-mediated death rate of B cells or increase these rates in tandem (see red arrows in Fig. 3.5A).



---

Figure 3.5 (*preceding page*): **ICB-targeted parameters can explain R vs. NR in melanoma and BCC.** A) Schematic representation of the model (C=cancer, B=B cells, M=anti-inflammatory macrophages). Red arrows indicate processes assumed to be upregulated by ICB. Grey arrow is inferred from our single-cell analysis. B-D) Comparison of melanoma (left) and BCC (right) B) Three trajectories with varying initial cancer population for each of melanoma and BCC. Color of the trajectory corresponds to initial cancer population. All axes are log scale. C) Contour plots showing varying number of equilibria as the death rate of B cells ( $y$ -axis) and killing rate by B cells of the cancer ( $x$ -axis) are varied. The “Bifurcation Line” indicates the values for which the bifurcations in Panel D are plotted. D) Bifurcation diagram in the killing rate,  $k$ , corresponding to the Bifurcation Line in Panel C. Possible starting and ending values for a non-responder and a responder are shown. Each Cancer and Macrophage axis is shown over two different scales for visualization purposes.

### **3.3.6 Multi-stability stemming from immune cell interactions in melanoma and BCC can explain heterogeneous response to checkpoint therapy**

To understand the effects of immunotherapy on cancer burden, we analyzed the steady states of our model within a certain biologically relevant subset of parameter space (Fig. 3.5C). Overall, our system displays multi-stability, a common concept in cancer state modeling[62], in both melanoma and BCC where the system can evolve towards two or more steady states depending on the level of cancer burden. Bifurcations in our model result in large changes to cancer burden (none, low, high, or very high). When immunotherapy pushes a patient through one of these bifurcations where the cancer burden decreases, the patient would likely have an observable response. When immunotherapy does not reach a bifurcation, the patient is classified as non-responsive even though the new steady-state cancer burden may have decreased. Other interpretations to the bifurcations in our model are explored in Figs. 3.8-3.11, located in the supplementary section.

We found that under the assumption that checkpoint therapy increases the killing rate of B cells, there exists divergent responses to immunotherapy in both melanoma and BCC (Fig. 3.5D). We observe the possibility for a responder in both cancers: one going from a very high cancer burden to one with a low cancer burden. In melanoma, this responder also experiences an increase in B cells and a large decrease of macrophages. In BCC, we see both immune populations drop after treatment, though the macrophage population in this case only decreases by one or two cells. On the other hand, we see the possibility for a non-responder in both cancers, namely only a slight decrease in an already very high cancer burden accompanied by a similarly small decrease in B cell population and an increase in macrophage population, potentially up to several orders of magnitude in the melanoma case.

In our chosen parameter regime, the value of the killing rate at which a patient would

become a responder is nearly identical for both BCC and melanoma, indicating a similar level of requirement in the killing rate activation for treating both melanoma and BCC with checkpoint therapy (Fig. 3.5D). However, when the cancer-mediated death rate of B cells is decreased, BCC has a lower killing rate threshold separating non-responders from responders compared to melanoma, and vice versa when the death rate is increased (Fig. 3.14). The fact that the death rate changes which cancer is more likely to respond to checkpoint therapy indicates the ODE model is insufficient to ascertain which cancer will exhibit a better response to immunotherapy.

### **3.3.7 Noise-induced cancer progression and regression potentially account for therapy-resistance in BCC**

In the highly complex cancer-immune interacting environment[10], the fluctuations in cell populations may induce stochastic transitions among meta-stable states[63]. Because of this, we next incorporated stochastic effects into our three-component dynamical model. In our model, the inclusion of random fluctuations in cell population dynamics allows for the spontaneous transitions between cancer states with various burdens, contributing another source to affect the checkpoint therapy outcome by resulting in the spontaneous progression or regression of cancer[72, 125, 87]. The likelihood of these transitions can be determined by therapy-relevant parameters, allowing predictions of which cancer is more amenable to checkpoint therapy or yields a better clinical outcome in the long-term.

In order to compare the relative stability of cancer states perturbed by noise, we constructed a cancer-state landscape to visualize the global structures of attractor basins in melanoma and BCC population dynamics. The landscape express the probability the system is in a given state in the long-term with lower values indicating higher probabilities. It agrees well with our bifurcation analysis of the non-stochastic system by showing two connected

potential wells representing the cancer states with relatively low and high tumor burdens (Fig. 3.6A). The connectivity between these cancer potential wells suggests the possibility of spontaneous transitions in both cancer types, which corresponds to the phenomenon of tumor progression and regression. We also observed that the potential well of the low-burden state in BCC is shallower than in melanoma.

A unique feature of stochastic systems is the possibility to have the system transition between stable states. The specific transition path the system follows can discriminate between growing and regressing cancers and can differentiate the relative ease of transitioning one way or the other. To study these transition paths, we implemented the geometric minimal action method (gMAM) which determines the likeliest of such paths on an arbitrary timescale. When the cancer undergoes regression from the high cancer burden state to the low cancer burden state, the B cell population is transiently increased before settling in the low equilibrium. This is in contrast to the progression case where the B cell population is initially unresponsive to the nascent growth of the cancer (Fig. 3.6B). This difference in transition paths could not be discerned from the previous non-stochastic system. This supports the necessity of immunosuppressive actors in the tumor microenvironment for a cancer to escape the immune system. That is, the underlying ODE predicts that a positive perturbation in the cancer population will result in a positive rate of change to the B cell population. However, the stochastic analysis predicts that such a path is unlikely to result in a transition to a higher cancer burden.

To quantify how checkpoint therapy affects the likelihood of spontaneous tumor progression and regression, we calculated the change in barrier heights between the two cancer states as the killing rate is increased (Fig. 3.6C). We find that the barrier height for regression in BCC is generally larger than in melanoma with similar killing rates, indicating BCC patients may only respond to a larger dose of checkpoint therapy. In addition, the introduction of checkpoint therapy significantly increases the barrier height for tumor progression in melanoma.

Counter-intuitively, it is very interesting to note that a slight increase in killing rate will instead reduce the tumor progression barrier height in BCC, increasing the possibility of immune escape by the cancer (Fig. 3.6C). This may provide a potential explanation for the unsatisfactory outcome of checkpoint therapy in BCC[87, 156, 125, 64, 38].

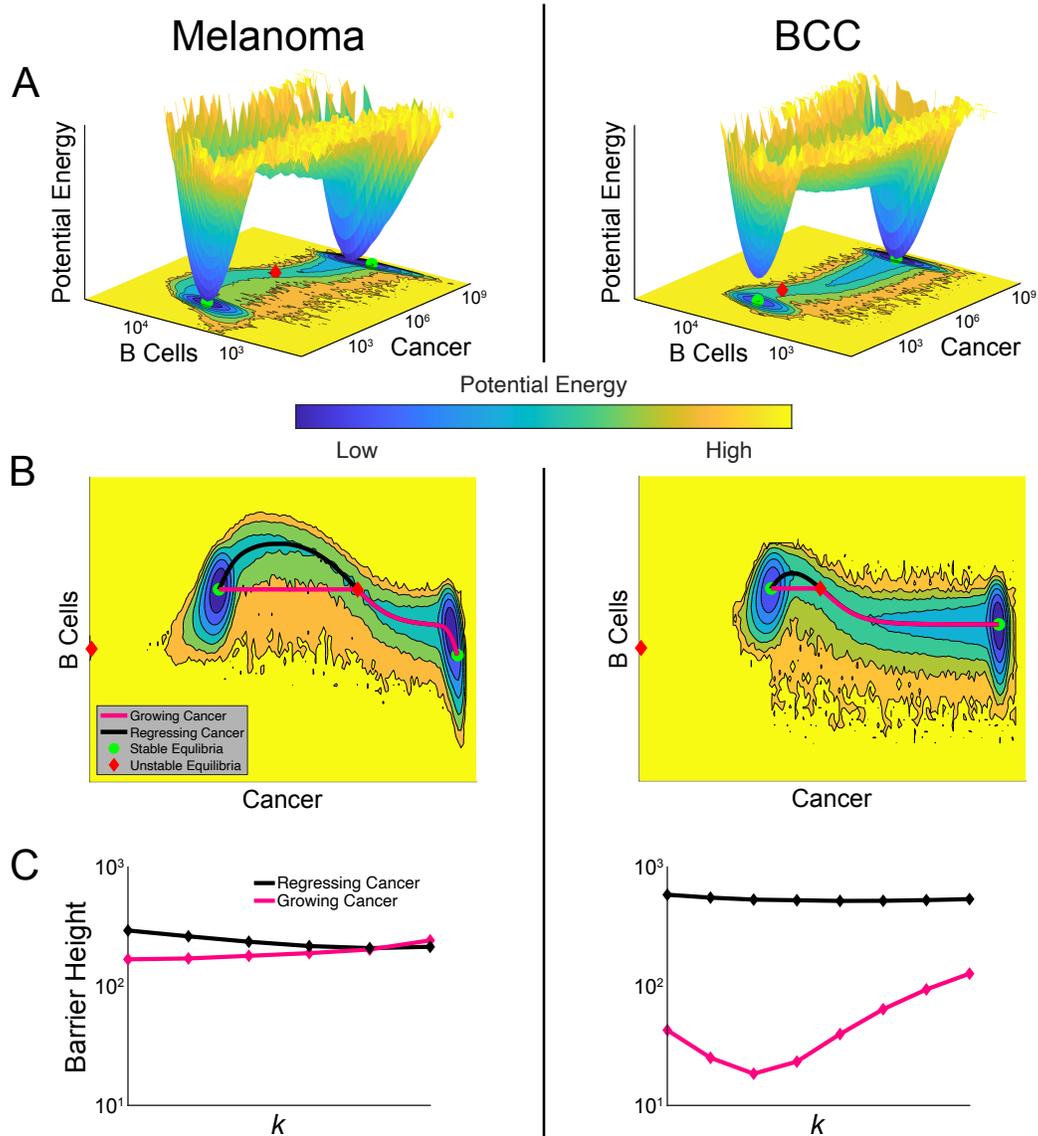


Figure 3.6: **Excess B cell counts predicts regressing cancer. Melanoma and BCC show different patterns for likelihood of progression.** A-C) Comparison of melanoma (left) and BCC (right) as modeled by our SDE. A) Energy landscape of both cancers with  $k = 10^{-4}$  and  $d_e = 1$ . Lower values indicate a higher probability of finding the system in that state. These values are the marginal probabilities having marginalized over macrophages. A contour plot is shown below along with the ODE-determined equilibria, both stable (green dots) and unstable (red diamonds). B) The transition paths between the stable equilibria plotted over the contour plots from Panel A. The black path is for a regressing cancer and the magenta path is for a growing cancer. C) The barrier height between the two stable equilibria as it varies with the killing rate,  $k$ . The black curve shows the variation of the barrier height for a regressing cancer, and the magenta for a growing cancer.

## 3.4 Methods

### 3.4.1 The three-component dynamical model

Based on the single-cell data analysis, we modeled the dynamics of memory B-cells, macrophages and cancer cells populations, which are emergent from their complex interactions. The assumptions on interactions between B cell and cancer cells are derived from existing literatures (Supplementary Materials) with B cells interacting with the cancer in ways similar to T cells. In this way we extend previous work. On the other hand, the inclusion of anti-inflammatory macrophages (“macrophages”) and their interactions with other cells constitutes the novel aspect of our work, as most previous work uses pro-inflammatory cells as a third state variable.

Whereas most models that add a third state variable to their cancer-effector cell model choose to include a second anti-tumor population (e.g. [30]), we added a pro-tumoral immune population. Derived from the single-cell data analysis, the macrophages act to down-regulate B cell proliferation, directly in opposition to the cancer-mediated upregulation of that very process. The macrophages are in turn influenced by the cancer and memory B cells by responding to the apoptotic signals from cancer cells as the memory B cells kill them[99].

With regard to plasma B cells, we chose not to include them for two reasons. First, their negative regulation of B cell proliferation is already effected by cancer cells. Second, we assumed that the suppressive nature of these plasma B cells are downstream of cancer signaling, so that a component of the cancer-mediated down-regulation of B cell proliferation can be viewed as the effect of plasma B cells already. Thus, we found it more desirable to keep a simpler model with which our analysis would be clearer rather than include a new state variable for plasma B cells.

Taken together, the model can be expressed in ordinary differential equations (ODE):

$$\frac{dC}{dt} = aC(1 - bC) - kCB \quad (3.1)$$

$$\frac{dB}{dt} = s - dB + b_e \frac{C}{\kappa_e + C} \frac{\kappa_m}{\kappa_m + M} B - d_e \frac{C}{\kappa_d + C} B \quad (3.2)$$

$$\frac{dM}{dt} = g - d_m M + p \frac{kCB}{\kappa_a + kCB} M \quad (3.3)$$

We conducted the non-dimensionalization to simplify our analysis (Supplementary Material). To perform equilibria and stability analysis, we solved the derived 5-degree polynomial of steady-state equation and determine the stability using the eigenvalues of the Jacobian (Supplementary Material). The bifurcation plot can be generated by tracking the change of equilibria with respect to one interested parameter.

To consider transitions among meta-stable cancer states, we included a time-independent noise term,  $\sigma(X_t)$ , and come up with a stochastic differential equation (SDE) model

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad (3.4)$$

with  $b(X_t)$  corresponding to the drift term in the ODE, and  $W_t$  being a standard Weiner process.

### 3.4.2 Energy landscape and transition paths

The cancer-state energy landscape can quantify the relative stability of different meta-stable states perturbed by noise, closely relevant to the notion of energy landscape, a mathematical realization of Waddington's epigenetics metaphor[80, 161, 149]. To generate the landscapes, we simulated a large number of trajectories with randomly chosen initial conditions. Initial conditions were uniformly distributed over the log scale of the state variables. Each

subsequent time step was binned based on the 3D coordinates and used to compute the probability a trajectory was in a particular bin. To arrive at the landscapes, we took the marginal probabilities over a given state variable and then computed the negative logarithm to arrive at our potential landscape.

To compute transition paths among meta-stable states, we applied the Freidlin and Wentzell's (FW) large deviation theory[41], which states that under a small noise assumption, the most probable path,  $\varphi^*(s)$ , transitioning from state  $x_1$  to  $x_2$  corresponds to the minimizer of the action functional:

$$S(\varphi) = \int_0^T (\varphi' - b(\varphi))^t D^{-1}(\varphi) (\varphi' - b(\varphi)) ds \quad (3.5)$$

where  $D$  is the diffusion tensor,  $D(x) = \sigma(x)\sigma^t(x)$  and

$$\varphi^* = \inf_{T>0} \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0, T)} S[\varphi] \quad (3.6)$$

We set  $x_1$  and  $x_2$  as stable fixed points of the ODE system. To tackle the numerical challenges introduced by critical points[49], we implemented a simplified geometric minimal action method (sgMAM) to solve the optimization problem[49]. We used the action functional for these paths to compute the barrier heights between stable equilibria. According to FW theory[41], the larger barrier height indicates the longer mean transition time between metastable states.

### 3.5 Discussion

Most modeling of cancer-immune interactions has focused on the role of T cells, and cytotoxic T lymphocytes in particular. As recent studies have shown, however, the presence of an active

B cell population in the tumor microenvironment can hold more predictive power of response to checkpoint therapy and so modeling such interactions has become an important item of study. Our initial work in this direction builds off of existing models of T cells interacting with cancer and extends them to include macrophages, which our data analysis has shown to be implicated in modulating cancer-B cell interactions.

We find that our model qualitatively agrees with our single-cell analysis as it pertains to divergent responses in both melanoma and BCC for responder and non-responders. This serves as an explanation for why some patients appear to have a naturally acquired resistance to immunotherapy. We also find that the model predicts important differences between melanoma and BCC patients. While melanoma responds to immunotherapy by having an increased likelihood to spontaneously regress, BCC actually displays therapy-induced resistance as the likelihood of spontaneous regression can in fact decrease on therapy.

### **3.5.1 Experimental and clinical implications of the model**

While altering many of the processes accounted for in our model can lead to a decrease in maximum tumor burden, only a subset leads to a reduction to a low or cancer free state (Figs. 3.12 and 3.13). In addition to the killing rate, these processes include the death and source rates of B cells, the cancer-mediated B cell proliferation rate, and the ability of cancer to increase B cell death (Supplementary Material). Targeting these pathways could prove promising in changing the tumor environment from tumor-supportive to tumor-rejecting.

Second, we see that the transition paths diverge and so the direction of change can be inferred from a snapshot of the state of the tumor microenvironment rather than requiring multiple samples to determine the direction of growth.

Third, as mentioned above, we see that the low cancer burden state in BCC is less stable

than its counterpart in melanoma. We see this in the fact that the low cancer burden well is shallower than the high cancer burden well in BCC only, suggesting that the biological noise inherent in the TME could more readily lead to progression than in the melanoma case.

We also see this in looking at the barrier height sensitivity to the killing rate, observing two differences between BCC and melanoma: 1) We see that the barrier height in BCC for progressing cancer remains well below that of regressing cancer, indicating checkpoint therapy is much more likely to be effective for a melanoma patient than a BCC patient. 2) When the killing rate is low, an increase in this parameter leads to a more likely progression of BCC indicating that an insufficiently strong ICB could have adverse effects on a BCC patient.

## **3.6 Supplementary Information**

### **3.6.1 Literature-based assumptions of the dynamical model**

In the three-component model (Fig. 3.5A), we follow a large body of theory describing cancer-immune interactions as a deterministic ODE[34]. Cancer cells grow logistically with a carrying capacity on the order of one billion cells and the immune system is the only force keeping it below this value. As recent studies suggest and our analysis affirms, this anti-tumor immune response is determined by the B cell response, specifically the memory B cell response[15, 56, 111]. That is, the memory B cell state variable is the sole immune component responsible for killing cancer cells in our model.

We also include the reciprocal effects of cancer on the pro-inflammatory immune system. First, the cancer can upregulate B cell proliferation as has been evidenced by the presence of tertiary lymphoid structures in tumors and the increased proliferation of B cells in such

areas[56, 111, 15]. This also allows for another similarity between this model and other cancer-immune models, even though these have typically looked at T cells[74, 35]. Second, the cancer can simultaneously exert negative regulation on the B cells via upregulating their removal from the system. This corresponds with the cancer shaping the TME towards an immunosuppressive phenotype resulting in poorly activated memory B cells[128] and immunosuppressive plasma B cells (Fig. 3.4).

### 3.6.2 Non-dimensionalization of the dynamical model and parameter selection

We non-dimensionalize our system to simplify the analysis, but present our results in terms of these original equations. These are the non-dimensionalized equations:

$$\frac{dx}{d\tau} = x(1 - x) - xy \tag{3.7}$$

$$\frac{dy}{d\tau} = \alpha - \beta y + \gamma \frac{x}{\delta + x} \frac{1}{1 + z} y - \varepsilon \frac{x}{\zeta + x} y \tag{3.8}$$

$$\frac{dz}{d\tau} = \eta - \nu z + \theta \frac{xy}{\lambda + \mu xy} z \tag{3.9}$$

Note that the time variable has been changed from  $t$  in the original equations to  $\tau$  here. Parameter values can be found in Table 3.1. In Table 3.2, we define the new parameters and state variables in terms of the original parameters and state variables.

Name	Description	Value		Units	Source
		Melanoma	BCC		
$a$	maximum proliferation rate of tumor cells	0.514		days <sup>-1</sup>	[112]
$b$	inverse carrying capacity of tumor	$1.02 \times 10^{-9}$		cells <sup>-1</sup>	[112]
$b_e$	maximum memory B cell proliferation rate	1.5		days <sup>-1</sup>	[22, 26]

$d$	death rate of memory B cells	2		days <sup>-1</sup>	[22]
$d_e$	maximum rate of cancer-mediated deactivation of memory B cells	1		days <sup>-1</sup>	Varied as bifurcation parameter
$d_m$	death rate of macrophages	3		days <sup>-1</sup>	Estimated from [151]
$g$	source rate of macrophages	30		cells · days <sup>-1</sup>	Estimated from [151]
$k$	memory B cell killing rate	10 <sup>-4</sup>		cells <sup>-1</sup> · days <sup>-1</sup>	[36, 45, 112, 155]; Varied as bifurcation parameter
$\kappa_a$	EC50 for apoptotic-signaling-induced proliferation of macrophages	4.2 × 10 <sup>7</sup>	10 <sup>9</sup>	cells · days <sup>-1</sup>	[99]; Our analysis
$\kappa_d$	EC50 for cancer-mediated upregulation of memory B cell deactivation	2.5 × 10 <sup>6</sup>	2.5 × 10 <sup>4</sup>	cells	Our analysis
$\kappa_e$	EC50 for cancer-mediated upregulation of memory B cell proliferation	500		cells	Estimated; [15, 56, 111]
$\kappa_m$	EC50 for macrophage-mediated downregulation of memory B cell proliferation	5 × 10 <sup>3</sup>		cells	Our analysis
$p$	maximum rate of apoptotic-signaling-induced proliferation of macrophages	4		days <sup>-1</sup>	Our analysis
$s$	source rate of memory B cells	5 × 10 <sup>3</sup>		cells · days <sup>-1</sup>	[2, 28]

Table 3.1: **Mathematical model parameters.** Some parameter values for memory B cells ( $s$ ,  $d$ , and  $b_e$ ) are chosen from literature on CTLs. This repurposing is in line with understanding memory B cells as working in tandem with CTLs via antibody production. The value of  $\kappa_e$  is estimated based on the effects of the process suggested by the given references. These show that even small tumors can stimulate B cell activation and proliferation in tertiary lymphoid structures (TLSs) located at the tumor site. The macrophage parameters  $g$  and  $d_m$  were estimated from [151].

$x = bC$	$y = \frac{k}{a}B$	$z = \frac{M}{\kappa_m}$
$\tau = at$	$\alpha = \frac{ks}{a^2}$	$\beta = \frac{d}{a}$
$\gamma = \frac{b_e}{a}$	$\delta = b\kappa_e$	$\varepsilon = \frac{d_e}{a}$
$\zeta = b\kappa_d$	$\eta = \frac{g}{a\kappa_m}$	$\theta = p$
$\lambda = b\kappa_a$	$\mu = a$	$\nu = \frac{d_m}{a}$

Table 3.2: Defining relations of non-dimensionalized ODE.

### 3.6.3 Equilibria and their stability of deterministic model

In solving for equilibria, we are able to simplify the set of equations into a union of two solution sets:

$$C = 0 \quad \text{or} \quad f(C) = 0$$

where  $f$  is a degree 5 polynomial. Of the six roots, we select only the sensible ones, i.e. those where  $(C, B, M)$  lies in the closed first octant of  $\mathbb{R}^3$ . We determine the stability of these fixed points using the eigenvalues of the Jacobian, looking for those with all eigenvalues having negative real part. In Figure 3.5B, we show the number of stable equilibria in the  $k - d_e$  plane. We choose to classify the stable fixed points based on the size of the cancer population. When  $C = 0$  is stable, this is complete elimination. When  $C$  is nonzero but several orders of magnitude smaller than its carrying capacity (usually around  $10^3$  cells), we describe this as a low cancer burden. When  $C$  is at least 1% of the carrying capacity (so  $\geq 10^7$  cells), we call this a high cancer state. When two such states are stable, we call the one with the larger cancer population very high. Usually, a high cancer burden is around  $10^8$  cells whereas a very high cancer burden is closer to  $10^9$  cells.

### 3.6.4 Adding stochastic effect to the model

We next turned to the reality of biological noise and considered how this could impact our model. Let  $X_t$  be the state vector of our system at time  $t$  and let  $b(X_t)$  be our time-independent ODE function. We take a generic, time-independent noise term,  $\sigma(X_t)$  and come up with the following SDE model:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad (3.10)$$

with  $W_t$  being a standard Weiner process. For our noise term,  $\sigma$ , we assume there are both additive and multiplicative sources of noise for each population. These are given by independent Weiner processes and so  $\sigma$  is a 3x6 matrix. Each row corresponds to a state variable. Each row will have exactly two nonzero entries: one constant providing the additive noise and one a linear form in that state variable providing the multiplicative noise. No column has two nonzero entries as we assume that any source of noise is only effect a single population. For determining the parameters of these functions, we chose them to be proportional to the parameters of our ODE (Table 3.3). For B cells and macrophages, these choices were tied to their source rates (additive coefficient) and proliferation rates (multiplicative coefficient). The energy landscapes shown in Figure 3.6A scaled these noise terms by  $1/a$ .

State Variable	Additive Proportionality Constant	Multiplicative Proportionality Constant
$C$	$100a$	$0.1a$
$B$	$0.01s$	$0.05b_e$
$M$	$0.01g$	$0.05p$

Table 3.3: **Noise parameters. With these parameters, the matrix  $\sigma$  is given by**

$$\sigma(C, B, M) = \begin{bmatrix} 100a & 0 & 0 & 0.1a \cdot C & 0 & 0 \\ 0 & 0.01s & 0 & 0 & 0.05b_e \cdot B & 0 \\ 0 & 0 & 0.01g & 0 & 0 & 0.05p \cdot M \end{bmatrix}$$

### 3.6.5 Model simulation methods and analysis

To analyze the ODE, we used MATLAB. For computing trajectories, we used the *ode45* function with nonnegativity constraints. To compute equilibria, we reduced the defining equation to one of two conditions: zero cancer or a degree five polynomial equation in cancer. Utilizing the Symbolic Toolbox in MATLAB, we could write down this polynomial explicitly as a function of the model parameters and then use the *roots* function to solve for the zeros. Selecting for the sensible equilibria—all state variables real and nonnegative—we got all the equilibria. By computing the Jacobian at the equilibria, we could then find which had eigenvalues with positive real part and label these as unstable equilibria while the rest were labeled as stable.

Having chosen a 3x3 grid in the  $k - d_e$  plane to explore further, we looked at how the cancer burden changed when perturbed from the center of this grid. Specifically, all parameters would be fixed at that grid point, except one would be varied and the stable equilibria with the maximal cancer size would be computed.

To analyze the SDE, we also used MATLAB and our own implementation of the Euler-Maruyama (EM) method. Since we desired to impose the condition that our state variables remained nonnegative, we implemented the following algorithm whenever an update step resulted in a negative component:

1. If any of the state variables that are negative after the update were larger than one before the update and if the time step was larger than a pre-specified tolerance, go on to Step 2. Otherwise, set the negative components to zero and continue solving.
2. Subdivide the interval into two equal time steps. Use two new vectors of random numbers satisfying conditions guaranteeing they add up to the original vector and have appropriately decreased variance.

Two comments on the above steps are in order. First, the time step tolerance guarantees that this process must terminate at some point since additive noise could force the system to a negative state. Second, the phrase “appropriately decreased variance” can be made explicit as follows. Let  $x$  and  $y$  be two independent random variables, normally distributed with mean zero and standard deviation  $\sigma_{dt/2} = \sqrt{dt/2}$  each representing one update of the Wiener process over half the time interval that resulted in a negative state variable. Letting  $z$  be the originally selected update that resulted in the negative state variable, we impose the condition  $x + y = z$  on the joint probability distribution of  $x$  and  $y$ . Computing the probability density function for  $x$  given this constraint, we arrive at

$$f(x) = f(x, y | x + y = z) \tag{3.11}$$

$$= K \frac{\exp\left(-\frac{1}{2}(x, z-x)^T \Sigma^{-1} (x, z-x)\right)}{2\pi \det \Sigma} \tag{3.12}$$

$$= K \frac{\exp\left(-\frac{1}{dt}(x^2 + (z-x)^2)\right)}{\pi dt^2/2} \tag{3.13}$$

$$= K \frac{\exp\left(-\frac{2}{dt}(x^2 - xz + z^2/2)\right)}{\pi dt^2/2} \tag{3.14}$$

$$= K \frac{\exp\left(-\frac{2}{dt}\left((x-z/2)^2 + z^2/4\right)\right)}{\pi dt^2/2} \tag{3.15}$$

$$= \tilde{K} \exp\left(-\frac{1}{2} \left(\frac{x-z/2}{\sqrt{dt/2}}\right)^2\right) \tag{3.16}$$

where  $K$  and  $\tilde{K}$  are constants of integration. From Equation 3.16, we see that we must modify the expected standard deviation,  $\sqrt{dt/2}$ , by dividing it by  $\sqrt{2}$  to account for the knowledge that the two must sum to  $z$ . This process would apply to each of the six randomly chosen numbers needed for one update to our EM method.

### 3.6.6 The simplified gMAM algorithm for transition path

To compute transition paths, we implemented a simplified geometric minimal action method (sgMAM)[49]. Unlike the original gMAM[58], the sgMAM takes the following formulation of the left invariant action functional on the space of curves,  $\hat{S}[\varphi]$ :

$$\hat{S}[\varphi] = \sup_{\vartheta: H(\varphi, \vartheta)=0} \int_0^1 \langle \varphi', \vartheta \rangle ds \quad (3.17)$$

Let  $E(\varphi, \vartheta) = \int_0^1 \langle \varphi, \vartheta \rangle ds$ ,  $E_*(\varphi) = \sup_{\vartheta: H(\varphi, \vartheta)=0} E(\varphi, \vartheta)$ , and  $\vartheta_*$  be the maximizer such that  $E_*(\varphi) = E(\varphi, \vartheta_*(\varphi))$ . It follows that  $\vartheta_*$  satisfies the Euler-Lagrange equation associated with the constrained optimization problem defined by  $E_*$ :

$$D_{\vartheta}E(\varphi, \vartheta_*) = \mu H_{\vartheta}(\varphi, \vartheta_*) \quad (3.18)$$

Where  $H$  is the associated Hamiltonian and  $\mu(s)$  is the Langrange multiplier used to enforce the constraint  $H(\varphi, \vartheta_*) = 0$  so that when  $\vartheta = \vartheta_*$  we have

$$\mu = \frac{\|D_{\vartheta}E\|^2}{\langle D_{\vartheta}E, H_{\vartheta} \rangle} = \frac{\|\varphi'\|^2}{\langle \varphi', H_{\vartheta} \rangle} \quad (3.19)$$

With the inner product and norm being the one induced by the diffusion tensor,  $D(x) = \sigma(x)\sigma(x)^T$ . We seek the minimizer of  $E_*$ ,  $\varphi_*$ , and so we compute the gradient (see [49] for details):

$$D_{\varphi}E_*(\varphi_*) = -\vartheta'_* - \mu H_{\varphi}(\varphi_*, \vartheta_*) \quad (3.20)$$

As this gradient can be ill-behaved around critical points where  $\varphi' = 0$ , we use  $\mu^{-1}$  as a pre-conditioner to ensure convergence.

Thus, the sgMAM can be broken down into two tasks:

1. For a given  $\varphi$ , find  $\vartheta_*(\varphi)$  by solving

$$\vartheta_*(\varphi) = \arg \max_{\vartheta: H(\varphi, \vartheta)=0} E(\varphi, \vartheta) \quad (3.21)$$

which is equivalent to solving

$$D_{\vartheta}E(\varphi, \vartheta_*) = \varphi' = \mu H_{\vartheta}(\varphi, \vartheta_*) \quad (3.22)$$

under the constraint  $H(\varphi, \vartheta_*) = 0$ .

2. Find  $\varphi_*$  by solving the optimization problem

$$\varphi_* = \arg \min_{\varphi \in \bar{C}_x^y(0,1)} E_*(\varphi) \quad (3.23)$$

by pre-conditioned gradient descent using as direction

$$\mu^{-1} \vartheta'_*(\varphi) + H_{\varphi}(\varphi, \vartheta_*(\varphi)) \quad (3.24)$$

while maintaining the constraint  $|\varphi'| = \text{const.}$

### 3.6.7 Other interpretations of immunotherapy

Our analysis of the data indicates that immunotherapy can both increase activation and energy. In the main text, we considered how increasing the killing rate (activation) of memory B cells can change the arrangement of stable steady states. Here, we consider what happens as the cancer-induced death of memory B cells varies as well as when both these two parameters vary.

First, by increasing the cancer-mediated death rate of memory B cells,  $d_e$ , we expect the cancer burden to increase. Thus, considering this in isolation, we are considering the worst case scenario our model predicts for immunotherapy. For both melanoma and BCC, there are bifurcations resulting in an increased cancer burden with a corresponding increase in macrophages and decrease in memory B cells (Fig. 3.7).

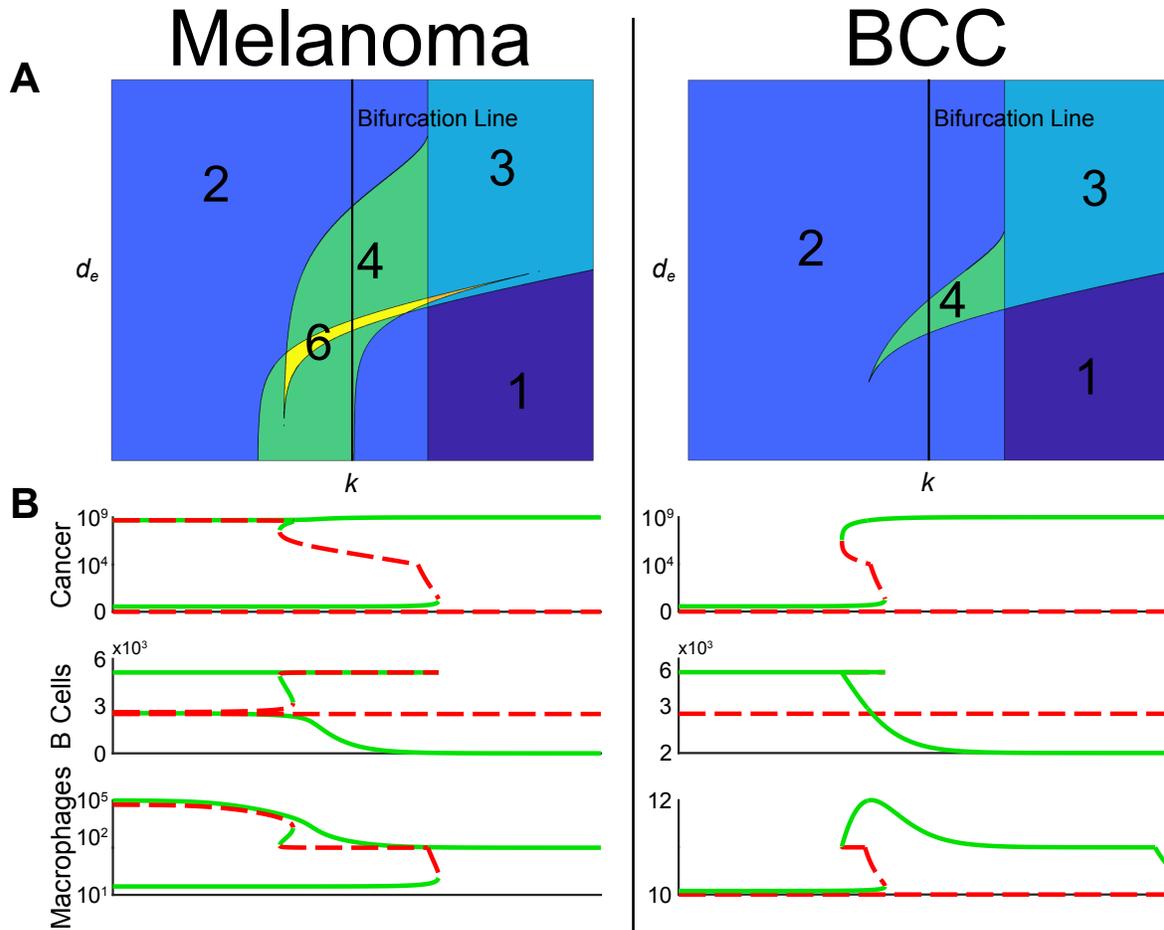


Figure 3.7: **Bifurcation in cancer-induced death rate, comparing melanoma and BCC.** A) Contour plots showing varying number of equilibria as the death rate of B cells ( $y$ -axis) and killing rate by B cells of the cancer ( $x$ -axis) are varied. The “Bifurcation Line” indicates the values for which the bifurcations in Panel B are plotted. B) Bifurcation diagram in the killing rate,  $d_e$ , corresponding to the Bifurcation Line in Panel A. Each Cancer and Macrophage axis is shown over two different scales for visualization purposes.

Second, we consider increasing both the activation ( $k$ ) and anergy ( $d_e$ ) scores simultaneously. We choose to do so by scaling  $k$  by some quantity  $c$  and  $d_e$  by  $c^m$  for a fixed value  $m$ . Below,

we show the result for two possible values of  $m$ : 0.5 and 2. When  $m = 0.5$ , this would mean that the negative effects of therapy (increased B cell death rate) are less pronounced than the positive effects. When  $m = 2$ , the opposite is true: the negative effects grow faster than the positive effects. In all plots, therapy results in an increase along the  $y$ -axis.

In both melanoma (Fig. 3.8) and BCC (Fig. 3.9), there are only favorable bifurcations when  $m = 0.5$  with responders showing all possible changes of cancer burden. This would indicate that the stronger the therapy, the better the response, similar to what was seen in just varying  $k$ .

In sharp contrast, when  $m = 2$  (Fig. 3.10 and Fig. 3.11), there are regions where immunotherapy can have varied effects. In the top row of both, therapy can create a stable, low cancer burden state and then a zero cancer burden that is stable. In melanoma, rows 2 and 3 have a high cancer burden become stable in a small window, allowing for a partial, yet meaningful response. Then, in both melanoma and BCC, a particular set of bifurcation lines show a clear “sweet spot” where therapy results in total clearance of all cancer cells, but going beyond this reintroduces a high cancer burden. Should such a patient exist, checkpoint blockade could well indeed be effective, but too strong a dosing regimen could result in little change in the cancer burden. Finally, the bottom panels of both show patients with similar responses to varying  $k$  alone: sufficient ICB results in a complete response.

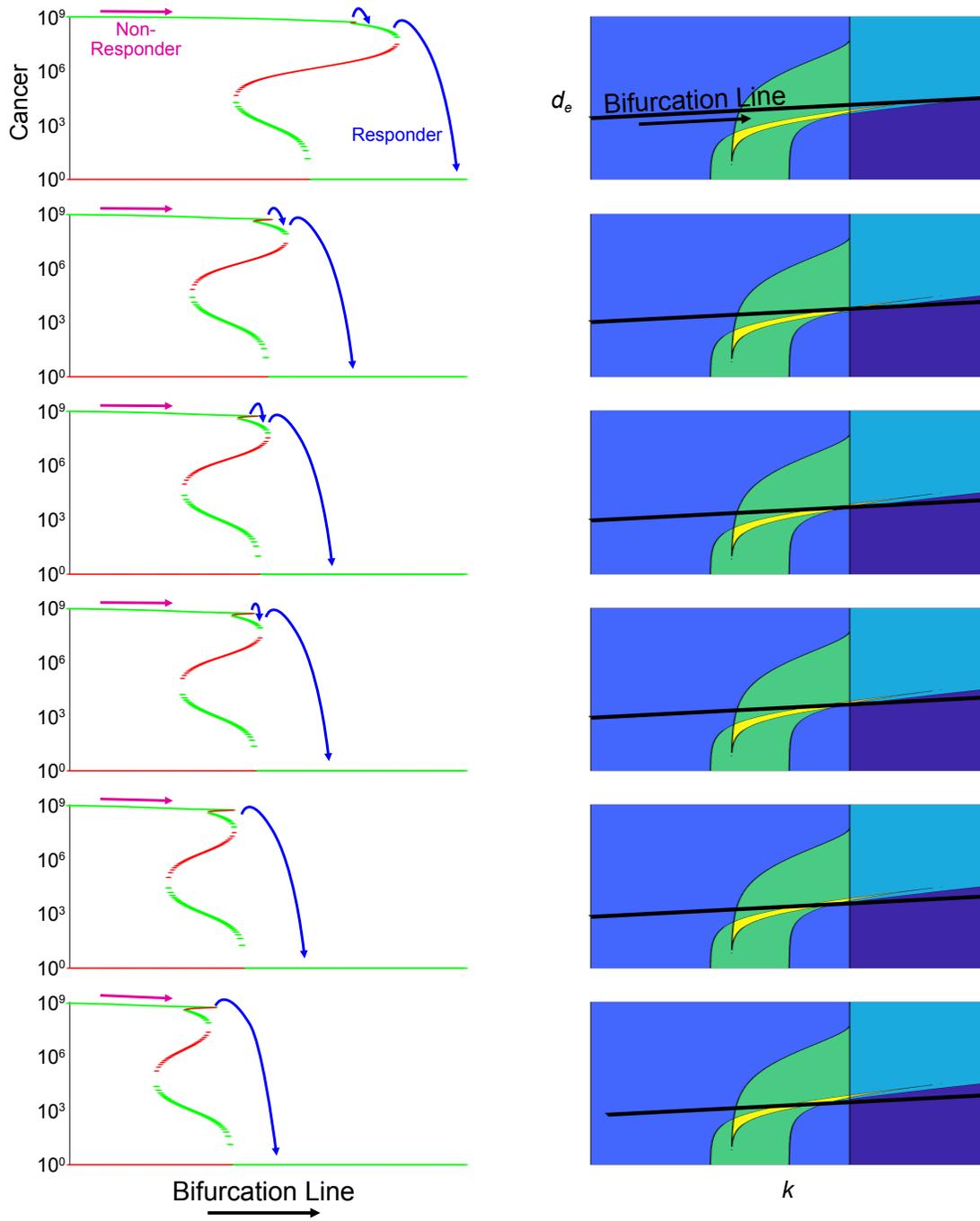


Figure 3.8: **Bifurcation for melanoma varying both  $k$  and  $d_e$  with  $m = 0.5$ .** Left panels show bifurcation in cancer population only. Green dots indicate stable equilibria and red dots unstable equilibria. Possible non-responders shown by pink arrows. Possible responders shown by blue arrows. Right panels show the path through parameter space represented along  $x$ -axis of left panels. In some panels, the entire line segment is not shown.

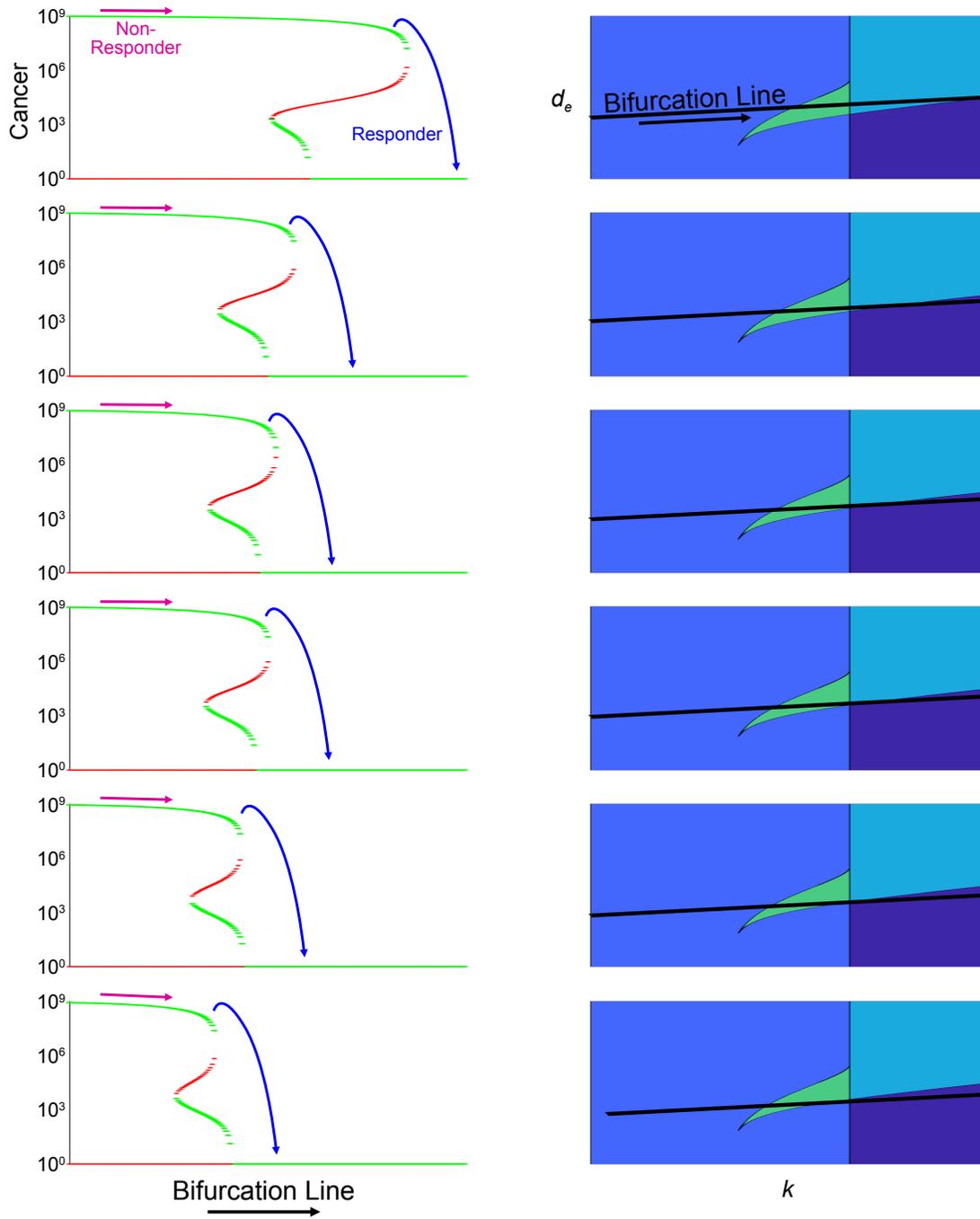


Figure 3.9: **Bifurcation for BCC varying both  $k$  and  $d_e$  with  $m = 0.5$ .** See caption of Figure 3.8 for details.

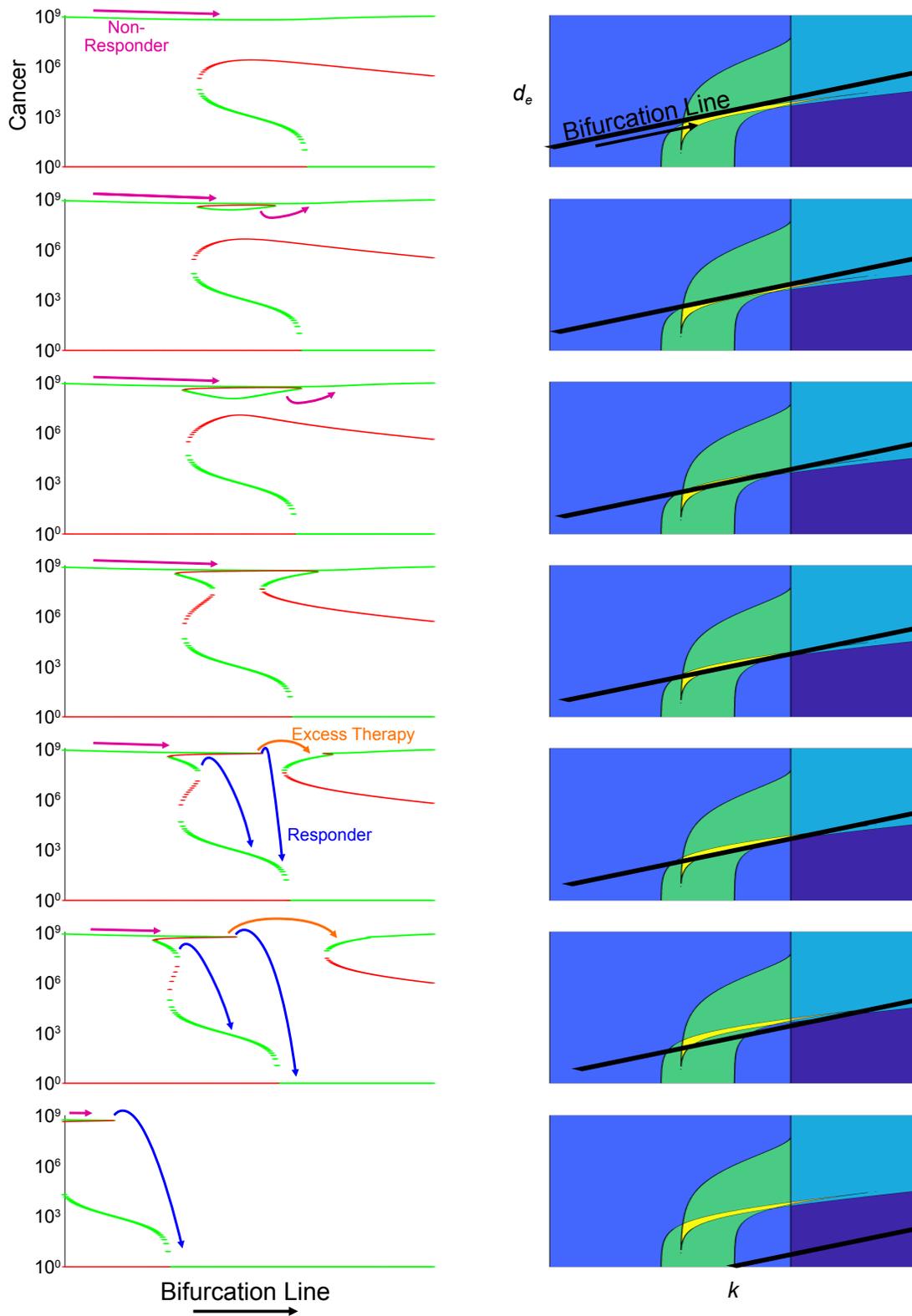


Figure 3.10: **Bifurcation for melanoma** varying both  $k$  and  $d_e$  with  $m = 2$ . See caption of Figure 3.8 for details. Orange arrows so the possibility of patients receiving too much therapy and thus not responding.

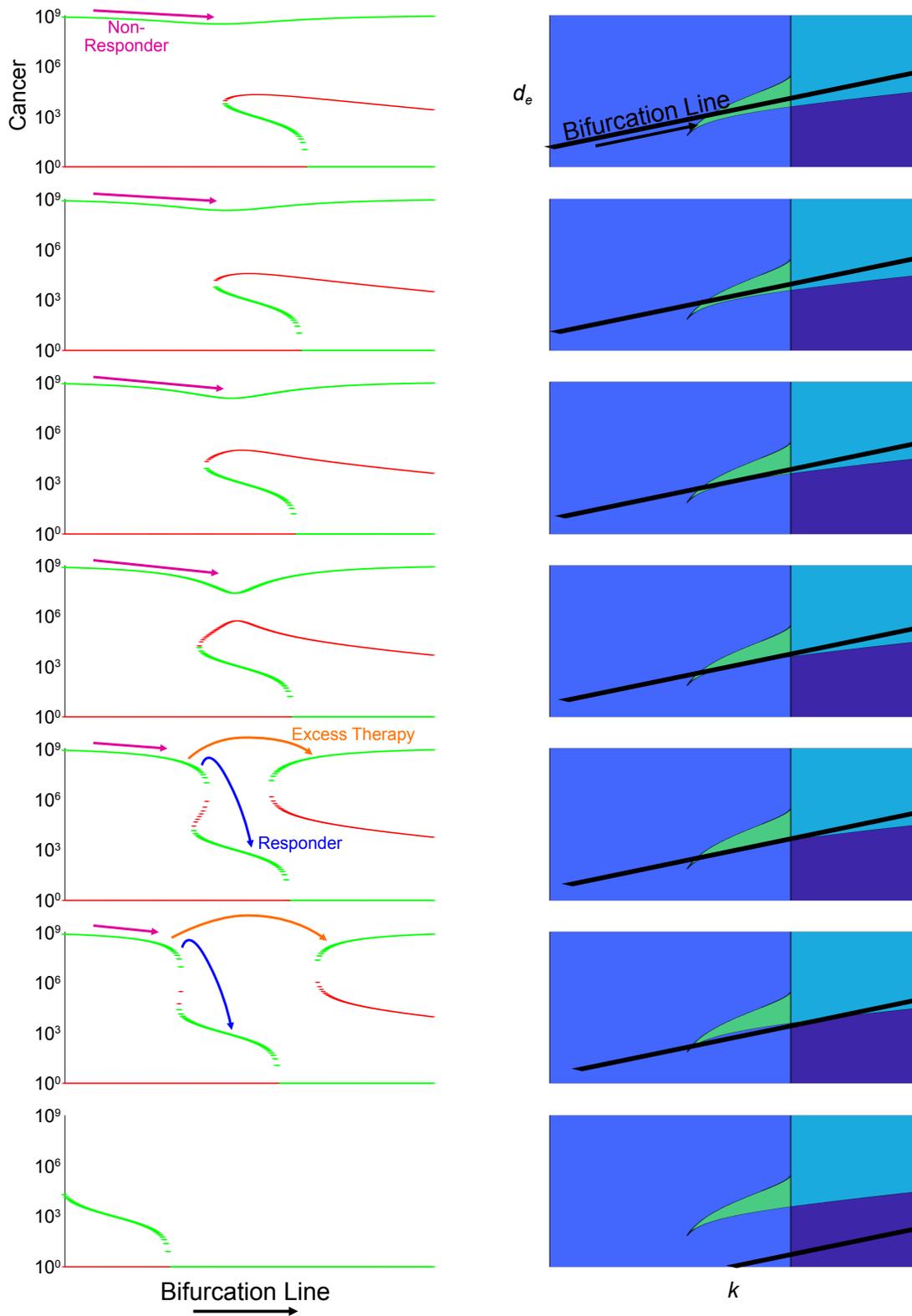


Figure 3.11: **Bifurcation for BCC varying both  $k$  and  $d_e$  with  $m = 2$ .** See caption of Figure 3.8 for details. Orange arrows so the possibility of patients receiving too much therapy and thus not responding.

### 3.6.8 Pathways to target

We have focused to this point on only two biological processes: the B cell-mediated killing of cancer cells and the cancer-induced death of B cells. However, our model also predicts other pathways can lead to strong responses to checkpoint therapy, where we take strong to mean the cancer moves from high or very high to low or none. The biological processes (and associated parameters) that can lead to such a response in both melanoma and BCC are as follows (Figs. 3.12 and 3.13): cancer proliferation ( $a$ ), cancer carrying capacity ( $b$ ), cancer-induced proliferation of B cells ( $b_e$ ), the death rate of B cells ( $d$ ), the killing rate ( $k$ ), and the natural source rate of B cells ( $s$ ).

Of these, the cancer-induced proliferation is the most intriguing. The activation score for B cells could be understood as an increase in their proliferative capacity as induced by recognition of the cancer. We chose to instead focus on the killing rate in our main analysis because varying  $d_e$  and  $b_e$  are explicitly opposing forces, making jointly varying them approximately like varying a single parameter.

### 3.6.9 Comparison of melanoma and BCC sensitivity to immunotherapy

From the ODE model, we can analyze the  $k$  value at which the largest predicted response to immunotherapy will occur. We can use this as a guide to which cancer will be more sensitive to immunotherapy. However, depending on the value of  $d_e$ , one or the other will come out more sensitive by this interpretation (Fig. 3.14). Here, we plot in state space the  $k$  value at which the largest drop happens for values of  $d_e$  small enough. Around  $d_e = 1$ , the two curves cross and the cancers switch which one is more sensitive. At higher values, melanoma is more amenable to checkpoint blockade, but BCC is more so at low values of  $d_e$ .

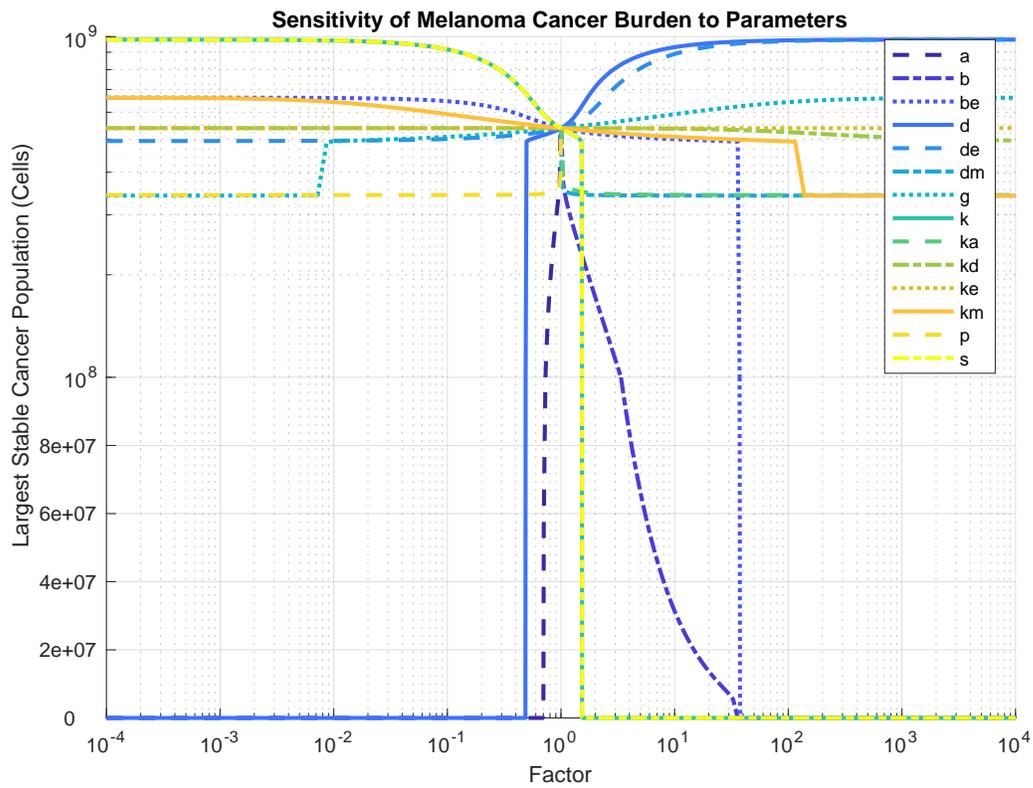


Figure 3.12: **Sensitivity of melanoma cancer burden to all parameters.** All parameters are varied by scaling them from their base initial value (Factor=1). The largest stable cancer burden is plotted.

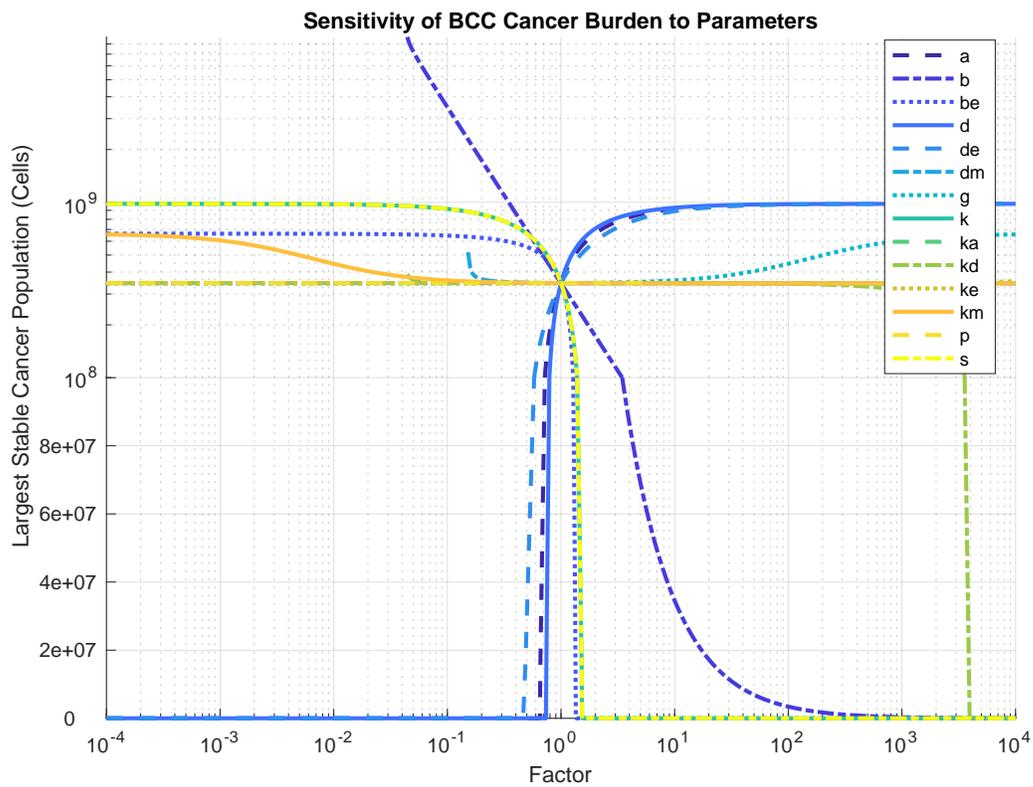


Figure 3.13: **Sensitivity of BCC cancer burden to all parameters.** All parameters are varied by scaling them from their base initial value (Factor=1). The largest stable cancer burden is plotted.

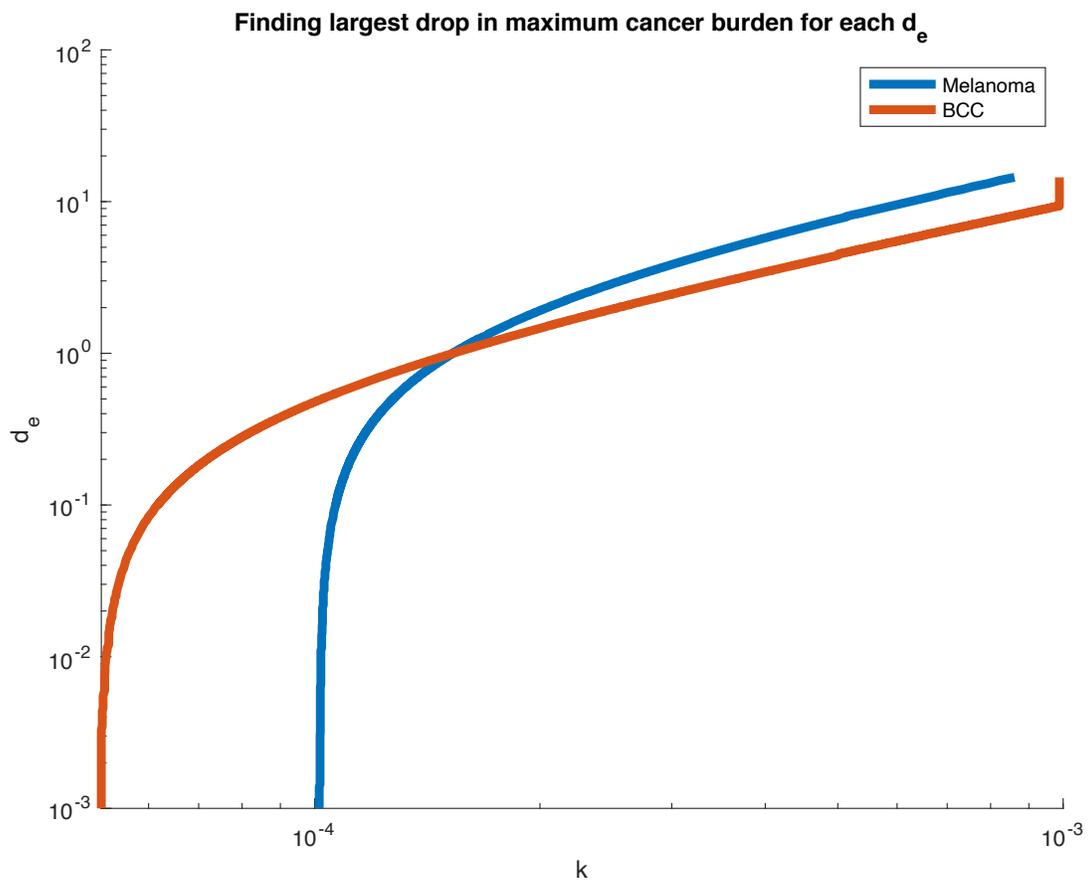


Figure 3.14: Comparison of melanoma and BCC sensitivity to immunotherapy.

# Chapter 4

## Parameter Inference

This chapter presents preliminary results using the framework of Bayesian inference to infer parameter values for a mathematical model. We build an ODE model and use published data to infer parameter values.

### 4.1 Summary

This chapter covers a Bayesian technique for parameter inference. A simple example is given of inferring parameter values ODE model from time series data.

### 4.2 Introduction

To make the most of modeling efforts and make them applicable to actual clinical scenarios, we often need to ensure some proximity to actually recorded data. Parameter inference is the process of using known data to inform the parameters of a mathematical model and thus

meet this goal. The standard method is to create a function whose input is a specific set of parameter values, computes an approximation to the data, and outputs how different this approximation is from the real thing. In the case of a discrete set of data points, the  $\ell^2$ -norm is common for this final step. If we call the function computing the approximation from the parameter values  $f$  and we call the data  $\hat{y}$ , then we are considering the following:

$$\text{parameters} \xrightarrow{f} \hat{y} \mapsto \|y - \hat{y}\|$$

From this point, the common technique is to employ some minimization algorithm to find the parameters minimizing this output.

What we choose to do instead is develop a distribution over possible parameter values, thus giving a fuller picture of what could be happening. To do this, we turn to Bayesian statistics. We will start with prior assumptions about the parameter distribution based on the standard minimization answer and then update accordingly.

The parameters are parameters to a model we built based on literature describing cancer-immune interactions. It involves the tumor, the pro-inflammatory immune cells, and the anti-inflammatory immune cells. To infer the parameter values of the model, we used a published data set of a murine tumor[17]. The experiment consisted of three distinct treatments to a tumor and the resulting average trajectories.

Our goal at the moment is just to demonstrate the capabilities of this approach, not make biologically relevant statements. Thus, we will only infer a subset of parameters at one time and leave the rest fixed.

## 4.3 Methods

### 4.3.1 The ODE model

We built a three-component ODE model using cancer ( $T$ ), pro-inflammatory immune cells ( $I_1$ ), and anti-inflammatory immune cells ( $I_2$ ) as our three populations. The cancer grows logistically in the absence of immune cells and is removed by  $I_1$  via a mass action term. This removal is modulated by the  $I_2$  population in the form of a decreasing Hill function. The  $I_1$  population has a natural source rate and death rate. There is also a natural transition of  $I_2 \rightarrow I_1$  and vice versa. The transition from  $I_2 \rightarrow I_1$  is up-regulated by cancer via a Hill function. The  $I_2$  population also has a source and death rate, except the source rate is an increasing Hill function of the tumor population. The full set of equations is thus as follows:

$$T' = rT(1 - T/K) - TI_1H_1^-(I_2) \quad (4.1)$$

$$I_1 = \sigma_1 - d_1I_1 + \alpha I_2 - I_1H_2^+(T) \quad (4.2)$$

$$I_2 = H_3^+(T) - d_2I_2 - \alpha I_2 + I_1H_2^+(T) \quad (4.3)$$

where  $H^\cdot$  is a generic Hill function with Hill coefficient 2 where the superscript indicates if the function increases or decreases with the input. That is,

$$H_i^\pm(x) = V_{\min}^i + \frac{V_{\max}^i - V_{\min}^i}{1 + \left(\frac{x}{K_i}\right)^{\mp 2}} \quad (4.4)$$

with range either  $[V_{\min}, V_{\max}]$  (for  $H^+$ ) or  $(V_{\min}, V_{\max}]$  (for  $H^-$ ).

### 4.3.2 Simulating therapy

The dataset that we are comparing against ran three therapies for treating a growing murine tumor: a control (Wild Type), one that results in a decreased  $I_2$  population (Removed  $I_2$ ), and one that also inhibits the  $I_1$  source rate into the system (Removed  $I_2$  and Inhibited  $I_1$  Source). The first intervention is administered after 11 days and second after 13 days. At the first time point, we increase the death rate of  $I_2$  ( $d_2$ ) by a fixed amount. At the second time point, we decrease the  $I_1$  source rate ( $\sigma_1$ ) by a fixed amount. These are additional parameters in the model.

### 4.3.3 Parameter inference

To infer parameter values of this model, we compare predicted trajectories of the model against known trajectories from published literature. These known trajectories are a sequence of time points at which the cancer population is known. Using Bayesian inference, we first make an initial guess for the probability distribution for a certain subset of parameters and then update that guess by comparing the *in silico* trajectories with known trajectories. We account for biological and measurement noise by assuming that each data point actually represents the mean of a normal distribution and all distributions have the same variance, which is a parameter we infer at the same time.

#### Prior distribution

To establish our initial guess for the parameter distributions—often called the prior distribution—we built and optimized an objective function that measured how far a simulated trajectory was from the known trajectory. These optimized parameter values were then used as the mean of a normal distribution with a large variance that was then truncated for biological

relevance, i.e. non-negative parameter values, and for convergence, i.e. the algorithm often became “stuck” in corners of the parameter space and did not converge. The large variance is chosen so as to not bias the algorithm too strongly towards this single point. These computations were done in MATLAB using their *fmincon* function.

## Posterior distribution estimation

To then update our prior distributions to create data-informed posterior distributions, we utilized Stan<sup>1</sup>, a software package for statistical analysis. The algorithm uses Hamiltonian Monte Carlo to explore parameter space, at each point computing a trajectory and then a likelihood for the known data given the parameters. The “Hamiltonian” in the name comes from the algorithm’s ability to use momentum to better explore parameter space, particularly “corners” that non-Hamiltonian versions of the algorithm miss.

The algorithm works by initializing several chains which evolve independently over many iterations. The goal is for each chain to converge on the same distribution, that is all chains end up exploring parameter space in equal ways up to the ordering in which they are explored. This does provide a means for checking convergence of the algorithm (Fig. 4.1AB). We have further confirmation that the algorithm did not work when we look at the trajectories that chain 4 produces and see that they do not closely reproduce the known trajectories (Fig. 4.1C).

---

<sup>1</sup>Stan Development Team. 2018. *RStan: the R interface to Stan*. R package version 2.17.3. <http://mc-stan.org>

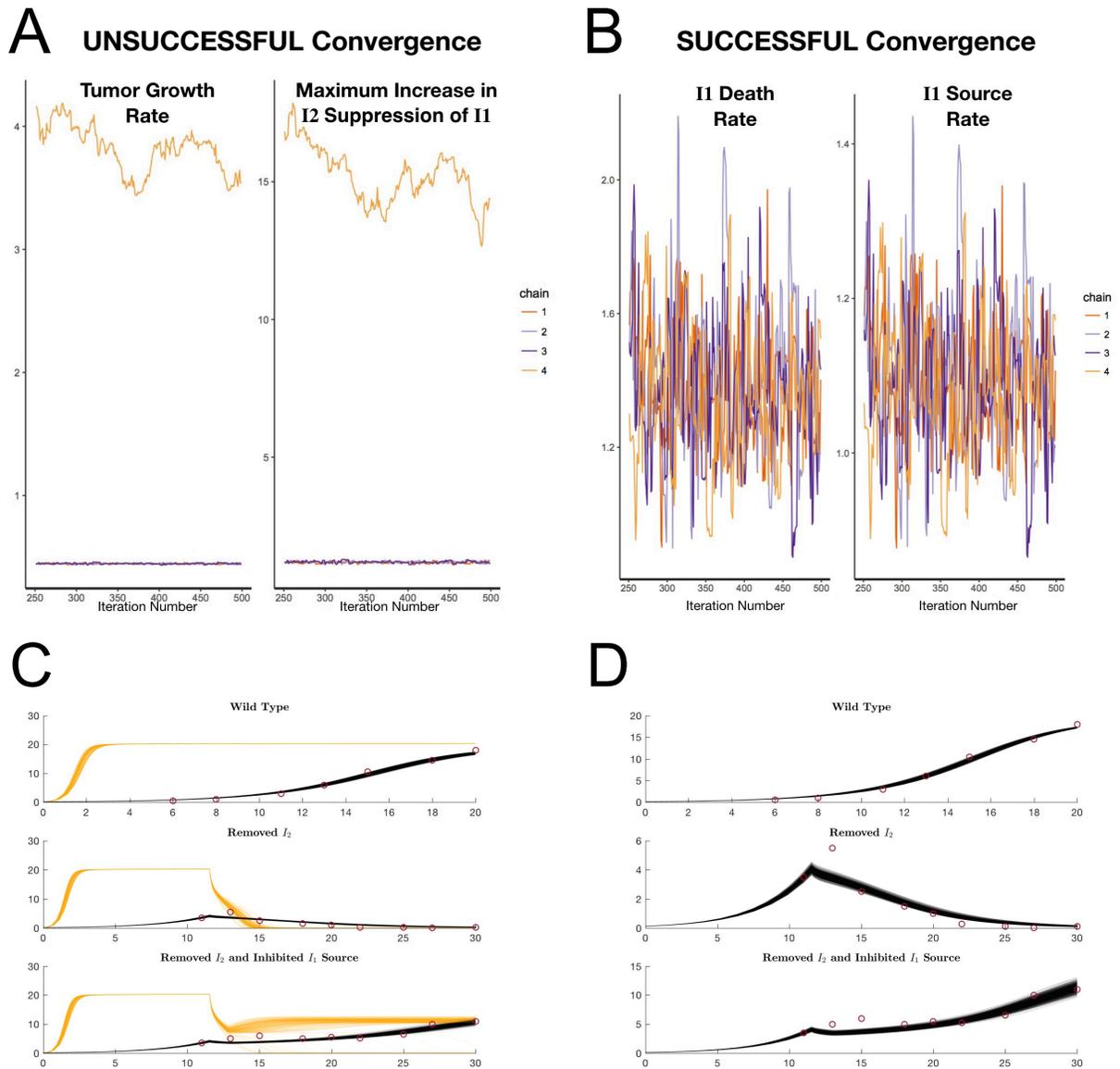


Figure 4.1: **Comparison of successful and unsuccessful convergence of HMC.** The first 250 iterations are discarded as warmup. In C and D, red circles indicate the known trajectories that we are trying to fit. A) Unsuccessful convergence. Notice chain 4 is has not converged on the same distribution as the other chains. B) Successful convergence. All four chains have converged on the same distribution of parameter values. C) The resulting trajectories from a sampling of each chain. Chains 1-3 are shown in black. Chain 4 is shown in yellow. D) The resulting trajectories from a sampling of each chain. All chains are shown in black.

## 4.4 Results

### 4.4.1 Confirmation of expected relationships and revelation of unexpected relationships between parameters

The algorithm reveals expected relationships between parameter, particularly those linked to processes that are additive or opposed. When we fix all parameters and just vary the source and death rates of  $I_1$ , a positive linear relationship between the two emerges (Fig. 4.2A). This is as expected because these two must be appropriately balanced to fit the data.

On the other hand, some surprising relationships between parameters can be uncovered by this. For example, when fixing all parameters other than the minimal source rate of  $I_2$  ( $V_{\min}^3$ ) and the maximal  $I_2$  suppression of  $I_1$  ( $V_{\min}^1$ ), we see two things (Fig. 4.2B). First, when  $V_{\min}^3$  is low,  $V_{\min}^1$  grows linearly with it. Second, once  $V_{\min}^3$  is high enough, then  $V_{\min}^1$  starts to decrease. That is, the parameters are connected in a non-monotonic pattern. The concentration of the density at low values of  $V_{\min}^3$  informs us that this is the most likely parameter values and that the positive correlation between the two is most likely. However, higher values are also possible and as Figure 4.2C shows, even at the extreme values the trajectories still fit the data.

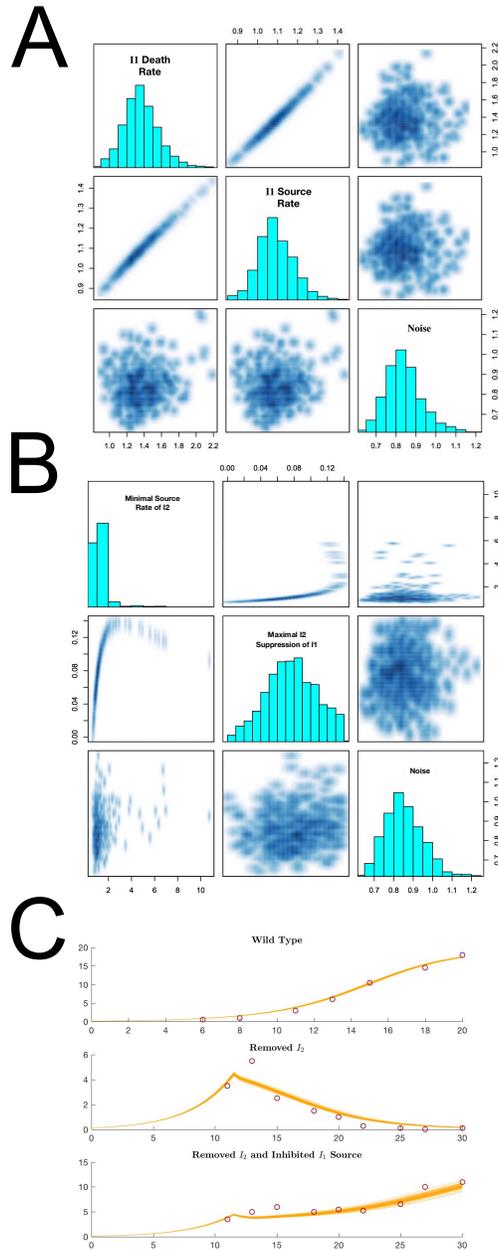


Figure 4.2: **Examples of parameter relationships.** A) Two parameters that are linearly, positively correlated. Given their directly opposing roles, this shows they balance one another out. B) Two parameters displaying a non-monotonic relationship. C) Trajectories sampled from the tail of the distribution where  $V_{\min}^3$  is high.

## 4.5 Discussion

This framework for creating a probability distribution on parameter space allows for more sophisticated parameter inference techniques. In some cases, the results merely confirm what we already expect, such as two directly opposing parameters to be positively correlated when fitting data. In other cases, surprising relationships between the parameters can be seen, including non-monotonic behavior between two parameters.

One drawback to this approach is the curse of dimensionality. The model studied here has 15 parameters (17 counting the therapy parameters) and to compute posterior distributions for all these at one time requires significant computing resources. It also was our experience that convergence was difficult to observe in these cases. However, when only a small subset of parameters can be identified as being of interest and others can be chosen from the literature, this approach can bring out new information that is overlooked by simple optimization.

# Bibliography

- [1] I. AHMAD, T. IWATA, AND H. Y. LEUNG, *Mechanisms of fgfr-mediated carcinogenesis*, Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1823 (2012), pp. 850–860.
- [2] C. L. ALTHAUS AND R. J. DE BOER, *Dynamics of immune escape during hiv/siv infection*, PLoS computational biology, 4 (2008).
- [3] G. AN AND S. KULKARNI, *An agent-based modeling framework linking inflammation and cancer using evolutionary principles: Description of a generative hierarchy for the hallmarks of cancer and developing a bridge between mechanism and epidemiological data*, Mathematical Biosciences, 260 (2015), pp. 16–24.
- [4] A. R. A. ANDERSON AND M. A. J. CHAPLAIN, *Continuous and discrete mathematical models of tumor-induced angiogenesis*, Bulletin of Mathematical Biology, 60 (1998), pp. 857–899.
- [5] M. ASHBURNER, C. A. BALL, J. A. BLAKE, D. BOTSTEIN, H. BUTLER, J. M. CHERRY, A. P. DAVIS, K. DOLINSKI, S. S. DWIGHT, J. T. EPPIG, ET AL., *Gene ontology: tool for the unification of biology*, Nature genetics, 25 (2000), pp. 25–29.
- [6] F. BALKWILL AND A. MANTOVANI, *Inflammation and cancer: back to Virchow?*, The Lancet, 357 (2001), pp. 539–545.
- [7] D. BARTIS, V. CSONGEI, A. WEICH, E. KISS, S. BARKO, T. KOVACS, M. AVDICEVIC, V. K. D’SOUZA, J. RAPP, K. KVELL, ET AL., *Down-regulation of canonical and up-regulation of non-canonical wnt signalling in the carcinogenic process of squamous cell lung carcinoma*, PloS one, 8 (2013).
- [8] S. BENZEKRY, C. LAMONT, D. BARBOLOSI, L. HLATKY, AND P. HAHNFELDT, *Mathematical Modeling of Tumor–Tumor Distant Interactions Supports a Systemic Control of Tumor Growth*, Cancer Research, 77 (2017), pp. 5183–5193.
- [9] H. BINDER, C. PORZELIUS, AND M. SCHUMACHER, *An overview of techniques for linking high-dimensional molecular data to time-to-event endpoints by risk prediction models*, Biometrical Journal, 53 (2011), pp. 170–189.
- [10] C. U. BLANK, J. B. HAANEN, A. RIBAS, AND T. N. SCHUMACHER, *The “cancer immunogram”*, Science, 352 (2016), pp. 658–660.

- [11] J. R. BRAHMER, C. G. DRAKE, I. WOLLNER, J. D. POWDERLY, J. PICUS, W. H. SHARFMAN, E. STANKEVICH, A. PONS, T. M. SALAY, T. L. McMILLER, ET AL., *Phase I study of single-agent anti-programmed death-1 (mdx-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates*, Journal of clinical oncology, 28 (2010), p. 3167.
- [12] D. T. BRAVO, Y.-L. YANG, K. KUCHENBECKER, M.-S. HUNG, Z. XU, D. M. JABLONS, AND L. YOU, *Frizzled-8 receptor is activated by the wnt-2 ligand in non-small cell lung cancer*, BMC cancer, 13 (2013), p. 316.
- [13] M. A. BRIONES-ORTA, L. LEVY, C. D. MADSEN, D. DAS, Y. ERKER, E. SAHAI, AND C. S. HILL, *Arkadia regulates tumor metastasis by modulation of the TGF- $\beta$  pathway.*, Cancer Research, 73 (2013), pp. 1800–1810.
- [14] A. BULMAN, M. NEAGU, AND C. CONSTANTIN, *Immunomics in skin cancer-improvement in diagnosis, prognosis and therapy monitoring*, Current Proteomics, 10 (2013), pp. 202–217.
- [15] R. CABRITA, M. LAUSS, A. SANNA, M. DONIA, M. S. LARSEN, S. MITRA, I. JOHANSSON, B. PHUNG, K. HARBST, J. VALLON-CHRISTERSSON, ET AL., *Tertiary lymphoid structures improve immunotherapy and survival in melanoma*, Nature, (2020), pp. 1–5.
- [16] R. J. G. B. CAMPELLO, D. MOULAVI, AND J. SANDER, *Density-Based Clustering Based on Hierarchical Density Estimates*, in Advances in Knowledge Discovery and Data Mining, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, eds., Lecture Notes in Computer Science, Berlin, Heidelberg, 2013, Springer, pp. 160–172.
- [17] R. CARRETERO, I. M. SEKTIOGLU, N. GARBI, O. C. SALGADO, P. BECKHOVE, AND G. J. HÄMMERLING, *Eosinophils orchestrate cancer rejection by normalizing tumor vessels and enhancing infiltration of cd8+ t cells*, Nature immunology, 16 (2015), pp. 609–617.
- [18] H.-C. CHEN, R. L. KODELL, K. F. CHENG, AND J. J. CHEN, *Assessment of performance of survival prediction models for cancer prognosis*, BMC medical research methodology, 12 (2012), p. 102.
- [19] H. CHO AND D. LEVY, *Modeling the Dynamics of Heterogeneity of Solid Tumors in Response to Chemotherapy*, Bulletin of Mathematical Biology, 79 (2017), pp. 2986–3012.
- [20] A. COLAPRICO, T. C. SILVA, C. OLSEN, L. GAROFANO, C. CAVA, D. GAROLINI, T. S. SABEDOT, T. M. MALTA, S. M. PAGNOTTA, I. CASTIGLIONI, ET AL., *Tcgabio: an R/bioconductor package for integrative analysis of tcga data*, Nucleic acids research, 44 (2016), pp. e71–e71.
- [21] G. O. CONSORTIUM, *The gene ontology resource: 20 years and still going strong*, Nucleic acids research, 47 (2019), pp. D330–D338.

- [22] J. M. CONWAY AND A. S. PERELSON, *A hepatitis c virus infection model with time-varying drug effectiveness: solution and analysis*, PLoS computational biology, 10 (2014).
- [23] K. CYLL, E. ERSVÆR, L. VLATKOVIC, M. PRADHAN, W. KILDAL, M. A. KJÆR, A. KLEPPE, T. S. HVEEM, B. CARLSEN, S. GILL, S. LÖFFELER, E. S. HAUG, H. WÆHRE, P. SOORIAKUMARAN, AND H. E. DANIELSEN, *Tumour heterogeneity poses a significant challenge to cancer biomarker research*, British Journal of Cancer, 117 (2017), pp. 367–375.
- [24] I. DAGOGO-JACK AND A. T. SHAW, *Tumour heterogeneity and resistance to cancer therapies*, Nature Reviews Clinical Oncology, 15 (2018), pp. 81–94.
- [25] W. DAMSKY, L. JILAVEANU, N. TURNER, C. PERRY, C. ZITO, M. TOMAYKO, J. LEVENTHAL, K. HEROLD, E. MEFFRE, M. BOSENBERG, ET AL., *B cell depletion or absence does not impede anti-tumor activity of pd-1 inhibitors*, Journal for immunotherapy of cancer, 7 (2019), p. 153.
- [26] M. P. DAVENPORT, R. M. RIBEIRO, D. L. CHAO, AND A. S. PERELSON, *Predicting the impact of a nonsterilizing vaccine against human immunodeficiency virus*, Journal of virology, 78 (2004), pp. 11340–11351.
- [27] C. DAVID, Y.-H. HUANG, M. CHEN, J. SU, Y. ZOU, N. BARDEESY, C. IACOBUZIO-DONAHUE, AND J. MASSAGUÉ, *TGF- $\beta$  Tumor Suppression through a Lethal EMT*, Cell, 164 (2016), pp. 1015–1030.
- [28] R. J. DE BOER, *Understanding the failure of cd8+ t-cell vaccination against simian/human immunodeficiency virus*, Journal of virology, 81 (2007), pp. 2838–2848.
- [29] M. DE LAURENTIIS, S. DE PLACIDO, A. R. BIANCO, G. M. CLARK, AND P. M. RAVDIN, *A prognostic model that makes quantitative estimates of probability of relapse for breast cancer patients*, Clinical Cancer Research, 5 (1999), pp. 4133–4139.
- [30] L. G. DE PILLIS AND A. E. RADUNSKAYA, *Modeling tumor-immune dynamics*, in Mathematical Models of Tumor-Immune System Dynamics, Springer, 2014, pp. 59–108.
- [31] K. E. DE VISSER, A. EICHTEN, AND L. M. COUSSENS, *Paradoxical roles of the immune system during cancer development*, Nature Reviews Cancer, 6 (2006), p. 24.
- [32] H. DILLEKÅS, M. S. ROGERS, AND O. STRAUME, *Are 90% of deaths from cancer caused by metastases?*, Cancer Medicine, 8 (2019), pp. 5574–5576.
- [33] A. DUTT, H. B. SALVESEN, T.-H. CHEN, A. H. RAMOS, R. C. ONOFRIO, C. HATTON, R. NICOLETTI, W. WINCKLER, R. GREWAL, M. HANNA, ET AL., *Drug-sensitive fgfr2 mutations in endometrial carcinoma*, Proceedings of the National Academy of Sciences, 105 (2008), pp. 8713–8717.

- [34] R. EFTIMIE, J. L. BRAMSON, AND D. J. EARN, *Interactions between the immune system and cancer: a brief review of non-spatial mathematical models*, Bulletin of mathematical biology, 73 (2011), pp. 2–32.
- [35] R. EFTIMIE, J. DUSHOFF, B. W. BRIDLE, J. L. BRAMSON, AND D. J. EARN, *Multi-stability and multi-instability phenomena in a mathematical model of tumor-immune-virus interactions*, Bulletin of mathematical biology, 73 (2011), pp. 2932–2961.
- [36] M. ELEMANS, A. FLORINS, L. WILLEMS, AND B. ASQUITH, *Rates of ctl killing in persistent viral infection in vivo*, PLoS computational biology, 10 (2014).
- [37] M. ESTER, H.-P. KRIEGEL, J. SANDER, X. XU, ET AL., *A density-based algorithm for discovering clusters in large spatial databases with noise.*, in Kdd, vol. 96, 1996, pp. 226–231.
- [38] G. S. FALCHOOK, R. LEIDNER, E. STANKEVICH, B. PIENING, C. BIFULCO, I. LOWY, AND M. G. FURY, *Responses of metastatic basal cell and cutaneous squamous cell carcinomas to anti-pd1 monoclonal antibody regn2810*, Journal for immunotherapy of cancer, 4 (2016), p. 70.
- [39] O. J. FINN, *Immuno-oncology: understanding the function and dysfunction of the immune system in cancer*, Annals of Oncology, 23 (2012), pp. viii6–viii9.
- [40] D. FRADKIN AND I. MUCHNIK, *Support vector machines for classification*, DIMACS series in discrete mathematics and theoretical computer science, 70 (2006), pp. 13–20.
- [41] M. I. FREIDLIN AND A. D. WENTZELL, *Random perturbations*, in Random perturbations of dynamical systems, Springer, 1998, pp. 15–43.
- [42] J. GALLAHER, A. BABU, S. PLEVITIS, AND A. R. A. ANDERSON, *Bridging Population and Tissue Scale Tumor Dynamics: A New Paradigm for Understanding Differences in Tumor Growth and Metastatic Disease*, Cancer Research, 74 (2014), pp. 426–435.
- [43] J. GALLAHER, A. HAWKINS-DAARUD, S. C. MASSEY, K. SWANSON, AND A. R. A. ANDERSON, *Hybrid approach for parameter estimation in agent-based models*, bioRxiv, (2017).
- [44] J. A. GALLAHER, P. M. ENRIQUEZ-NAVAS, K. A. LUDDY, R. A. GATENBY, AND A. R. A. ANDERSON, *Spatial Heterogeneity and Evolutionary Dynamics Modulate Time to Recurrence in Continuous and Adaptive Cancer Therapies*, Cancer Research, 78 (2018), pp. 2127–2139.
- [45] V. V. GANUSOV, N. GOONETILLEKE, M. K. LIU, G. FERRARI, G. M. SHAW, A. J. MCMICHAEL, P. BORROW, B. T. KORBER, AND A. S. PERELSON, *Fitness costs and diversity of the cytotoxic t lymphocyte (ctl) response determine the rate of ctl escape during acute and chronic phases of hiv infection*, Journal of virology, 85 (2011), pp. 10518–10528.

- [46] A. GELMAN, J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI, AND D. B. RUBIN, *Bayesian data analysis*, CRC press, 2013.
- [47] G. GIBSON, *Decanalization and the origin of complex disease*, Nature Reviews Genetics, 10 (2009), pp. 134–140.
- [48] F. GINHOUX, J. L. SCHULTZE, P. J. MURRAY, J. OCHANDO, AND S. K. BISWAS, *New insights into the multidimensional concept of macrophage ontogeny, activation and function*, Nature immunology, 17 (2016), pp. 34–40.
- [49] T. GRAFKE, T. SCHÄFER, AND E. VANDEN-EIJNDEN, *Long term effects of small random perturbations on dynamical systems: Theoretical and computational tools*, in Recent Progress and Modern Challenges in Applied Mathematics, Modeling and Computational Science, Springer, 2017, pp. 17–55.
- [50] J. M. GREENE, D. LEVY, K. L. FUNG, P. S. SOUZA, M. M. GOTTESMAN, AND O. LAVI, *Modeling intrinsic heterogeneity and growth of cancer cells*, Journal of Theoretical Biology, 367 (2015), pp. 262–277.
- [51] J. M. GREENE, D. LEVY, S. P. HERRADA, M. M. GOTTESMAN, AND O. LAVI, *Mathematical Modeling Reveals That Changes to Local Cell Density Dynamically Modulate Baseline Variations in Cell Growth and Drug Response*, Cancer Research, 76 (2016), pp. 2882–2890.
- [52] F. R. GRETEN AND S. I. GRIVENNIKOV, *Inflammation and cancer: Triggers, mechanisms, and consequences*, Immunity, 51 (2019), pp. 27–41.
- [53] R. L. GROSSMAN, A. P. HEATH, V. FERRETTI, H. E. VARMUS, D. R. LOWY, W. A. KIBBE, AND L. M. STAUDT, *Toward a shared vision for cancer genomic data*, New England Journal of Medicine, 375 (2016), pp. 1109–1112.
- [54] Y. GUO, Q. NIE, A. L. MACLEAN, Y. LI, J. LEI, AND S. LI, *Multiscale modeling of inflammation-induced tumorigenesis reveals competing oncogenic and onco-protective roles for inflammation*, Cancer research, (2017), pp. canres–1662.
- [55] J. HABETS, B. TANK, V. D. VUZEVSKI, E. C. VAN REEDE, E. STOLZ, AND T. VAN JOOST, *Characterization of the mononuclear infiltrate in basal cell carcinoma: a predominantly t cell-mediated immune response with minor participation of leu-7+(natural killer) cells and leu-14+(b) cells.*, Journal of investigative dermatology, 90 (1988).
- [56] B. A. HELMINK, S. M. REDDY, J. GAO, S. ZHANG, R. BASAR, R. THAKUR, K. YIZHAK, M. SADE-FELDMAN, J. BLANDO, G. HAN, ET AL., *B cells and tertiary lymphoid structures promote immunotherapy response*, Nature, (2020), pp. 1–7.
- [57] M. P. HERON, *Deaths: leading causes for 2017*, Centers for Disease Control, (2019).
- [58] M. HEYMAN AND E. VANDEN-EIJNDEN, *The geometric minimum action method: A least action principle on the space of curves*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 61 (2008), pp. 1052–1117.

- [59] T. HONG, K. WATANABE, C. H. TA, A. VILLARREAL-PONCE, Q. NIE, AND X. DAI, *An *ovol2-zeb1* mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states*, PLoS Computational Biology, 11 (2015), p. e1004569.
- [60] B. HU, E. ELINAV, S. HUBER, C. J. BOOTH, T. STROWIG, C. JIN, S. C. EISENBARTH, AND R. A. FLAVELL, *Inflammation-induced tumorigenesis in the colon is regulated by caspase-1 and *nlr4**, Proceedings of the National Academy of Sciences, 107 (2010), pp. 21635–21640.
- [61] Z. HU, X. CHEN, Y. ZHAO, T. TIAN, G. JIN, Y. SHU, Y. CHEN, L. XU, K. ZEN, C. ZHANG, ET AL., *Serum microrna signatures identified in a genome-wide serum microrna expression profiling predict survival of non-small-cell lung cancer*, J Clin Oncol, 28 (2010), pp. 1721–1726.
- [62] S. HUANG, I. ERNBERG, AND S. KAUFFMAN, *Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective*, in Seminars in cell & developmental biology, vol. 20, Elsevier, 2009, pp. 869–876.
- [63] S. HUANG AND S. KAUFFMAN, *How to escape the cancer attractor: rationale and limitations of multi-target drugs*, in Seminars in cancer biology, vol. 23, Elsevier, 2013, pp. 270–278.
- [64] S. IKEDA, A. M. GOODMAN, P. R. COHEN, T. J. JENSEN, C. K. ELLISON, G. FRAMPTON, V. MILLER, S. P. PATEL, AND R. KURZROCK, *Metastatic basal cell carcinoma with amplification of *pd-l1*: exceptional response to anti-*pd1* therapy*, NPJ genomic medicine, 1 (2016), pp. 1–5.
- [65] M. K. JOLLY, D. JIA, M. BOARETO, S. A. MANI, K. J. PIENTA, E. BEN-JACOB, AND H. LEVINE, *Coupling the modules of *emt* and *stemness*: A tunable ‘stemness window’ model*, Oncotarget, 6 (2015), p. 25161.
- [66] A. JONES, A. E. TESCHENDORFF, Q. LI, J. D. HAYWARD, A. KANNAN, T. MOULD, J. WEST, M. ZIKAN, D. CIBULA, H. FIEGL, ET AL., *Role of dna methylation and epigenetic silencing of *hand2* in endometrial cancer development*, PLoS medicine, 10 (2013).
- [67] E. KIM, J.-Y. KIM, M. A. SMITH, E. B. HAURA, AND A. R. A. ANDERSON, *Cell signaling heterogeneity is modulated by both cell-intrinsic and -extrinsic mechanisms: An integrated approach to understanding targeted therapy*, PLoS Biology, 16 (2018), p. e2002930.
- [68] P. D. W. KIRK, T. THORNE, AND M. P. H. STUMPF, *Model selection in systems and synthetic biology*, Current Opinion in Biotechnology, 24 (2013), pp. 767–774.
- [69] J. P. KLEIN, H. C. VAN HOUWELINGEN, J. G. IBRAHIM, AND T. H. SCHEIKE, *Handbook of survival analysis.*, CRC Press, 2016.

- [70] J. E. KORKOLA, J. HOULDSWORTH, D. R. FELDMAN, A. B. OLSHEN, L.-X. QIN, S. PATIL, V. E. REUTER, G. J. BOSL, AND R. CHAGANTI, *Identification and validation of a gene expression signature that predicts outcome in adult men with germ cell tumors*, *Journal of Clinical Oncology*, 27 (2009), p. 5240.
- [71] N. KRAMER, J. SCHMÖLLERL, C. UNGER, H. NIVARTHI, A. RUDISCH, D. UNTERLEUTHNER, M. SCHERZER, A. RIEDL, M. ARTAKER, I. CRNCEC, ET AL., *Autocrine wnt2 signaling in fibroblasts promotes colorectal cancer progression*, *Oncogene*, 36 (2017), pp. 5460–5472.
- [72] P. KUCEROVA AND M. CERVINKOVA, *Spontaneous regression of tumour and the role of microbial infection—possibilities for cancer treatment*, *Anti-Cancer Drugs*, 27 (2016), p. 269.
- [73] M. V. KULESHOV, M. R. JONES, A. D. ROUILLARD, N. F. FERNANDEZ, Q. DUAN, Z. WANG, S. KOPLEV, S. L. JENKINS, K. M. JAGODNIK, A. LACHMANN, ET AL., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update*, *Nucleic acids research*, 44 (2016), pp. W90–W97.
- [74] V. A. KUZNETSOV, I. A. MAKALKIN, M. A. TAYLOR, AND A. S. PERELSON, *Non-linear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis*, *Bulletin of mathematical biology*, 56 (1994), pp. 295–321.
- [75] E. LABBÉ, A. LETAMENDIA, AND L. ATTISANO, *Association of smads with lymphoid enhancer binding factor 1/t cell-specific factor mediates cooperative signaling by the transforming growth factor- $\beta$  and wnt pathways*, *Proceedings of the National Academy of Sciences*, 97 (2000), pp. 8358–8363.
- [76] B. LAMBERT, A. L. MACLEAN, A. G. FLETCHER, A. N. COMBES, M. H. LITTLE, AND H. M. BYRNE, *Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis.*, *Journal of Mathematical Biology*, 10 (2018), p. 106.
- [77] O. LAVI, J. M. GREENE, D. LEVY, AND M. M. GOTTESMAN, *The Role of Cell Density and Intratumoral Heterogeneity in Multidrug Resistance*, *Cancer Research*, (2013).
- [78] J. E. LEE, S.-H. SHIN, H.-W. SHIN, Y.-S. CHUN, AND J.-W. PARK, *Nuclear fgfr2 negatively regulates hypoxia-induced cell invasion in prostate cancer by interacting with hif-1 and hif-2*, *Scientific reports*, 9 (2019), pp. 1–12.
- [79] K.-M. LEUNG, R. M. ELASHOFF, AND A. A. AFIFI, *Censoring issues in survival analysis*, *Annual review of public health*, 18 (1997), pp. 83–104.
- [80] C. LI AND J. WANG, *Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths*, *PLoS computational biology*, 9 (2013).

- [81] Q. LI, A. KANNAN, F. J. DEMAYO, J. P. LYDON, P. S. COOKE, H. YAMAGISHI, D. SRIVASTAVA, M. K. BAGCHI, AND I. C. BAGCHI, *The antiproliferative action of progesterone in uterine epithelium is mediated by hand2*, *Science*, 331 (2011), pp. 912–916.
- [82] A. LIBERZON, C. BIRGER, H. THORVALDSDÓTTIR, M. GHANDI, J. P. MESIROV, AND P. TAMAYO, *The molecular signatures database hallmark gene set collection*, *Cell systems*, 1 (2015), pp. 417–425.
- [83] A. LIBERZON, A. SUBRAMANIAN, R. PINCHBACK, H. THORVALDSDÓTTIR, P. TAMAYO, AND J. P. MESIROV, *Molecular signatures database (msigdb) 3.0*, *Bioinformatics*, 27 (2011), pp. 1739–1740.
- [84] J. LIM AND J. P. THIERY, *Epithelial-mesenchymal transitions: insights from development*, *Development*, 139 (2012), pp. 3471–3486.
- [85] J.-H. LIM, K.-E. LEE, K.-S. HAHN, AND K.-W. PARK, *Analyzing survival data as binary outcomes with logistic regression*, *Communications for Statistical Applications and Methods*, 17 (2010), pp. 117–126.
- [86] H. LINARDOU AND H. GOGAS, *Toxicity management of immunotherapy for patients with metastatic melanoma*, *Annals of translational medicine*, 4 (2016).
- [87] E. J. LIPSON, M. T. LILO, A. OGURTSOVA, J. ESANDRIO, H. XU, P. BROTHERS, M. SCHOLLENBERGER, W. H. SHARFMAN, AND J. M. TAUBE, *Basal cell carcinoma: Pd-l1/pd-1 checkpoint expression and tumor regression after pd-1 blockade*, *Journal for immunotherapy of cancer*, 5 (2017), p. 23.
- [88] J. LIU, T. LICHTENBERG, K. A. HOADLEY, L. M. POISSON, A. J. LAZAR, A. D. CHERNIACK, A. J. KOVATICH, C. C. BENZ, D. A. LEVINE, A. V. LEE, ET AL., *An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics*, *Cell*, 173 (2018), pp. 400–416.
- [89] P.-K. LO, J. S. LEE, X. LIANG, L. HAN, T. MORI, M. J. FACKLER, H. SADIK, P. ARGANI, T. K. PANDITA, AND S. SUKUMAR, *Epigenetic inactivation of the potential tumor suppressor gene foxf1 in breast cancer*, *Cancer research*, 70 (2010), pp. 6047–6058.
- [90] W. W. LOCKWOOD, S. K. CHANDEL, G. L. STEWART, H. ERDJUMENT-BROMAGE, AND L. J. BEVERLY, *The novel ubiquitin ligase complex, scffbxw4, interacts with the cop9 signalosome in an f-box dependent manner, is mutated, lost and under-expressed in human cancers*, *PloS one*, 8 (2013).
- [91] Y. LOUZOUN, C. XUE, G. B. LESINSKI, AND A. FRIEDMAN, *A mathematical model for pancreatic cancer growth and treatments*, *Journal of Theoretical Biology*, 351 (2014), pp. 74–82.
- [92] A. L. MACLEAN, T. HONG, AND Q. NIE, *Exploring intermediate cell states through the lens of single cells*, *Current Opinion in Systems Biology*, 9 (2018), pp. 32–41.

- [93] L. MCINNEN, J. HEALY, AND J. MELVILLE, *Umap: Uniform manifold approximation and projection for dimension reduction*, arXiv preprint arXiv:1802.03426, (2018).
- [94] A. J. MIKELS AND R. NUSSE, *Purified wnt5a protein activates or inhibits  $\beta$ -catenin-tcf signaling depending on receptor context*, PLoS biology, 4 (2006).
- [95] A. J. MILLER AND M. C. MIHM JR, *Melanoma*, New England Journal of Medicine, 355 (2006), pp. 51–65.
- [96] N. MORIS, C. PINA, AND A. M. ARIAS, *Transition states and cell fate decisions in epigenetic landscapes*, Nature Reviews Genetics, 17 (2016), p. 693.
- [97] M. D. MORRIS, *Factorial sampling plans for preliminary computational experiments*, Technometrics, 33 (1991), pp. 161–174.
- [98] S. MUKHERJEE, *The emperor of all maladies: a biography of cancer*, Simon and Schuster, 2010.
- [99] C. MURDOCH, M. MUTHANA, S. B. COFFELT, AND C. E. LEWIS, *The role of myeloid cells in the promotion of tumour angiogenesis*, Nature reviews cancer, 8 (2008), pp. 618–631.
- [100] V. MURILLO-GARZÓN, I. GORROÑO-ETXEARRIA, M. ÅKERFELT, M. C. PUUSTINEN, L. SISTONEN, M. NEES, J. CARTON, J. WAXMAN, AND R. M. KYPTA, *Frizzled-8 integrates wnt-11 and transforming growth factor- $\beta$  signaling in prostate cancer*, Nature communications, 9 (2018), pp. 1–16.
- [101] Q. NIE, *Stem cells: a window of opportunity in low-dimensional emt space*, Oncotarget, 9 (2018), p. 31790.
- [102] M. A. NIETO, R. Y.-J. HUANG, R. A. JACKSON, AND J. P. THIERY, *Emt: 2016*, Cell, 166 (2016), pp. 21–45.
- [103] R. NUSSE AND H. CLEVERS, *Wnt/ $\beta$ -catenin signaling, disease, and emerging therapeutic modalities*, Cell, 169 (2017), pp. 985–999.
- [104] M. OPPER AND O. WINTHER, *Gaussian processes for classification: Mean-field algorithms*, Neural computation, 12 (2000), pp. 2655–2684.
- [105] N. OSHIMORI, D. ORISTIAN, AND E. FUCHS, *TGF- $\beta$  Promotes Heterogeneity and Drug Resistance in Squamous Cell Carcinoma*, Cell, 160 (2015), pp. 963–976.
- [106] M. R. OWEN, I. J. STAMPER, M. MUTHANA, G. W. RICHARDSON, J. DOBSON, C. E. LEWIS, AND H. M. BYRNE, *Mathematical Modeling Predicts Synergistic Antitumor Effects of Combining a Macrophage-Based, Hypoxia-Targeted Gene Therapy with Chemotherapy*, Cancer Research, 71 (2011), pp. 2826–2837.
- [107] A. K. PALUCKA AND L. M. COUSSENS, *The basis of oncoimmunology*, Cell, 164 (2016), pp. 1233–1247.

- [108] D. M. PARDOLL, *The blockade of immune checkpoints in cancer immunotherapy*, Nature Reviews Cancer, 12 (2012), p. 252.
- [109] A. PASSARELLI, F. MANNAVOLA, L. S. STUCCI, M. TUCCI, AND F. SILVESTRIS, *Immune system and melanoma biology: a balance between immunosurveillance and immune escape*, Oncotarget, 8 (2017), p. 106132.
- [110] H. PEINADO, D. OLMEDA, AND A. CANO, *Snail, zeb and bhlh factors in tumour progression: an alliance against the epithelial phenotype?*, Nature Reviews Cancer, 7 (2007), pp. 415 EP –.
- [111] F. PETITPREZ, A. DE REYNIÈS, E. Z. KEUNG, T. W.-W. CHEN, C.-M. SUN, J. CALDERARO, Y.-M. JENG, L.-P. HSIAO, L. LACROIX, A. BOUGOÛIN, ET AL., *B cells are associated with survival and immunotherapy response in sarcoma*, Nature, (2020), pp. 1–5.
- [112] L. G. D. PILLIS, A. E. RADUNSKAYA, AND C. L. WISEMAN, *A Validated Mathematical Model of Cell-Mediated Immune Response to Tumor Growth*, Cancer Research, 65 (2005), pp. 7950–7958.
- [113] P. POLAKIS, *Wnt signaling and cancer*, Genes & development, 14 (2000), pp. 1837–1851.
- [114] M. A. POSTOW, M. K. CALLAHAN, AND J. D. WOLCHOK, *Immune checkpoint blockade in cancer therapy*, Journal of clinical oncology, 33 (2015), p. 1974.
- [115] C. PRUNIER, D. BAKER, P. TEN DIJKE, AND L. RITSMA, *Tgf- $\beta$  family signaling pathways in cellular dormancy*, Trends in cancer, 5 (2019), pp. 66–78.
- [116] C. J. A. PUNT, M. KOOPMAN, AND L. VERMEULEN, *From tumour heterogeneity to advances in precision treatment of colorectal cancer*, Nature Reviews Clinical Oncology, 14 (2017), pp. 235–246.
- [117] C. RACKAUCKAS, T. SCHILLING, AND Q. NIE, *Mean-independent noise control of cell fates via intermediate states*, iScience, 3 (2018), pp. 11–20.
- [118] C. E. RASMUSSEN AND C. K. WILLIAMS, *Gaussian processes for machine learning, vol 14, no 2*, 2006.
- [119] N. P. RESTIFO, M. E. DUDLEY, AND S. A. ROSENBERG, *Adoptive immunotherapy for cancer: harnessing the t cell response*, Nature Reviews Immunology, 12 (2012), p. 269.
- [120] A. RIBAS AND J. D. WOLCHOK, *Cancer immunotherapy using checkpoint blockade*, Science, 359 (2018), pp. 1350–1355.
- [121] D. RICOL, D. CAPPELLEN, A. EL MARJOU, S. GIL-DIEZ-DE MEDINA, J.-M. GIRAULT, T. YOSHIDA, G. FERRY, G. TUCKER, M.-F. POUPON, D. CHOPIN, ET AL., *Tumour suppressive properties of fibroblast growth factor receptor 2-iiib in human bladder cancer*, Oncogene, 18 (1999), pp. 7234–7243.

- [122] B. RUFFELL AND L. M. COUSSENS, *Macrophages and therapeutic resistance in cancer*, *Cancer cell*, 27 (2015), pp. 462–472.
- [123] B. RUFFELL, D. G. DENARDO, N. I. AFFARA, AND L. M. COUSSENS, *Lymphocytes in cancer development: polarization towards pro-tumor immunity*, *Cytokine & Growth Factor Reviews*, 21 (2010), pp. 3–10.
- [124] B. M. RYAN AND J. M. FAUPEL-BADGER, *The hallmarks of premalignant conditions: a molecular basis for cancer prevention*, in *Seminars in oncology*, vol. 43, Elsevier, 2016, pp. 22–35.
- [125] F. SABBATINO, A. MARRA, L. LIGUORI, G. SCOGNAMIGLIO, C. FUSCIELLO, G. BOTTI, S. FERRONE, AND S. PEPE, *Resistance to anti-pd-1-based immunotherapy in basal cell carcinoma: a case report and review of the literature*, *Journal for immunotherapy of cancer*, 6 (2018), p. 126.
- [126] M. SADE-FELDMAN, K. YIZHAK, S. L. BJORGAARD, J. P. RAY, C. G. DE BOER, R. W. JENKINS, D. J. LIEB, J. H. CHEN, D. T. FREDERICK, M. BARZILY-ROKNI, ET AL., *Defining t cell states associated with response to checkpoint immunotherapy in melanoma*, *Cell*, 176 (2019), p. 404.
- [127] A. SÁNCHEZ-DANÉS, J.-C. LARSIMONT, M. LIAGRE, E. MUÑOZ-COUSELO, G. LAPOUGE, A. BRISEBARRE, C. DUBOIS, M. SUPPA, V. SUKUMARAN, V. DEL MARMOL, ET AL., *A slow-cycling lgr5 tumour population mediates basal cell carcinoma relapse after therapy*, *Nature*, 562 (2018), p. 434.
- [128] A. SARVARIA, J. A. MADRIGAL, AND A. SAUDEMONT, *B cell regulation in cancer and anti-tumor immunity*, *Cellular & molecular immunology*, 14 (2017), pp. 662–674.
- [129] R. D. SCHREIBER, L. J. OLD, AND M. J. SMYTH, *Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion*, *Science*, 331 (2011), pp. 1565–1570.
- [130] R. SERRE, S. BENZEKRY, L. PADOVANI, C. MEILLE, N. ANDRÉ, J. CICCOLINI, F. BARLESI, X. MURACCIOLE, AND D. BARBOLOSI, *Mathematical Modeling of Cancer Immunotherapy and Its Synergy with Radiotherapy*, *Cancer Research*, 76 (2016), pp. 4931–4940.
- [131] Y. SHA, D. HAENSEL, G. GUTIERREZ, H. DU, X. DAI, AND Q. NIE, *Intermediate cell states in epithelial-to-mesenchymal transition*, *Physical biology*, (2018).
- [132] —, *Intermediate cell states in epithelial-to-mesenchymal transition*, *Physical biology*, 16 (2019), p. 021001.
- [133] P. SHARMA, S. HU-LIESKOVAN, J. A. WARGO, AND A. RIBAS, *Primary, adaptive, and acquired resistance to cancer immunotherapy*, *Cell*, 168 (2017), pp. 707–723.
- [134] A. H. SHARPE, *Introduction to checkpoint inhibitors and cancer immunotherapy*, *Immunological reviews*, 276 (2017), p. 5.

- [135] SHERRATT JONATHAN A. AND NOWAK MARTIN A., *Oncogenes, anti-oncogenes and the immune response to cancer : a mathematical model*, Proceedings of the Royal Society of London. Series B: Biological Sciences, 248 (1992), pp. 261–271.
- [136] C. SHI, Y. CHEN, Y. CHEN, Y. YANG, W. BING, AND J. QI, *Cd4+ cd25+ regulatory t cells promote hepatocellular carcinoma invasion via tgf- $\beta$ 1-induced epithelial–mesenchymal transition*, OncoTargets and therapy, 12 (2019), p. 279.
- [137] T. SHIBUE AND R. A. WEINBERG, *EMT, CSCs, and drug resistance: the mechanistic link and clinical implications*, Nature Reviews Clinical Oncology, 14 (2017), pp. 611–629. Number: 10 Publisher: Nature Publishing Group.
- [138] A. SHWARTZ AND A. WEISS, *Large deviations for performance analysis: queues, communication and computing*, vol. 5, CRC Press, 1995.
- [139] H. SOHIER, J.-L. FARGES, AND H. PIET-LAHANIER, *Improvement of the representativity of the morris method for air-launch-to-orbit separation*, IFAC Proceedings Volumes, 47 (2014), pp. 7954–7959.
- [140] C. H. TA, Q. NIE, AND T. HONG, *Controlling stochasticity in epithelial-mesenchymal transition through multiple intermediate cellular states*, Discrete and continuous dynamical systems. Series B, 21 (2016), p. 2275.
- [141] M. TAMURA, Y. SASAKI, R. KOYAMA, K. TAKEDA, M. IDOGAWA, AND T. TOKINO, *Forkhead transcription factor foxf1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness*, Oncogene, 33 (2014), pp. 4837–4846.
- [142] *The Cancer Genome Atlas Program*, June 2018. Archive Location: nciglobal,ncicenterprise Library Catalog: [www.cancer.gov](http://www.cancer.gov).
- [143] S. TERRY, P. SAVAGNER, S. ORTIZ-CUARAN, L. MAHJOUBI, P. SAINTIGNY, J.-P. THIERY, AND S. CHOUAIB, *New insights into the role of emt in tumor immune escape*, Molecular Oncology, 11 (2017), pp. 824–846.
- [144] S. L. TOPALIAN, C. G. DRAKE, AND D. M. PARDOLL, *Immune checkpoint blockade: a common denominator approach to cancer therapy*, Cancer cell, 27 (2015), pp. 450–461.
- [145] J. VANHATALO, J. RIIHIMÄKI, J. HARTIKAINEN, P. JYLÄNKI, V. TOLVANEN, AND A. VEHTARI, *Gpstuff: Bayesian modeling with gaussian processes*, Journal of Machine Learning Research, 14 (2013), pp. 1175–1179.
- [146] S. S. VARADHAN, *Large deviations and applications*, vol. 46, Siam, 1984.
- [147] A. VEHTARI, A. GELMAN, AND J. GABRY, *Practical bayesian model evaluation using leave-one-out cross-validation and waic*, Statistics and computing, 27 (2017), pp. 1413–1432.

- [148] H. WANG, L. NIU, S. JIANG, J. ZHAI, P. WANG, F. KONG, AND X. JIN, *Comprehensive analysis of aberrantly expressed profiles of lncrnas and mirnas with associated cerna network in muscle-invasive bladder cancer*, *Oncotarget*, 7 (2016), p. 86174.
- [149] J. WANG, K. ZHANG, L. XU, AND E. WANG, *Quantifying the waddington landscape and biological paths for development and differentiation*, *Proceedings of the National Academy of Sciences*, 108 (2011), pp. 8257–8262.
- [150] S. WANG, M. KARIKOMI, A. L. MACLEAN, AND Q. NIE, *Cell lineage and communication network inference via optimization for single-cell transcriptomics*, *Nucleic acids research*, 47 (2019), pp. e66–e66.
- [151] Y. WANG, T. YANG, Y. MA, G. V. HALADE, J. ZHANG, M. L. LINDSEY, AND Y.-F. JIN, *Mathematical modeling and stability analysis of macrophage activation in left ventricular remodeling post-myocardial infarction*, *BMC genomics*, 13 (2012), p. S21.
- [152] D. WARNE, R. E. BAKER, AND M. J. SIMPSON, *Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art*, *Journal of The Royal Society Interface*, 16 (2019), p. 20180943.
- [153] J. N. WEINSTEIN, E. A. COLLISSE, G. B. MILLS, K. R. M. SHAW, B. A. OZENBERGER, K. ELLROTT, I. SHMULEVICH, C. SANDER, J. M. STUART, C. G. A. R. NETWORK, ET AL., *The cancer genome atlas pan-cancer analysis project*, *Nature genetics*, 45 (2013), p. 1113.
- [154] J. WEST, L. YOU, J. BROWN, P. K. NEWTON, AND A. R. ANDERSON, *Towards multi-drug adaptive therapy*, *bioRxiv*, (2018), p. 476507.
- [155] W. D. WICK, O. O. YANG, L. COREY, AND S. G. SELF, *How many human immunodeficiency virus type 1-infected target cells can a cytotoxic t-lymphocyte kill?*, *Journal of virology*, 79 (2005), pp. 13579–13586.
- [156] J. WINKLER, R. SCHNEIDERBAUER, C. BENDER, O. SEDLACZEK, S. FRÖHLING, R. PENZEL, A. ENK, AND J. HASSEL, *Anti-programmed cell death-1 therapy in non-melanoma skin cancer*, *British Journal of Dermatology*, 176 (2017), pp. 498–502.
- [157] D. M. WITTEN AND R. TIBSHIRANI, *Survival analysis with high-dimensional covariates*, *Statistical methods in medical research*, 19 (2010), pp. 29–51.
- [158] K. WOODS, A. PASAM, A. JAYACHANDRAN, M. C. ANDREWS, AND J. CEBON, *Effects of epithelial to mesenchymal transition on t cell targeting of melanoma cells*, *Frontiers in Oncology*, 4 (2014), p. 367.
- [159] Z. YING, X. F. HUANG, X. XIANG, Y. LIU, X. KANG, Y. SONG, X. GUO, H. LIU, N. DING, T. ZHANG, ET AL., *A safe and potent anti-cd19 car t cell therapy*, *Nature medicine*, (2019), p. 1.

- [160] K. E. YOST, A. T. SATPATHY, D. K. WELLS, Y. QI, C. WANG, R. KAGEYAMA, K. L. MCNAMARA, J. M. GRANJA, K. Y. SARIN, R. A. BROWN, ET AL., *Clonal replacement of tumor-specific t cells following pd-1 blockade*, Nature medicine, 25 (2019), pp. 1251–1259.
- [161] P. ZHOU AND T. LI, *Construction of the landscape for multi-stable systems: Potential landscape, quasi-potential, a-type integral and beyond*, The Journal of chemical physics, 144 (2016), p. 094109.