

Mathematical modelling of tumour evolution and radiation response: The impact of heterogeneity



Jacob G. Scott
St Anne's College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2016

Dedication

This work is dedicated to my patients, who get up every day and face an uncertain life unfairly perturbed by cancer; and to my wife and children, who get up every day and face an uncertain life with an academic for a husband and father.

Acknowledgements

As with any scientific work, I owe a great deal of the credit to my mentors and collaborators. Firstly, to Professors Philip Maini and Alexander Anderson and Doctor Alexander Fletcher, my academic mentors and now friends, and Doctors Robert Gatenby and Craig Stevens, my professional mentors and patrons, without whom this would not have come to pass. I also owe my two closest collaborators and colleagues, Doctors David Basanta and Philip Gerlee, a great deal for their patience and ideas, as well as to countless others who have made themselves available to me for questions and camaraderie within the Wolfson Centre for Mathematical Biology. I am also grateful to the Moffitt Cancer Center for making this work possible by allowing me to take time away from the clinic and for funding my studies. Further, without my family, including my loving wife Sarah, and children Maren and Rhys's support, I would have never been able to take this valuable time away from my clinical career to come to Oxford in the first place. And last, but not least, my mom and dad. Thanks.

Abstract

This thesis seeks to use mathematical and computational models to develop measures of clinically available data to deepen our understanding, and improve our treatments, of cancer. We consider two broad characteristics of cancer: heterogeneity, in the form of differences in cellular phenotype, and the physical microenvironment; and evolution, which has become accepted as a driver of tumour progression. To ensure that the conclusions drawn are as translatable as possible, we will attempt to use data types that are clinically available. Using a hybrid discrete-cell-based model in two spatial dimensions, we focus on these fundamental aspects of cancer, with the hope of generating new understanding and useful hypotheses to benefit current patients and oncologists.

First, we model a tumour growing under the rules of the cancer stem cell hypothesis and a neutral model of evolution, and ask if we can infer the underlying biological proliferative structure. Specifically, we work toward predicting the symmetric division probability of our simulated tumours from clinically relevant observables, as this is a key driving parameter of tumour progression and therapeutic response. We focus on measures of clonal diversity, group size and shape, and a suite of statistical measures of the phylogenetic trees resulting from the tumour's evolution in different regions of parameter space. We find strikingly different patterns in these measures for changing symmetric division probability which hinge on the inclusion of spatial constraints. These results give us insight into differences between solid and liquid tumours, and also generate a number of actionable clinical and biological hypotheses.

Second, we explicitly consider the physical microenvironment of tumours invading into healthy tissue, and model oxygen transport, uptake and cellular competition. We then explore the effect of spatial organisation of blood vessels within the tumour on tumour growth kinetics and cellularity. Finding wide variability in the distribution of oxygen across tumours dependent on both vascular organisation and density, we proceed to explore the utility of spatial measures of vessels on radiation efficacy. Our results offer a novel

hypothesis as to the failure of vascular normalisation therapy and radiation, and a possible clinical solution.

Statement of Originality

This thesis is submitted to the University of Oxford in fulfillment of the requirements of the degree of Doctor of Philosophy. This thesis contains no material which has been previously submitted for a degree or diploma at this University or any other. I generated all *in silico* data and simulations. Some basic elements of the spatial cellular automaton model implemented in the java programming language were adapted from previous work I conducted with Dr David Basanta, but all extensions and further modifications presented in this thesis were done solely by me. I wrote all python and MATLAB code and performed all analyses presented herein.

Contents

1	Introduction	1
1.1	Why study mathematical oncology?	1
1.1.1	Cancer is horrible	1
1.1.2	Cancer incidence is rising and death rates are flat	2
1.1.3	A new perspective is needed	3
1.2	Cancer is heterogeneous on many scales	4
1.2.1	Microenvironmental heterogeneity	4
1.2.2	Proliferative heterogeneity: the cancer stem cell hypothesis	7
1.2.3	Intratumoural heterogeneity: clonal evolution	9
1.3	Scope of thesis	11
2	Contributions to mathematical oncology as a member of a larger team	15
2.1	Introduction	15
2.2	Metastasis on networks	15
2.3	Learning and perturbing evolution	18
2.3.1	Markov model	19
2.3.2	Evolutionary game theory	20
2.4	Other work modeling stem cell dynamics	22
2.4.1	Ordinary differential equations	22
2.4.2	Cellular automaton models	23
2.5	Clinical radiation biology	23
3	Inferring tumour proliferative organisation from measures of diversity and spatial heterogeneity	26
3.1	Introduction	26
3.1.1	Previous work	27
3.1.2	Aim of work	28
3.1.3	Structure of chapter	29
3.2	Model development	30

3.2.1	Non-spatial TIC-driven model of neutral evolution	30
3.2.2	Spatial model	33
3.3	Results	35
3.3.1	Ascertaining the effect of space on tumour diversity	35
3.3.2	The effect of space and proliferative heterogeneity on colony size distribution	40
3.3.3	Inferring symmetric division rate from spatial measures of clonal architecture	43
3.3.4	Inferring symmetric division probability from imperfect information	46
3.3.5	The influence of biopsy size on diversity estimation	48
3.3.6	Comparison of diversity prediction from biopsy in TIC-driven and non-TIC-driven tumours	48
3.4	Discussion	52
4	Inferring tumour proliferative organisation from phylogenetic tree mea- sures	55
4.1	Introduction	55
4.1.1	Phylogenetic reconstruction in human disease	55
4.1.2	Aim of work	57
4.1.3	Chapter structure	57
4.2	Methods	57
4.2.1	Phylogenetic tree reconstruction algorithm	58
4.2.2	Qualitative comparison of reconstructed trees	58
4.2.3	Candidate tree-based measures for model comparison	58
4.3	Results	62
4.3.1	Measuring trees from simulation	62
4.3.2	Variation of tree-based measures with symmetric division probability	64
4.3.3	Dynamics of tree-based measures during tumour growth	66
4.3.4	Dependence of tree-based measures on mutation probability	67
4.3.5	Changes in tree measures with spatial constraints	68
4.3.6	Utility of imperfect information in estimating tree measures	69
4.4	Discussion	70
4.4.1	Translating findings to the clinic	72
4.4.2	Conclusions	72

5	Quantifying microenvironmental heterogeneity in a hybrid cellular automaton model of cancer in a vascularised domain	74
5.1	Introduction	74
5.1.1	Previous work	75
5.1.2	Aim of work	76
5.1.3	Structure of chapter	76
5.2	Hybrid cellular automaton model of tumour and normal tissue	76
5.2.1	Modelling the continuous oxygen field	77
5.2.2	Individual cell behaviour	79
5.2.3	Cellular automaton rules and updates	81
5.3	Understanding normal tissue behaviour	88
5.3.1	Varying vascular density with regular patterning	88
5.3.2	Varying vascular density with irregular patterning	89
5.4	Vascularised tumour growth and invasion	90
5.4.1	Tumour growth in a regularly vascularised domain	90
5.4.2	Tumour growth in heterogeneous vascular domains	97
5.5	Conclusions	98
6	Effects of vascular patterning on radiation treatment efficacy	101
6.1	Introduction	101
6.1.1	Previous work	101
6.1.2	Aim of work	103
6.1.3	Structure of chapter	103
6.2	Established radiobiology	104
6.2.1	Proposed mechanisms for DNA damage	104
6.2.2	The linear-quadratic model of cell kill	104
6.2.3	Dependence of cell kill on oxygen concentration	105
6.2.4	Tumour control probability	107
6.3	The effect of heterogeneous vascular patterns on surviving cell number	108
6.3.1	Surviving cell number after radiation varies widely with vessel pattern	109
6.3.2	A simplified case: two vessels at varying distance	109
6.3.3	Vessel density is insufficient to predict radiation effect	112
6.4	Methods for quantifying vessel organisation	113
6.4.1	Spatial autocorrelation	113
6.4.2	A simple metric of vascular organisation	114
6.4.3	Voronoi tessellation	114
6.4.4	Ripley's K	117

6.5	Using spatial metrics to predict radiation effect	119
6.5.1	Aggregation coefficient	120
6.5.2	Voronoi polygon area distribution	120
6.5.3	Ripley's L	121
6.5.4	Comparison across metrics	122
6.6	Conclusions	123
6.6.1	Effect of vessel normalisation on radiation response	124
6.6.2	Future directions: translation to the clinic	126
7	Conclusions and future work	128
7.1	Part 1: Conclusions and future directions	128
7.1.1	Diversity, colony size and shape, and imperfect information: comparisons of models across symmetric division probability	128
7.1.2	Inferring symmetric division probability from phylogenetic measures	129
7.1.3	Future work	130
7.2	Part 2: Conclusions and future directions	131
7.2.1	Understanding tumour progression and cell-oxygen distributions using a hybrid cellular automaton model with vasculature	131
7.2.2	Predicting radiation response from spatial measures of vascular architecture	133
7.2.3	Future work	133
7.3	Concluding remarks	134
A	Chapter 2 supplemental material	136
A.1	Size and number of biopsies: diminishing returns	136
A.2	Shape normalisation	137
B	Chapter 3 supplemental information	138
B.1	Temporal dynamics	138
B.2	Non-spatial Model comparative statistics	139
B.3	Comparative pairwise statistics of spatial model	141
B.4	Raw vs. pruned trees	143
B.5	Algorithm for developing individual cell 'genomes' from mutational flag and life history.	143

C Chapter 4 supplemental information	145
C.1 Analytic solution to diffusion equation	145
C.2 Calibration of model using tumour growth in an avascular domain: a tumour spheroid simulation	147
D Chapter 5 supplemental information	149
D.1 Predicting carrying capacity	149
D.1.1 Aggregation coefficient	149
D.1.2 Voronoi Polygon Area	151
D.1.3 Ripley's K	153
Bibliography	154

List of Figures

1.1	Cancer is a highly heterogeneous, destructive and deadly disease.	2
1.2	Cancer deaths are rising worldwide, and are nearing the number one cause of death in the USA.	3
1.3	Cancer is complex, and exhibits heterogeneity over multiple spatial and temporal scales.	5
1.4	A comparison between the cancer stem cell hypothesis and the stochastic hypothesis.	8
1.5	The evolution of our thinking about clonal evolution in cancer.	10
1.6	Branched evolution from Darwin to cancer.	11
2.1	A simplified representation of the human vascular system connecting representative organs with a primary tumour in the liver.	16
2.2	Studying tumour and metastasis population dynamics on the vascular network reveals likelihood of primary self-seeding is low, and suggests secondary seeding as an alternate hypothesis.	17
2.3	Filter-flow.	18
2.4	The Red Queen hypothesis states that against evolution, as fast as we run (make new drugs), we'll never get ahead.	19
2.5	Constructing a Markov model of evolution on a hypercubic landscape from bacterial antibiotic response data provides clues to rational therapy sequencing.	20
2.6	Evolutionary games with three strategies have illustrated the effect of the timing, duration and sequence of therapies shaping the composition of heterogeneous tumours.	21
2.7	Stem cell driven tumours have characteristic growth curves in and out of therapy that can be used to infer parameters of an underlying model. . .	22
2.8	Microenvironmental variables must influence intrinsic phenotypic parameters of cancer stem cells to affect tumourigenicity.	23
2.9	Deriving GARD from a clinical sample.	24

2.10	GARD score distribution and density within putative 60 Gy (left) and 70 Gy (right) dose groups, by disease site.	25
3.1	Colony size distribution is observable in mouse models and colony shape is observable in clinical samples for some neutral mutations.	29
3.2	A Hierarchical model of stem cell-driven tissues.	31
3.3	Schematic for mutation in our model of neutral evolution.	32
3.4	Definition of lattice position neighbourhood for cell division.	34
3.5	Schematic of the discrete time updating algorithm for the hierarchical cellular automaton.	35
3.6	Temporal evolution of the spatial model reveals observable morphologic differences between TIC-driven and non-TIC-driven tumours, as observed by others.	36
3.7	Example simulations of the spatial cellular automaton model for different probabilities of symmetric division.	38
3.8	Spatial constraints on proliferation affect the Shannon index of colony size distribution across a range of symmetric division probabilities.	39
3.9	Effect size of changes in symmetric division probability on diversity between the spatial and non-spatial model.	40
3.10	Distributions of moments of colony size distribution for both spatial and non-spatial models.	41
3.11	Effect size of changes in symmetric division probability on the moments of colony size distribution between the spatial and non-spatial model.	42
3.12	Distribution of distance to clonal group centroid as a function of symmetric division probability.	44
3.13	Comparison of distribution of colony ellipse eccentricities for a range of symmetric division probabilities.	45
3.14	Method for biopsying the cellular automaton model.	47
3.15	Increasing biopsy size, with fixed position, decreases average deviation from the tumour Shannon index.	49
3.16	Individual biopsies taken from non-hierarchical tumours are better representations of true diversity than those from hierarchical tumours.	50
3.17	The accuracy of diversity prediction between hierarchical and non-hierarchical tumours changes with increasing number of biopsies.	51
3.18	Comparison of trade-off between biopsy size and number for different symmetric division probabilities.	52

4.1	Phylogenetic reconstruction in the study of evolution in infectious diseases and cancers.	56
4.2	Three example simulations with increasing symmetric division probability and their associated phylogenetic trees.	59
4.3	Example phylogenetic trees and their measures.	62
4.4	A summary of four tree indices measured over a range of symmetric division probability.	63
4.5	Effect size of symmetric division for four tree indices.	65
4.6	Comparing phylogenetic tree measures across symmetric division probability through tumour growth.	66
4.7	Comparing phylogenetic tree measures across symmetric division probability and mutation probability.	69
4.8	With clinically relevant sample numbers and sizes, inferred trees miss the majority of the complexity of the ‘true’ phylogeny.	71
5.1	A summary of oxygen-based cell fate thresholds.	81
5.2	Schematic of the discrete time updating algorithm for the hybrid cellular automaton model.	82
5.3	Definition of lattice position neighbourhood for cell division and oxygen diffusion.	83
5.4	Oxygen concentration plotted as a function of time and space in the computational model in comparison to the analytic solution.	87
5.5	Varying vascular density with regular spacing affects the carrying capacity and cell-oxygen distribution in simulated normal tissue.	90
5.6	Varying vascular density with irregular patterning affects the carrying capacity of normal tissue and the cell-oxygen distribution.	91
5.7	Increasing spacing between vessels slows tumour growth and creates areas of necrosis.	92
5.8	Qualitative parameter sensitivity: an analytic representation.	93
5.9	Size of necrotic area increases monotonically with oxygen uptake and distance between vessels.	95
5.10	Equilibrium cell density versus vessel density in the hybrid cellular automaton model.	98
5.11	Wide variation is present in cell-oxygen distribution between simulations with constant vessel density.	99
6.1	A schematic describing the difference between direct and indirect radiation damage.	105

6.2	A schematic describing the linear-quadratic model of radiation effect. . .	106
6.3	Surviving fraction is a non-linear function of oxygen concentration. . . .	107
6.4	Surviving cell number versus vessel density in the hybrid cellular automa- tion model.	110
6.5	Simulation results for a simplified two-vessel case.	111
6.6	Carrying capacity and cellular oxygen distribution are competing effects on cell survival after radiation.	112
6.7	Varying the vascular aggregation affects the carrying capacity of normal tissue independent of vascular density.	115
6.8	Examples of Voronoi polygon plots with Voronoi polygon area distributions.	116
6.9	Ripley's K edge correction.	118
6.10	Examples of Ripley's L for specific vessel distributions.	119
6.11	Aggregation coefficient versus surviving cells after radiation.	121
6.12	Voronoi polygon area distribution skewness versus surviving cells after ra- diation.	122
6.13	Ripley's L versus surviving cells after radiation.	123
6.14	Comparison of the correlation of the three measures of vascular patterning with radiation effect.	124
6.15	Normalising vascular patterning has opposite effects on radiation efficacy in low versus high vascular density.	125
6.16	Schematic of a method for spatially and temporally optimising radiation therapy treatment and VNT scheduling.	127
7.1	Signatures of positive selection can be seen in bulk sequencing data and phylogenies.	132
7.2	A simple model of a hierarchical tumour with irregular vasculature shows differing TIC dynamics in different TIC niches within the same tumour. .	134
A.1	Exhaustive exploration of biopsy size changes.	136
A.2	Plot of colony size versus frequency and mean distance to centroid versus colony size.	137
B.1	Tumour growth dynamics for a range of symmetric division rates.	138
B.2	Raw and pruned trees give rise to qualitatively similar summary measures with rank preserved.	144
C.1	Tumour growth in an avascular domain with oxygen diffusion from the outside displays characteristics of tumour 'spheroid' growth.	147

D.1	Aggregation measure versus carrying capacity.	150
D.2	Comparison of correlation of Voronoi polygon area distribution moments and carrying capacity.	151
D.3	VPA distribution skewness versus carrying capacity.	152
D.4	Ripley's L versus carrying capacity.	153

Chapter 1

Introduction

1.1 Why study mathematical oncology?

1.1.1 Cancer is horrible

My reason for entering into this course of study is to make a positive change in cancer research by bringing more mathematical thinking and modelling to clinically relevant research. Cancer is a disease that is growing in its reach, and affects people of all ages, colours, genders, socioeconomic status and cultural backgrounds. It dismantles families by taking loved ones and by breaking them financially due to the exorbitant cost of (often futile) treatments. Originating from our own cells, a cancer causes destruction where it arises, then spreads and causes multi-organ damage until it eventually kills us. Cancer exists in between multi-cellularity and unicellularity. It is more than an aberration of physiology, like diabetes or other errors in metabolism or growth; a cancer is almost a life form of its own, evolving in parallel with, but without regard for, the host. While a cancer shares and obeys many of the rules of biology, it is not a free living organism and therefore is exempt from many of the rules as well.

It is also important to remember, before we begin to discuss the facts and figures related to cancer, the concepts that drive our understanding of the biology, or the mathematical models that I will use to elucidate its mechanisms, the patients who suffer from it. As a radiation oncologist, during my time in the clinic I see a neverending line of patients, young and old, seeking treatment and hope. Each is as different from the next as are two snowflakes in myriad aspects. Each tumour is different, not just in histologic subtype, but in location, in aggressiveness, and in its underlying genetics. Each patient is affected differently: financially; psychologically; physically. Figure 1.1 depicts three patients whom I have treated in the last year, all of whom are deceased, and all of whom had different cancers. I chose these three, from the hundreds of cases I have treated, as reminders of the human aspects of this disease. Each of these three has a different form

of cancer, treated in different ways: but, all three had their faces, their most human body part, destroyed by the disease before it ultimately killed them.



Figure 1.1: Cancer is a highly heterogeneous, destructive and deadly disease. Every patient experiences cancer differently due to differences in location of their tumour, their tumour's biology, their own healthy tissue biology and the treatment. Here are three representative, deceased, patients of mine from the last year all experiencing cancer that affects their most human aspect: their face. From left to right there is a clear cell sarcoma affecting the floor of mouth, an angiosarcoma affecting the scalp and skull, and a merkel cell carcinoma diffusely affecting the skin of the face. All patients, before their death, signed consent for use of pictures taken during clinical care as educational materials, including publication.

1.1.2 Cancer incidence is rising and death rates are flat

Cancer has been moving rapidly up the list of causes of human death over the last century. We have made amazing strides in the treatment of infectious diseases through vaccines, sanitation and antibiotics; and of cardiovascular disease, through surgical and dietary advances; but, with few exceptions, we have done little to reduce the rate of cancer death to date. In the USA in particular, cancer has even risen to become the number one cause of death in some states, with continually rising incidence (Figure 1.2). A large proportion of the increase in cancer incidence is blamed on cigarette smoking, which is beginning to decline in the US, but is still increasing worldwide, and this trend is not predicted to slow down [117].

Oncology entered into a new era of 'precision medicine' in 2001, with the approval of imatinib for chronic myelogenous leukemia [62] (and only just recently for radiation therapy with an extension of classical radiobiological models to genomics [211]), in which therapies targeted to patients' individual tumour biology have promised to provide cures. This promise, however, has not yet come to fruition. This is largely due to two reasons. It

is now accepted that cancer is an evolutionary process [159], following the rules of selection and mutation first posited by Darwin [50]. The first result of this evolutionary process is, sadly, that cancers inevitably find solutions to whatever therapeutic intervention we throw at them. This is particularly true in the case of newer, targeted therapies [93], which target single aberrations within tumours. A second consequence of this evolutionary process is that cancers are highly heterogeneous. This is especially the case in solid tumours, where microenvironmental pressures change over the length scale of a tumour. Thus, while a targetable mutation may be present in a tumour, it is unlikely to be in every cell (and is usually in only a small proportion). So, while the targeted drug is highly efficacious in some proportion of the tumour, a cure does not result.

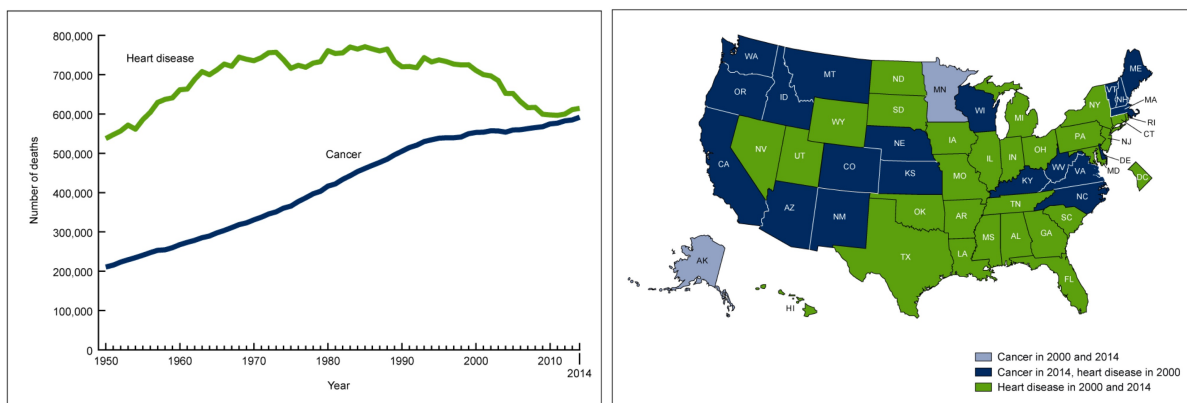


Figure 1.2: Cancer deaths are rising worldwide, and are nearing the number one cause of death in the USA. While modern medicine has solved many of the problems of the last century, we have made very few advances which have translated into higher cure rates for cancer. This disparity between cancer and other diseases means that cancer is rising on the list of killers, nearly equaling heart disease (left). In some states in the USA, cancer has become the leading cause of death (right). Image adapted from the United States Centers for Disease Control website (www.cdc.gov), which kindly provides all information as public domain.

1.1.3 A new perspective is needed

While the aforementioned reasons for our failures to produce cures in clinical oncology are becoming accepted, the standard practice in oncology research of searching for ‘druggable targets’ has not yet abated. Driven by a combination of skewed funding systems, the pressures of academia, and the false promise of finding a ‘silver bullet’ (not to mention getting rich because of finding it), this remains the mantra of oncologists the world over. It was this situation, much akin to the parable of the emperor’s new clothes, where many people understand the problem, but are unwilling to change their behaviour in response to that understanding, that drove me to mathematical oncology. It is here, outside of standard oncology research, where I believe the next fundamental advances will be made, and it is for these reasons that I have pursued this course of study.

Oncology requires a team science approach: While a new, mathematical perspective is sorely needed in oncology, the research requires such broad and deep knowledge of cell and molecular biology, clinical medicine and human physiology that carrying it out in a vacuum is no longer possible. Because of this, a strong dedication to team science is necessary. During the course of my thesis, while my mathematical skills were building, I served as the cancer expert for a number of mathematical oncology projects. Many of these projects have led to scientific publications, which I have co-authored. Since my contribution to these works was largely non-mathematical, I have chosen not to include them for examination as part of this thesis. However, to reflect their (and my) contribution to the larger mathematical oncology literature, I present a survey of that research in Chapter 2.

1.2 Cancer is heterogeneous on many scales

In recent years it has become increasingly important to understand the heterogeneity that exists in cancers [118]. This heterogeneity occurs at many scales: at the population scale between patients; at the molecular scale within tumours; between lesions within the same patient, in the case of metastatic disease [90]; between individual cells at the genotypic and phenotypic level [18, 107, 152]; and, in the case of the microenvironment, in stromal organization and chemical cues [142]. This heterogeneity extends beyond differences in cell type and genetic makeup and involves changes in spatial organisation on both the microscopic [15] and macroscopic [72] scales (see Figure 1.3).

These different scales of heterogeneity emerge from different mechanisms. They include different macroenvironments in the case of metastases in different organs and microenvironments in the case of invading solid tumours. They also include cell-specific changes such as varying cell division rates and varying mutation rates due to impaired error repair mechanisms, and tissue-specific changes such as different architectures and cell types across organs. The combined effects of increased cell division, increased mutation rate and selection are captured by the concept of clonal evolution [94], which we address in Section 1.5.

1.2.1 Microenvironmental heterogeneity

At the cell scale, microenvironmental heterogeneity is becoming widely accepted as a key player in tumour progression and therapeutic response [118, 152]. Nutrients, growth factors, extracellular matrix and other cell types are all part of the normal tissue that surrounds and pervades a solid tumour, each of which has been shown to vary widely across

different tumour stages and types. This is, in part, due to the interplay between the tumour and its microenvironment, each influencing the other in a spatially and temporally heterogeneous manner. A schematic of a tumour growing in the presence of non-cancerous elements is depicted in Figure 1.3, illustrating the many elements comprising the heterogeneous milieu, including stromal cells, blood vessels, healthy surrounding tissue, and extracellular matrix [15].

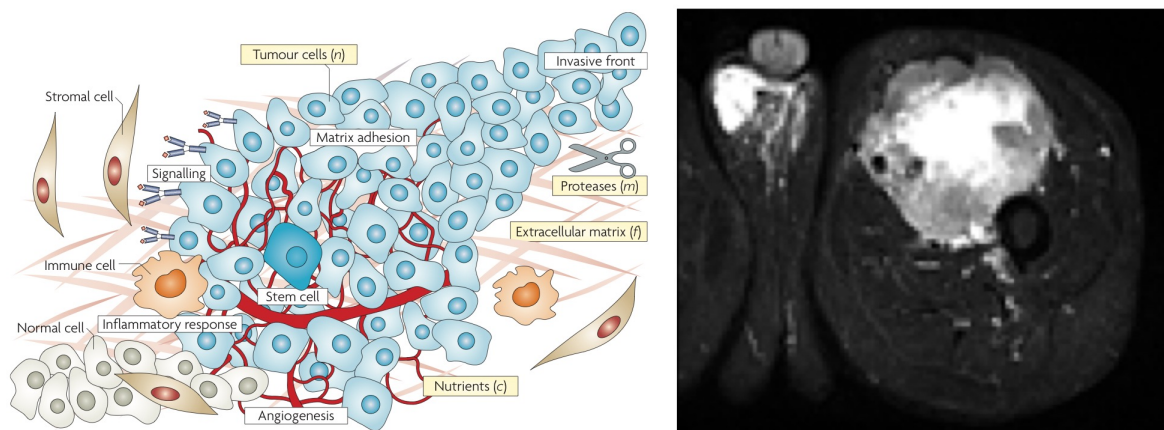


Figure 1.3: Cancer is not just a collection of cells, but a complex, organised combination of healthy and cancerous components which exhibit phenotypic and spatial heterogeneity over multiple spatial and temporal scales. Different types of tumour cells interact with healthy stromal cells, vasculature, healthy normal cells and a plethora of extracellular components on a length scale the order of a cell (left). Tumours also exhibit spatial heterogeneity on longer length scales, as is evidenced by many imaging modalities. On the right we show an example of a highly heterogeneous soft tissue sarcoma using Magnetic Resonance Imaging (MRI), where the tumour in the anterior thigh enhances over a wide range. Figures adapted from [15] (left) and [72] (right), with permission.

Many aspects of microenvironmental tumour heterogeneity have been studied alone and in combination using mathematical models. Complex systems like tumours are well suited for mathematical modelling. To test biological hypotheses experimentally quickly becomes intractable when more than only a few elements are included, whereas experiments simple enough to be done are often too simple to be enlightening. This is where mathematical models show their strengths, through either computational power or analysis, these models are able to dissect the relative contributions of different components. While the results of mathematical models are rarely the end of the story, they can provide insights that allow experimentalists to significantly reduce the space in which they need to search, and can provide hypotheses that would otherwise have gone untested.

Here, we provide a brief overview of some of these efforts for individual aspects of microenvironmental heterogeneity. We delve more deeply into individual aspects as they become relevant in the body of the thesis. For thorough, recent reviews of mathemat-

ical oncology as a whole, the work of Altrock et al. [11], Byrne [33], and Rejniak and Anderson [194] are excellent resources.

Altered metabolism: It has been long known that cancer cells have altered metabolisms compared to normal tissue. This results in changes in the way sugars are utilised, as well as other alterations, and can change division rates of cancer cells. The Warburg effect is a physiologic shift towards a glycolytic phenotype that is triggered by lack of oxygen but is maintained when normoxia returns, and was first posited by Otto Warburg as a central component of cancer in 1956 [240]. This sentiment has been echoed over the years and has more recently been formalised as a ‘Hallmark of Cancer’ by Hanahan and Weinberg [101, 102]. How this energetically less fit phenotype would persist, and grow to dominate a tumour, is not intuitively obvious. In addition to affecting the way that cancer cells metabolise sugars, the Warburg shift also creates a situation in which the cells produce an excess of lactic acid, which slowly poisons the surrounding tissue. Many theoretical studies have sought to better understand how this shift, which appears on the surface to be detrimental, could be so pervasive. One of the earliest works modelling the effect of acid production in tumours was by Gatenby and Gawlinski [81], who built a partial differential equation (PDE) model of tumour growth in the presence of healthy tissue. They assumed that tumour cells could produce lactic acid, and that this acid allowed them a competitive advantage against healthy tissue and subsequently promoted invasion. Their model predicted travelling waves of acid leading an invading front of tumour cells and they showed regions in parameter space in which this competition favoured the tumour and found evidence in pathologic samples of this phenomenon. The authors termed this the acid-mediated invasion hypothesis. This study was followed by Patel et al. [184] who created a hybrid cellular automaton (CA) model of a tumour growing in a vascularised domain. They too modelled the production of acid by tumour cells, but were able to show a delicate dependence of the production parameters from the earlier study on vessel density, a finding later validated in a three-dimensional model by Smallbone et al. [217].

Microenvironmental influence on phenotype: The importance of acid having been established for tumour growth, at least theoretically [22], the questions being asked turned toward how these cells came to become acid producers. To answer this question, Gerlee and Anderson [87, 88] built a hybrid CA model of cancer growth that included oxygen and glucose diffusing from the outside of the tumour. Further, they modelled the stiffness of the microenvironment, as tumours are known to be stiffer than surrounding tissue, due to a process called desmoplasia which involves the creation of extra-cellular

matrix by tumour cells as well as the surrounding healthy cells. The authors modelled cells as agents, each of which had a neural network within it to make cell fate decisions based on the environmental conditions, and could therefore adopt a number of phenotypes. They found that the glycolytic phenotype emerged most reliably and quickly when the environment was stiffest, and when nutrients were most scarce. These results, and those of others [14, 16], suggest the multi-step process of carcinogenesis, based on phenotypic and microenvironmental change is well within the plastic capabilities of a cell, i.e. without need for genetic mutation.

Angiogenesis: While many models assume that nutrients diffuse into a tumour from the outside, like in a tumour spheroid experiment, it is well known that tumours, like normal tissue, are vascularised. The quantity, quality and even source of the vasculature within tumours, however, are all open questions in oncology. The creation of vessels in healthy tissue, and similarly in tumours, is predominated by a process called angiogenesis. This is a topic on which immeasurable biological study has been completed and has also been an active topic in theoretical studies. For a recent review of the state of our biological knowledge and the role of treatment in this field, see Yadav et al. [250]. From a mathematical standpoint both continuous and discrete approximations of new vessel growth have been employed [13, 231]. Questions studied have included how blood flow dynamics change in the face of changing vascular trees and its effect on therapeutic delivery [153, 154, 185, 230]; the role of angiogenesis in tumour progression [229, 232]; and the effects of vascular pruning [180], sometimes called vascular normalisation [113], another mechanism by which the same biological process can improve, rather than just add to, vascular networks.

1.2.2 Proliferative heterogeneity: the cancer stem cell hypothesis

Beyond heterogeneity in the microenvironment surrounding cells within a tumour, there exists significant heterogeneity between cancer cells within the same tumour. This heterogeneity occurs at both the phenotypic and genetic scales. One aspect of phenotypic heterogeneity that has been studied in great detail is the difference between putative cancer stem cells, often termed tumour initiating cells (TICs) and their non-stem, or transit amplifying cell (TAC) progeny. TICs have been shown to be highly tumourigenic in xenograft models and have some properties of normal stem cells [158]. In a simplified view of the tumour cell hierarchy, TICs can divide symmetrically, to produce two TIC daughters, or asymmetrically, to produce a TIC daughter and a more differentiated progeny [166, 195]. The hypothesis that cancers follow this hierarchical organisation is

termed the cancer stem cell hypothesis (CSCH), and stands in contrast to the previously espoused stochastic model in which all cells had a low probability of recapitulating the tumour, but were otherwise identical. These two models are schematised in Figure 1.4.

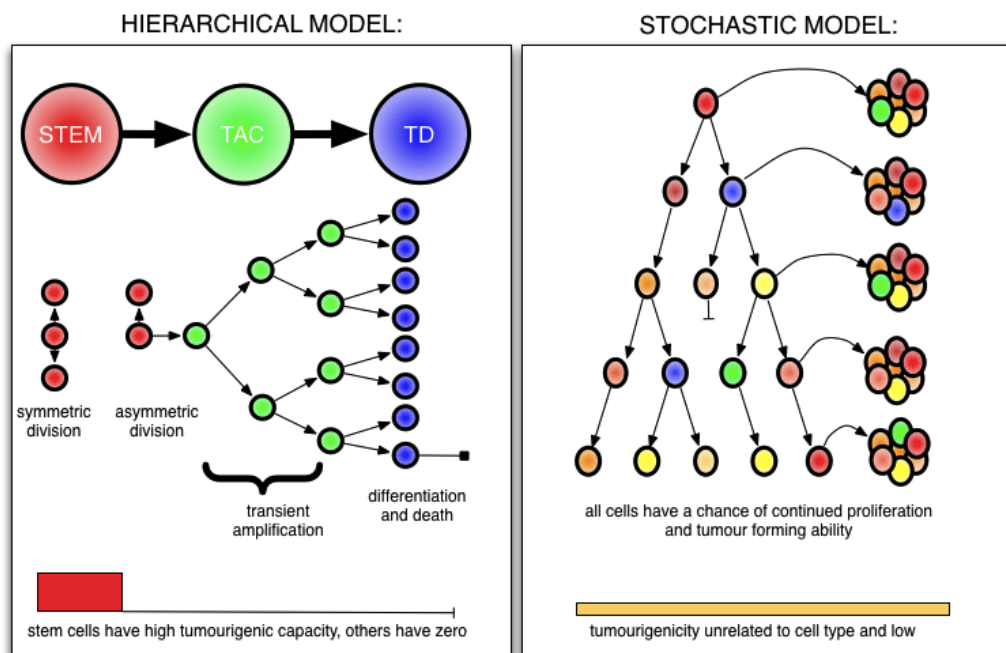


Figure 1.4: A comparison between the cancer stem cell hypothesis (left) and the stochastic hypothesis (right). In the CSCH, only a small subset of tumour cells, the TICs, are capable of infinite replication, and therefore recapitulation of the tumour phenotype. TICs can divide symmetrically, resulting in two stem daughters, or asymmetrically, resulting in a stem daughter and a non-stem, TAC daughter. TACs are also capable of symmetric division, resulting in two TAC daughters, but only for a certain number of divisions before they differentiate into terminally differentiated cells which are incapable of further division.

TICs, first identified in leukemia [30], have also been implicated in many other cancers, including, but not limited to, brain [216], breast [1], head and neck [193], lung [67], melanoma [144], colon [175, 196, 251] and others. These tumour types can be broadly classed as ‘hierarchical’ in the sense that they are posited to have a hierarchical proliferative organisation similar to normal adult tissues with well-defined stem cell populations. In these hierarchical tumours, TICs can divide to give cells with reduced proliferative capacity or self-renew to promote tumour maintenance. As TICs have been demonstrated to be resistant to a wide variety of therapies including radiation [23, 55] and chemotherapy [42, 51], and to play a role in the development of metastasis [183], the TIC hypothesis has important implications for patient treatment [57]. In particular, the effect of current strategies on the tumour cell hierarchy should be established, as therapies targeting TICs specifically are likely to provide greater benefits for cancer patients [58, 135]. While the evidence for the CSCH continues to build, its existence continues to be an open question in the community, and the model is under active revision [237].

1.2.3 Intratumoural heterogeneity: clonal evolution

While the CSCH is debated, the fact that cancers contain cells with many different genotypes is not. The fact that cancer is an evolving and adaptive process undergoing clonal evolution has been known for several decades [173]. The original model of clonal evolution considers lineages of cancer cells that acquire mutations in a sequential manner, each fitter clone then taking over a large fraction of the tumour bulk [94, 173]. This description of clonal sweeps does not allow for significant heterogeneity, however, and thus is gradually being usurped by a model which allows a wider variety of subclones, now detectable with modern sequencing technologies, which may fix, but do not necessarily sweep to domination [108]. The most recent iteration of the model, dubbed the ‘Big Bang’ model [221], posits that many of the important, ‘driver’, mutations occur early in tumour development. In this model, which was made possible through the advent of multiregion sequencing (see Figure 1.6, lower left), subsequent beneficial mutations may never have the opportunity to become significant proportions of the population, a phenomenon dubbed ‘survival of the first, not the fittest’. These three models of clonal evolution are schematised in Figure 1.5.

Each of these models depends not just on mutation, but also on selection forces to determine which are the fittest clones to sweep to fixation in each population. While this form of selective, Darwinian evolution is now widely accepted in cancer [189, 207], neutral mutation still plays a significant role in the development of heterogeneity [177]. For example, Williams et al. [246] have recently shown that spatial sampling indicates that many tumours exhibit neutral, or nearly neutral, dynamics for the majority of their history. The neutral theory of evolution, first posited in 1968 by Kimura [123], states that the vast majority of genetic alterations that accrue within an individual (cell/organism) are neutral; that is, they have no direct effect on fitness. While it is clear in cancer that some mutations are beneficial (drivers) and some are deleterious (passengers) [155], the relative rates of each type of mutation are unknown, and the community has yet to come to a consensus.

Measuring evolutionary histories with phylogenetics

To describe the evolutionary relationship between members of a species, or larger groups of life forms, biologists often formulate tree diagrams that represent their specific hypotheses about relatedness. While tree diagrams have been in use since medieval times to describe genealogies, their use to describe animal species was not popularized until the early 1800s. The most famous of these is likely the drawing by Charles Darwin in his notes that led to the *Origin of a Species* (Figure 1.6, upper left). This method has since

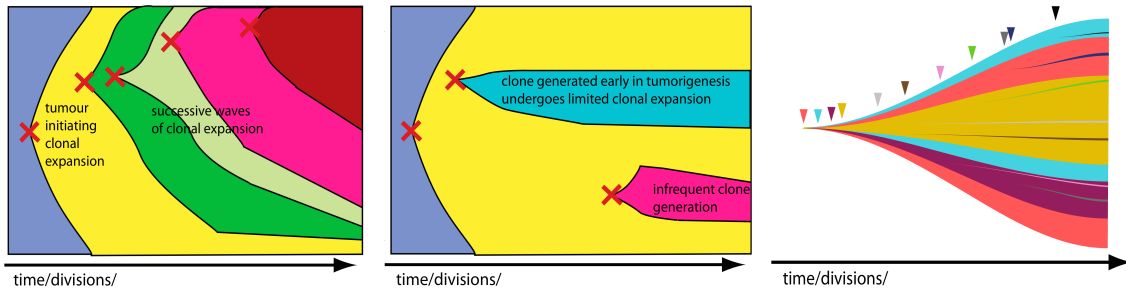


Figure 1.5: The evolution of our thinking about clonal evolution in cancer. The long and widely held paradigm for clonal evolution, often dubbed the ‘clonal sweep’ model (left) posited that each subsequent fitter mutation would (eventually) sweep to fixation in the tumour population. As the ability to sequence tumours improved, the paradigm evolved, and the conceptual model began to change to include more limited clonal expansions (middle), with the end state being more heterogeneous. Most recently, evidence from spatial sampling in colorectal cancers has led to the hypothesis that much of the meaningful evolution takes place at the beginning in a single ‘big bang’ (right) leading to a scenario where many of the measurable mutations which occur do so very early in the evolution of the tumour. Figure adapted from [108] and [221], with permission.

been widely adopted and we can now place ourselves in a tree diagram for all of life in the ‘tree of life’ (Figure 1.6, upper right).

While these trees were originally made on the basis of gross morphological differences (or similarities) and were called phenograms or cladograms, in the past several decades we have begun to define these differences based on genetic information. The field of phylogenetics, born in the 1980s, seeks to use objective, genetic information to build trees. When populations are sampled, a common method of understanding the clonal evolution is through phylogenetic reconstruction, a method of inferring, usually from genetic sequence similarity, the evolutionary life history of a given life form. This has classically been applied in scientific fields such as zoology, and it has become a branch of bioinformatics all of its own, even spawning a branch of discrete mathematics called T-theory [60].

This field of phylogenetics has, in the last decade, begun to be applied to cancers, giving rise to subfield recently dubbed ‘PhyloOncology’ by Somarelli and colleagues [220]. The application of PhyloOncology to (largely) genetic data in cancer seeks to answer a number of questions: (1) How do tumours across a cohort of patients cluster genetically together (and what does this teach us about the biology, or clinical outcome) [222]?; (2) From where, and when, did spatially distinct metastatic lesions arise [90, 255]?; and, (3) what mutations confer selective advantages, and how much [35, 247]?

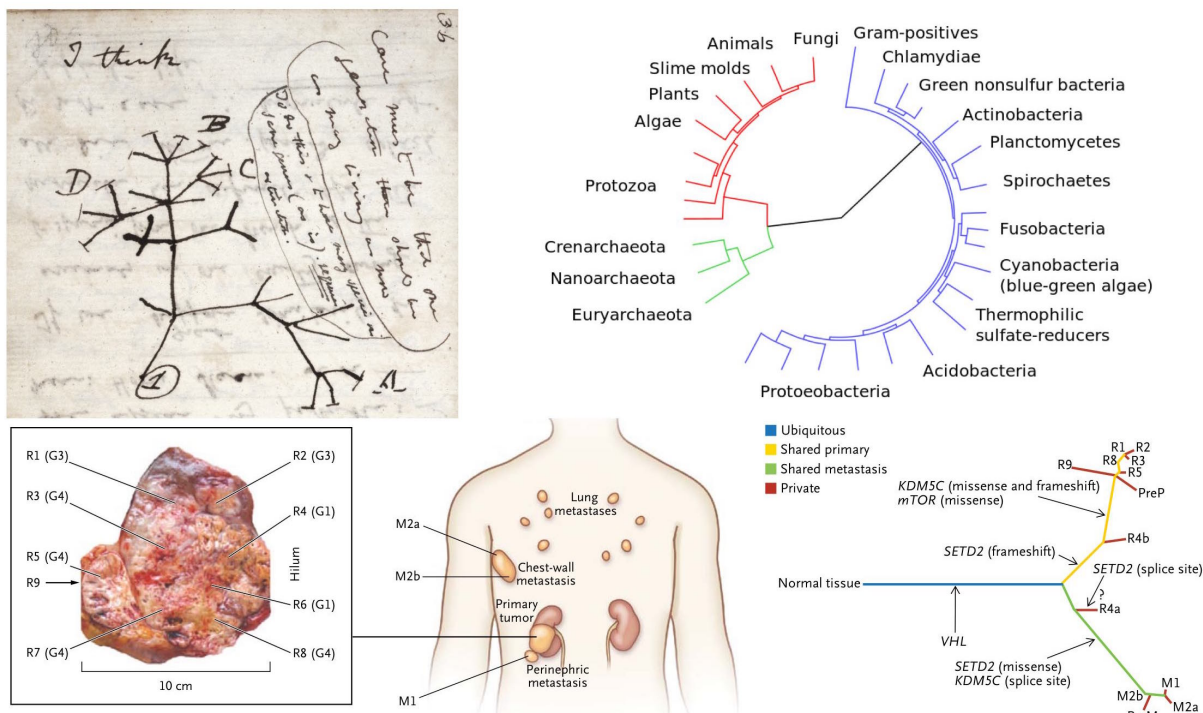


Figure 1.6: Branched evolution from Darwin to cancer. The history of branched evolution in biology and now oncology. Upper left, we show the first example of an evolutionary tree as sketched by Charles Darwin in preparation for his *Origin of a Species*. Upper right we show one representation of the ‘tree of life’ in which the relationship between all living things is represented. The two lower images show the first use of these methods, together with multi-region tumour biopsies (lower left) and next generation sequencing, to create a phylogenetic life history of a cancer in a single patient (lower right). Figures adapted from Darwin [50], Wikipedia and Gerlinger [90].

1.3 Scope of thesis

In this thesis, we aim to use mathematical models to develop clinically translatable biomarkers to improve our understanding, and eventually, our treatment, of cancer. We focus on building simple models of tumour growth using the previously reviewed concepts of microenvironmental heterogeneity, proliferative heterogeneity (in the form of the CSCH) and clonal evolution. After giving an overview, in Chapter 2, of my broader contributions to mathematical oncology, I will present the work upon which I rest my application for a Doctor of Philosophy in mathematics. We proceed in two parts, each composed of two research chapters. While we have given a brief overview of the relevant biological background and previous modelling work that preceded our work in the earlier sections of this introduction, each chapter will begin with a literature review more specific to the work presented therein.

Part I: In the first part of the thesis, we work towards methods of using clinically available data from tissue biopsies to answer two questions: (1) is the cellular proliferation within a given tumour hierarchically organised?; and, if so, (2) what is the underlying symmetric division probability? The answers to these questions will help oncologists better classify tumours into ‘stem-cell-driven’ or ‘non stem-cell-driven’ categories, which will move our understanding forward and also suggest in which situations certain therapies would work better or worse. Being able to measure relative symmetric division probabilities between two samples, either of the same cancer at two different time points, or between two different cancers, will allow for further classification, as above, but also allow us to understand the effect of therapies on symmetric division probability.

To answer these two questions, we build and analyse a pair of mathematical models of tumour growth with proliferative heterogeneity and neutral clonal evolution in Chapter 3, with special consideration paid to the resultant diversity, clonal dynamics and spatial patterns. Then, in Chapter 4, we build an analytic pipeline to reconstruct the phylogenetic trees which result from our simulations and ask what deeper insight we can gain from these structures.

Chapter 3: Inferring tumour proliferative organisation from measures of diversity and spatial heterogeneity

In Chapter 3, we encode the classic CSCH into a non-spatial, individual-based stochastic model with neutral mutation. We then extend this model to explicitly consider the inclusion of spatial constraints on growth by embedding the model in a two-dimensional lattice. The spatial model follows the same rules as the non-spatial model, with the exception of contact inhibition. We then explore the effect of varying symmetric division probability in the two models on clonal diversity and colony size distribution, and find that they behave quite differently. Of note, we find that different conclusions can be drawn from each analysis, depending on whether the simulation is spatially constrained (like a solid tumour) or not (like a liquid tumour), suggesting that data from these tumour types should be viewed quite differently. In particular, we find that diversity measures are sufficient to discern the symmetric division probabilities for non-spatially constrained tumours, but not in spatially constrained ones. We end Chapter 3 by introducing a method to ‘biopsy’ the simulation and ask whether the conclusions we drew about the differences between the models, and the underlying symmetric division probabilities, can still be observed.

Chapter 4: Inferring symmetric division probability from phylogenetic analyses in a stochastic model.

In Chapter 4 we work to discern between differing symmetric division probabilities in our spatial model using measures that could be applied to currently available data. After creating an algorithm to generate phylogenetic trees from our simulation outputs, we explore a number of statistical measures of the topology of the resulting phylogenetic trees. We find that unlike the diversity, and spatial colony shape and size metrics from Chapter 3, several of the statistical measures of tree topology from the spatial model are able to help us discern between simulations based on symmetric division probability. Encouragingly, our ability to discern between simulations is robust to changes in tumour size and mutation rate. Disappointingly, however, we are unable to discern between the trees which we infer from biopsies of the simulations only – and we leave this problem to future work.

Part II: In the second part of the thesis, we turn to the issue of microenvironmental heterogeneity, in the form of oxygen, with an eye to understanding radiation sensitivity changes. The majority of the work in these two chapters was published in *PLoS Computational Biology*, in January of 2016, as a manuscript titled “Spatial Metrics of Tumour Vascular Organisation Predict Radiation Efficacy in a Computational Model” [212].

Chapter 5: Quantifying microenvironmental heterogeneity in a hybrid cellular automaton model of cancer in a vascularised domain

In Chapter 5, we extend our spatial CA model to include both a continuous approximation of oxygen, diffusing from vessels and taken up by cells, and a field of surrounding, healthy tissue. After carefully parameterising our model, we examine the effect of vascular density, in regularly organised domains, on tumour growth velocity. We find several dynamic regimes, and perform simplified analysis to understand the relationships between parameters governing these regimes. We then move on to observing a summary metric of oxygenation, which we term the cell-oxygen distribution. This distribution describes the number of cells in a given domain that experience a given oxygen concentration. We find that, in domains with equal vascular density, but different vascular patterning, there is a wide range in the carrying capacity (number of cells supported), as well as significant changes in the cell-oxygen distribution. Both of these quantities, the carrying capacity and the cell-oxygen distribution, have direct bearing on the probability of success after radiation therapy, which we explore in Chapter 6.

Chapter 6: The effects of vascular patterning on radiation treatment efficacy

In Chapter 6, we explore the effect of the observations concerning cell-oxygen distribution changes in Chapter 5 given differing vascular patterning on radiation efficacy. Here, we develop an algorithm to measure the surviving fraction of a population of cells experiencing a given distribution of oxygen, after a given dose of radiation. To make this calculation, we consider the well studied ‘oxygen enhancement ratio’, which is an increasing function of cell kill with increasing oxygen concentration. We then explore the relationship between the number of surviving cells after radiation and several spatial measures of vascular organisation. For all measures, we find the same pattern of correlation: we find that the relationship between organisation and cell kill is opposite for high and low vessel densities. This suggests that vascular normalisation therapy (VNT), which is a novel explanation for how drugs acting to block the angiogenesis pathways, could have opposing effects – even within the same tumour if there existed areas of both high and low vascular density. Further, it suggests a method of temporally and spatially scheduling radiation and VNT to optimise therapeutic response.

Conclusions

We close the thesis with some final remarks in Chapter 7 highlighting opportunities to strengthen our results, and to move forward.

Chapter 2

Contributions to mathematical oncology as a member of a larger team

2.1 Introduction

As discussed in Chapter 1, cancer research is a complex and wide ranging discipline. To apply mathematical reasoning and techniques to this research requires multi-disciplinary teams with members from very disparate fields including, but not limited to: clinical oncology, cell and molecular biology, chemistry and genetics, in addition to mathematical modellers. The field of mathematical oncology, then, is by definition one that requires a team science approach. During the course of my work on this thesis, in addition to honing my own mathematical modelling skills, and pursuing the work presented in Chapters 3-6, I also participated in several other mathematical oncology endeavours as part of a larger team. While my contribution to most of these was significant, I prefer to have my official thesis comprised of work I completed in entirety with my own hands. That being said, because of the impact of this work on the larger mathematical and clinical oncology communities, I highlight some of it here.

2.2 Metastasis on networks

The process of metastasis, where cells break away from the primary tumour and travel through the blood (hematogenous) or lymphatic system to colonise distant sites, is the cause of $\approx 90\%$ of cancer mortality. I have published a book chapter on the mathematical modelling endeavours in this field [210]. Together with collaborators, I proposed and explored a novel view of the hematogenous spread of cancer based on the network comprised of the veins and arteries of the human vascular system. Beginning with a perspective published in *Nature Reviews Cancer* [206], I suggested, together with my

co-authors Alexander Anderson, Peter Kuhn and Paul Newton, that this viewpoint could help explain some of the patterns of metastasis that we see by taking into consideration both the biological heterogeneity and physical factors present in each organ's capillary bed, schematised in Figure 2.1.

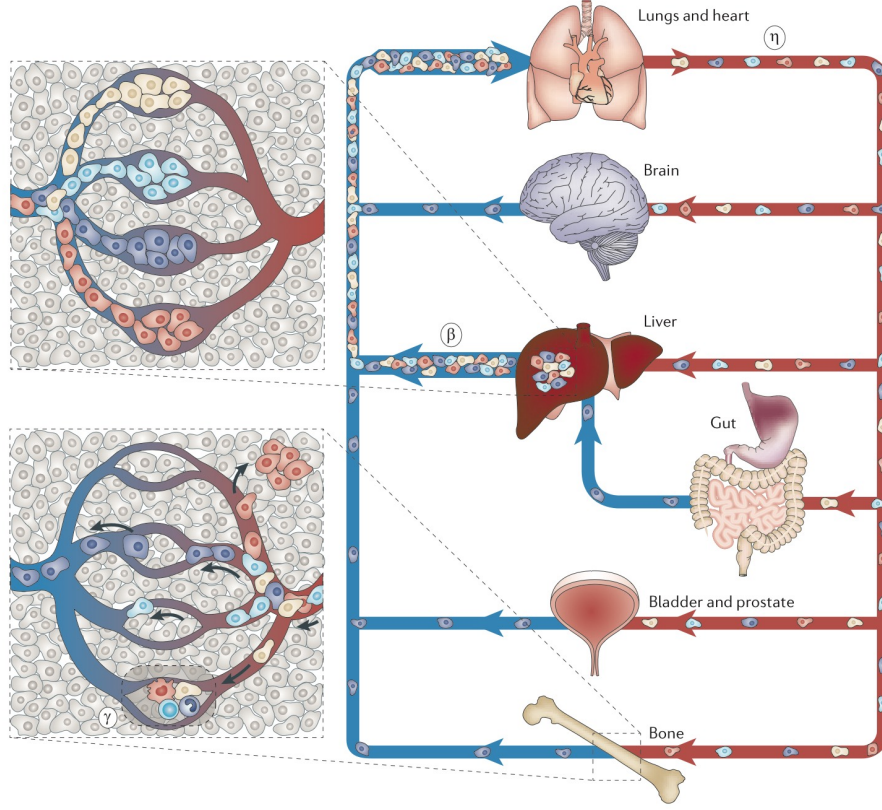


Figure 2.1: A simplified representation of the human vascular system connecting representative organs with a primary tumour in the liver. Four phenotypes representing putatively different metastatic clones are represented by the four different colours. The shedding term, β (middle), represents the rate and type of cells exiting the primary tumour. The clearance rate, γ (lower left), represents the rate at which cells are removed from the system by an immune response or other processes. The filtration fraction, η (upper right), represents the proportion of cells that arrest in an organ per pass and can be thought of as the concentration of cells coming out of an organ divided by the concentration going into the organ. In this example, the red cells are the proper ‘seed’ for the ‘soil’ of the bone, and are shown successfully growing after extravasation. The dark blue cells, which are able to flow through capillary beds, represent a circulator phenotype. Figure adapted from Scott et al. [206].

After proposing this simple network-based representation, we endeavoured to use the idea, and the parameters we posited, including shedding rate from the primary tumour (β), filtration fraction in a given capillary bed (η) and clearance rate (γ) by the immune system, together with a simple model of population growth to tease apart two competing hypotheses [208]. The two hypotheses were formulated to explain experimental results which showed, convincingly, that cells that originated in a primary tumour could leave the primary, enter the circulation, and then make it back to the primary tumour [122].

The first hypothesis suggested that shed cells could come back to the primary tumour and ‘re-seed’ it in a new spatial location (though still in the primary site) and benefit the primary tumour by boosting its growth rate. The second, competing hypothesis, was that this process actually happened through secondary metastatic lesions, seeded by the primary, which themselves could grow and shed cells to secondarily boost the primary tumour growth. Through an exploration of a biologically plausible parameter space, including error propagation, we found that the ‘secondary-seeding’ hypothesis was far more likely. An overview of our model assumptions, and the range of uncertainty for each hypothesis, is shown in Figure 2.2.

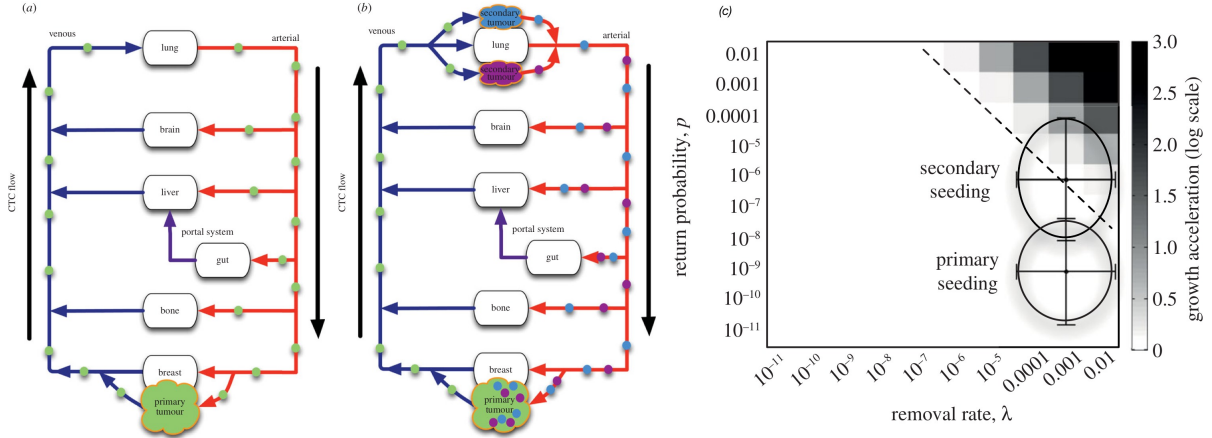


Figure 2.2: Studying tumour and metastasis populations dynamics on the vascular network reveals likelihood of primary self-seeding is low, and suggests secondary seeding as an alternate hypothesis. (a,b) We represent the primary and secondary seeding hypotheses on our simplified vascular network. (c) After modelling population dynamics and propagating error, we find that secondary seeding is far more likely to drive primary tumour growth. Figure adapted from Scott et al. [208].

After exploring the seeding hypotheses, we sought to separate the physical aspects of blood flow and filtration in the capillary beds – which we dubbed the “filter-flow model” [209] – from the biological “seed and soil” model which was first posited at the end of the 19th century [73]. To do this, we built upon our population dynamics model in the previously discussed paper, and also tuned it to be sensitive to blood flow to individual organs. We then used our network model of metastasis, combined with the blood flow data taken from the physiology literature and defined a metric, which we called ‘Metastatic Efficiency Index’ (MEI). The MEI is the ratio of metastatic incidence from a given primary tumour site (i) to a given organ (j), N_{ij} , taken from autopsy data, to circulating tumour cell outflow in the same direction (ϕ_{ij}), derived from our population dynamics model and blood flow data. We then considered cases where, like in the previous paper [208], subclinical secondary metastatic lesions existed, and observed how the MEIs changed. The resultant MEIs, with and without subclinical metastases, were compared to

previously published indices of metastatic efficiency taken from purely biological datasets. We found striking differences in prediction of MEI for organ-metastatic site pairs that had the liver in between, and otherwise reasonable agreement. Our results suggest that neither the physical, nor the biological information alone are sufficient to predict metastatic patterns.

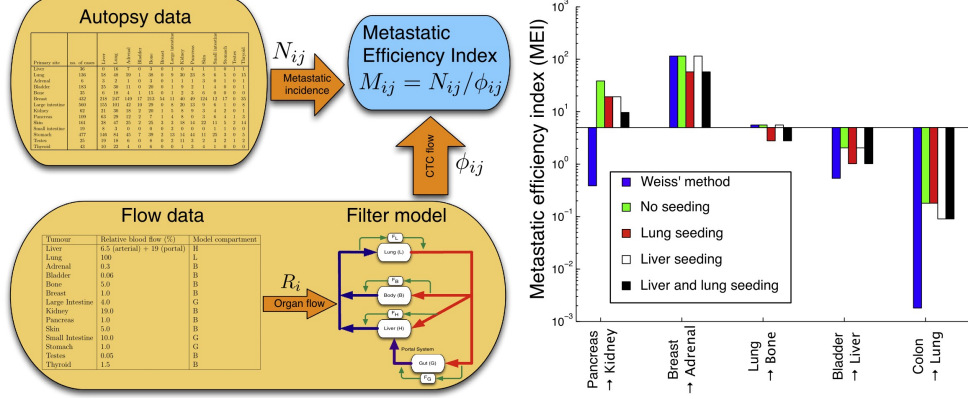


Figure 2.3: Filter-flow. We introduced a new term, the ‘Metastatic Efficiency Index’ (MEI), which is defined as the ratio of metastatic incidence from a given primary tumour site (i) to a given organ (j), N_{ij} , taken from autopsy data, to circulating tumour cell outflow in the same direction (ϕ_{ij}), derived from our population dynamics model and blood flow data (Left). We then compared several different primary-metastasis organ pairs and found striking differences between the predictions for metastasis based on existence of subclinical disease (Right). Figure adapted from Scott et al. [209].

2.3 Learning and perturbing evolution

Beyond the process of metastasis itself, the problem of drug resistance presents one of the most major challenges in oncology, and also in infectious disease. In both disciplines, very effective medical treatments are already in use, but in each case the tumour, or the infection, is able to evolve resistance to those drugs – eventually causing the death of the host. While many people are working on this problem in the clinic, the research usually conforms to a common pattern. Mutations are identified that correlate with (or cause) resistance in patients. The consequences of those mutations are identified, and a drug is found or, in some cases designed, to target those consequences, and a clinical trial is born. While this approach does work, on occasion, and new drugs are found that can prolong survival, it is a never-ending race. Like the Queen of Hearts told Alice, in Louis Carroll’s *Through the Looking Glass* [38]: “it takes all the running you can do, to keep in the same place”; no matter how fast we make new drugs, the diseases will figure out an answer. Applied to cancer, and infectious disease, this interpretation of Van Valen’s Red

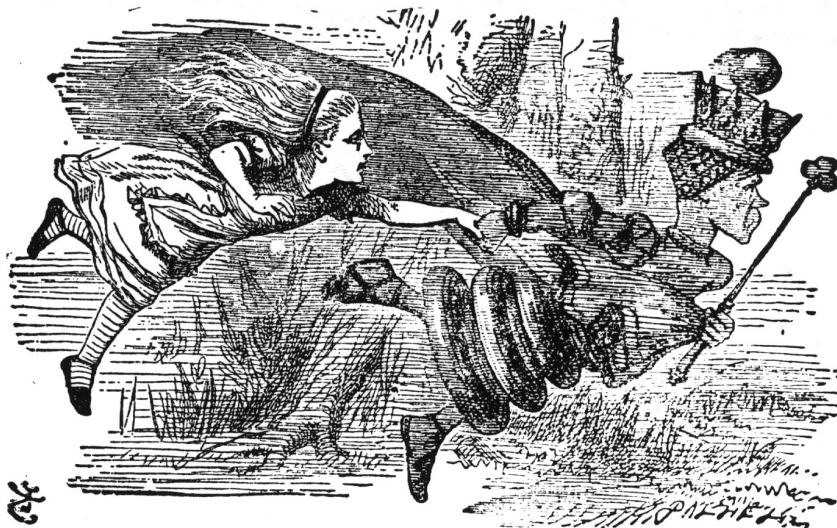


Figure 2.4: The Red Queen hypothesis states that against evolution, as fast as we run (make new drugs), we’ll never get ahead. Credit: Illustration by Sir John Tenniel from Lewis Carroll’s *Through the Looking Glass* [38], 1871, reproduced from Wikipedia, public domain.

Queen hypothesis [235] is just as appropriate as in the context of speciation in which he first postulated it.

I propose instead that we study the process of the evolution of resistance itself. There are several mathematical formalisms that lend themselves toward studying evolution of populations. During the course of my thesis studies, I participated in two longer term (ongoing) collaborative projects studying this evolution of resistance. The first of these projects involved using a Markov chain model to study evolution on sequentially ordered fitness landscapes, and the second, intratumoural competition using evolutionary game theory, also considering sequentially ordered therapies, and exploring the role of spatial constraints in this competition.

2.3.1 Markov model

In the first of these works, I led a collaborative effort with an Oxford student, Dan Nichol as first author, together with other members of Oxford Computer Science (Peter Jeavons) and the Moffitt Cancer Center (Alexander Anderson and Robert Gatenby) and the Cleveland Veteran’s hospital (Robert Bonomo) to study antibiotic therapy using the fitness landscape metaphor. Daniel Nichol, has presented this work for his thesis, so I will only review it here. Briefly, we constructed a Markov chain model of evolution by considering a biased random walk on N -dimensional hypercubes representing the genotype of a given population under selection by a given drug. Using data measured empirically for

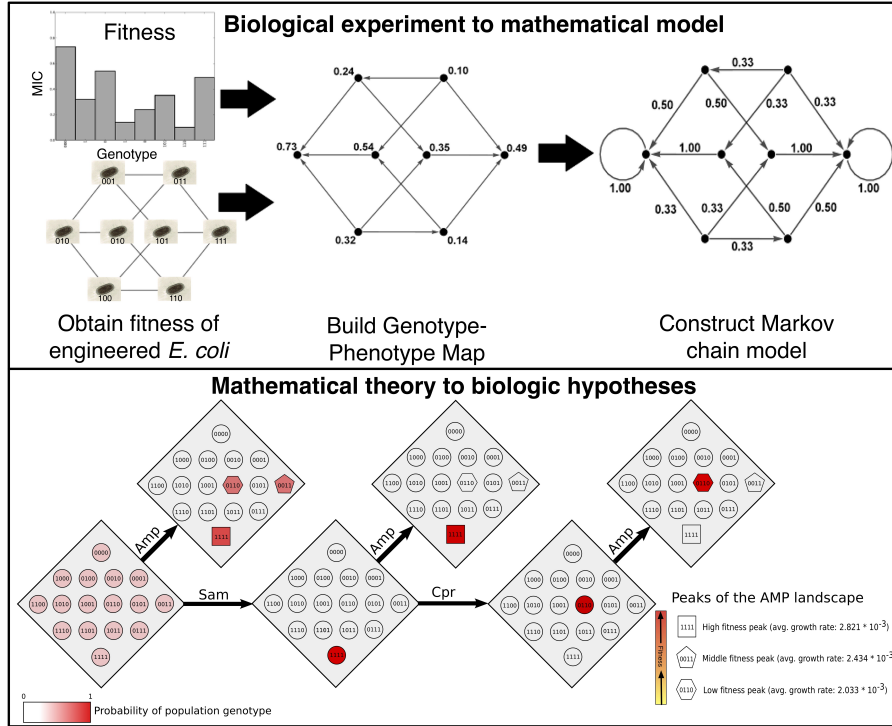


Figure 2.5: Constructing a Markov model of evolution on a hypercubic landscape from bacterial antibiotic response data provides clues to rational therapy sequencing. We used data from bacterial growth experiments in a large set of antibiotics to parameterise a Markov model of evolution (top) [171]. While we found that most combinations of antibiotics sped up the emergence of resistance, we were able to find optimal sequences of antibiotic therapy to minimise the evolution of resistance (bottom). Reproduced from Nichol et al. [171], available through Creative Commons.

15 β -lactam antibiotics by Mira [161] we constructed 4-allele fitness landscapes. Invoking the strong selection, weak mutation (SSWM) assumption [92], we considered only a monotypic population (single genotype), but were able to exhaustively enumerate all possible 2-, 3-, and 4-drug sequences. While we found that the majority of drug sequences increased the probability of resistance emerging, we also found that through rational ordering, the emergence of resistance could be reliably avoided [171]. A schematic of this method, and a rational sequence we found, is shown in Figure 2.5.

2.3.2 Evolutionary game theory

In an attempt to understand how the order (as well as timing) of therapy influences tumour evolution, I worked with another set of collaborators, chiefly David Basanta and Alexander Anderson from the Moffitt Cancer Center, to build an evolutionary game theory (EGT) model of this process. By considering a group of competing players, or strategies, we define the fitness that each has in the presence of a given amount of each other. This was then simplified by normalising each individual population by the total

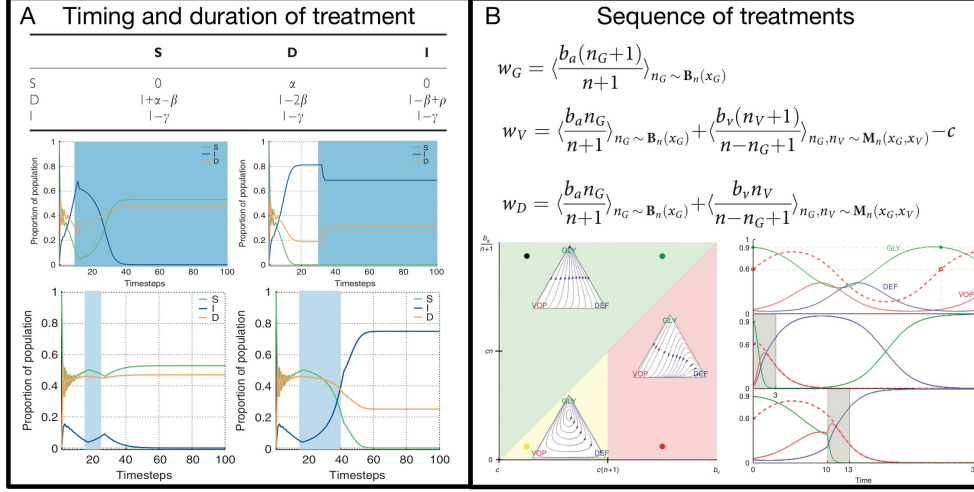


Figure 2.6: Evolutionary games with three strategies have illustrated the effect of the timing, duration and sequence of therapies shaping the composition of heterogeneous tumors. (A) The timing and duration of therapy were found to drastically change the evolutionary dynamics in a three strategy game of prostate cancer cells [26]. (B) In a model considering public goods, we showed that two therapies, given for the same duration, but in a different order, could affect the long term outcomes of a population of tumour cells [120]. Figure adapted from Basanta et al. [26] and Kaznatcheev et al. [120], with permission.

population such that we need only consider frequencies [218]. The resulting model takes the form of a couple system of ODEs. We first began to study the effect of timing and duration of hormone therapy in a three strategy game representing prostate cancer. The cell types we modelled were: those susceptible to hormone therapy, termed dependent (D) as they depended on the availability of androgens in the environment; stromal cells (S) and cells that were resistant to hormone therapy, whose fitness was independent (I) of hormone. In Figure 2.6A, we show the payoff matrix, which represents the mathematical model at the top, and then examples of therapy timing (middle) and duration (bottom) affecting overall population dynamics, and also the resultant evolutionary steady states (ESS) [26].

In Figure 2.6B, a second model is illustrated in which we studied the evolutionary dynamics of three populations of cancer cells, two of which produce factors (goods) beneficial to others in the population. This model is a three-strategy extension of the popular ‘public goods’ game with some modifications to consider proximity of cells to one another [120]. We did not explicitly consider space in this model, but we have made attempts to spatialise these games before, but only in the context of two player games [119]. The resulting evolutionary dynamics in this model provided a theoretical illustration of the concept that the strategies themselves could be targeted to influence overall population dynamics, rather than simply targeting cell types as we do in oncology now.

2.4 Other work modeling stem cell dynamics

While in this thesis I will present work on the cancer cell stem hypothesis (CSCH) for consideration for the qualification of D.Phil., I have also contributed significantly to other works on this topic during this time which I will highlight here.

2.4.1 Ordinary differential equations

Before deciding on a final form of the thesis, I spent some time working on a multi-compartment formalism of the CSCH as an extension of existing work from Werner et al. [243]. While my own work on this ended up being one of many dead ends during my time as a DPhil student, it sparked a collaboration, leading to publication [244]. We used a multi-compartment ODE model, and explored the dynamics of a cancer stem-cell (CSC) driven tumour under therapy. We found characteristic qualitative behaviours of both the total population and stem-cell fraction curves (Figure 2.7 B,C). These characteristic curves, if fit from patient data, can be used to infer the parameters of the underlying ODE model [244]. This information, can in turn, be used to assess the efficacy of a given therapy.

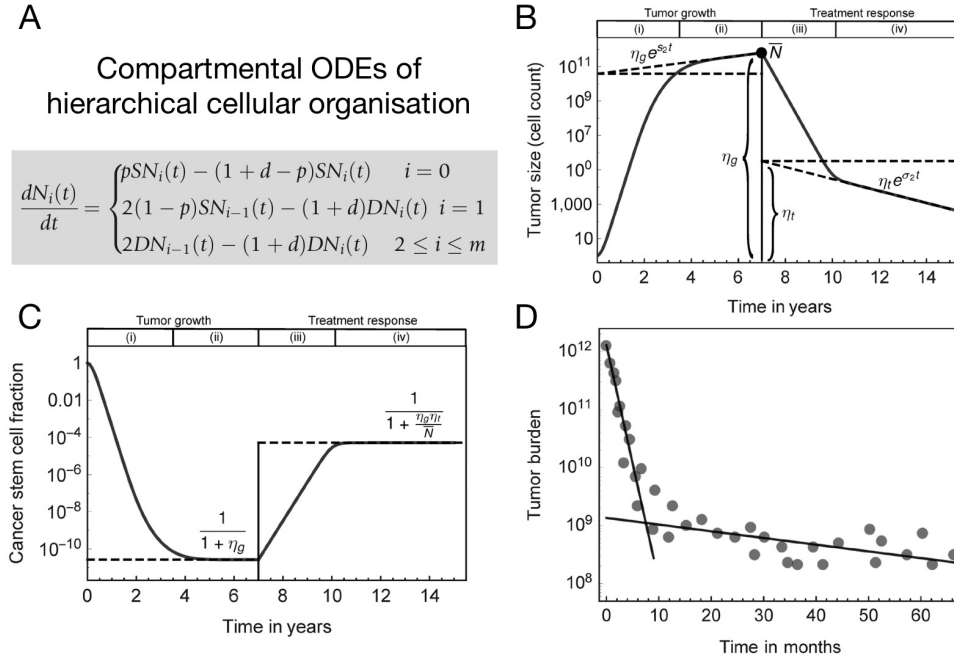


Figure 2.7: Stem cell driven tumours have characteristic growth curves in and out of therapy that can be used to infer parameters of an underlying model. (A) A system of coupled ODEs describing a hierarchically organised tissue was adopted from [243]. Characteristic curves for total population (B) and cancer stem cell fraction (C) can be used to infer model parameters from patient data (D). Figure adapted from Werner et al. [244], with permission.

2.4.2 Cellular automaton models

A second model, which directly preceded the thesis work presented hereafter, was constructed to study how the microenvironment, in the form of local oxygen concentration, affected the ability of CSCs to form tumours. Together with David Basanta, Alexander Anderson and Prakash Chinnaiyan from the Moffitt Cancer Center, I developed and explored a stochastic, hybrid discrete-continuous cellular automaton model of a stem-driven cancer growing in a heterogeneously vascularised domain. The model served as the basis for the model I eventually used to conduct the lion’s share of the work on spatial models presented in this thesis, but required significant modifications as discussed in detail in later chapters. In this work however, we studied the effect of changing vascular density, the symmetric division rate of stem cells and the proposed lifetime for transient amplifying cells [213]. We identified combinations of microenvironmental variables and cell intrinsic phenotypic parameters that coded for loss of homeostasis, and suggested that to fully understand tumourigenesis, cancer stem cells must also be placed in the specific microenvironmental context. Further, our study supports the concept of stemness itself being affected by changes in oxygen concentration.

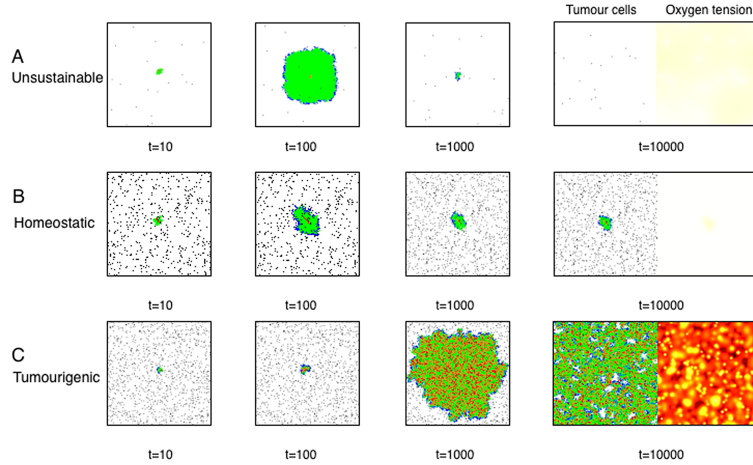


Figure 2.8: Microenvironmental Variables Must Influence Intrinsic Phenotypic Parameters of Cancer Stem Cells to Affect Tumourigenicity. We explored the parameters of a hybrid cellular automaton of cancer stem cells in a heterogeneous environment. We conclude that future theoretical models of stem-cell driven tumours must include specific feedback from the microenvironment onto the individual cellular behaviour, and identified several cell intrinsic parameters which govern loss of homeostasis into a state of uncontrolled growth. Figure reproduced from Scott et al. [213], available through Creative Commons.

2.5 Clinical radiation biology

Radiation response has been modelled mathematically for decades, beginning with the linear-quadratic model of Lea [132]. At the heart of this is the linear-quadratic equation,

which was originally used to describe the results of colony formation assays. This simple equation describes the surviving fraction (SF) of cells after a given dose of radiation in terms of two cell-intrinsic biological parameters, and has been used as the basis of clinical trial dosing schemas, and allows us to understand how radiation doses add together when given in different fractionation schedules. The linear-quadratic model, which we will discuss in more detail in Chapter 6, is given by

$$SF = e^{-nd(\alpha+\beta d)}, \quad (2.1)$$

where n represents the number of doses of d Gy of radiation therapy, and α and β represent the linear and quadratic biological parameters, which we will discuss later.

In work published with collaborators from the Moffitt Cancer Center, led by Javier Torres-Roca, I noticed that a gene signature derived from *in vitro* experiments of 2 Gy fractions of radiation could be incorporated into this model by simple rearrangement. In particular, substituting the numeric score of the radiation sensitivity index (RSI) [69] into equation (2.1) for SF, and assuming that β is a constant, allows us to solve for a ‘genomic α ’. This is the first time that directly calculating a radiobiologic parameter from a patient sample has been possible. Secondly, once a patient’s ‘genomic α ’ has been ascertained, one can calculate the effect of a given dose of radiation which we dubbed the genomically adjusted radiation dose (GARD), as illustrated in Figure 2.9.

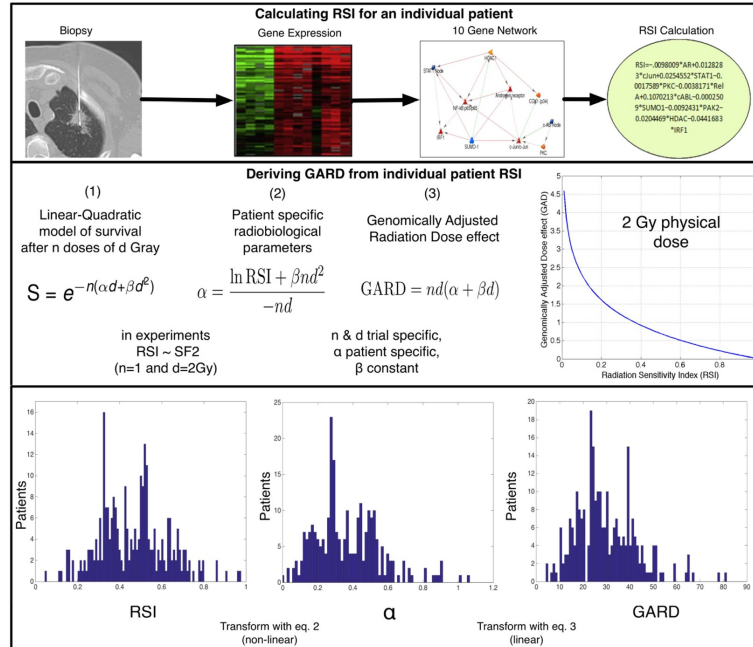


Figure 2.9: Deriving GARD from a clinical sample. After biopsying the patient, a measure of gene expression is made and transformed to an RSI score (top). This score is then mathematically transformed using equation (2.1) (middle), and can then be applied to patient level data (bottom). Figure adapted from Scott et al. [211].

Armed with this information, we then proceeded to analyse an institutional database with $\approx 9,000$ patient tumour samples. We found a high degree of heterogeneity in estimated α values which corresponded to variability in the efficacy of the dose received (Figure 2.10). Thus, while two patients may have been given the same physical dose of radiation, the biological effect of that dose could be very different. We also showed that these differences can explain differences in clinical outcome, and can be used to guide personalised dosing decisions [211]. This work represents the first step to bringing radiation oncology into the era of personalised medicine.

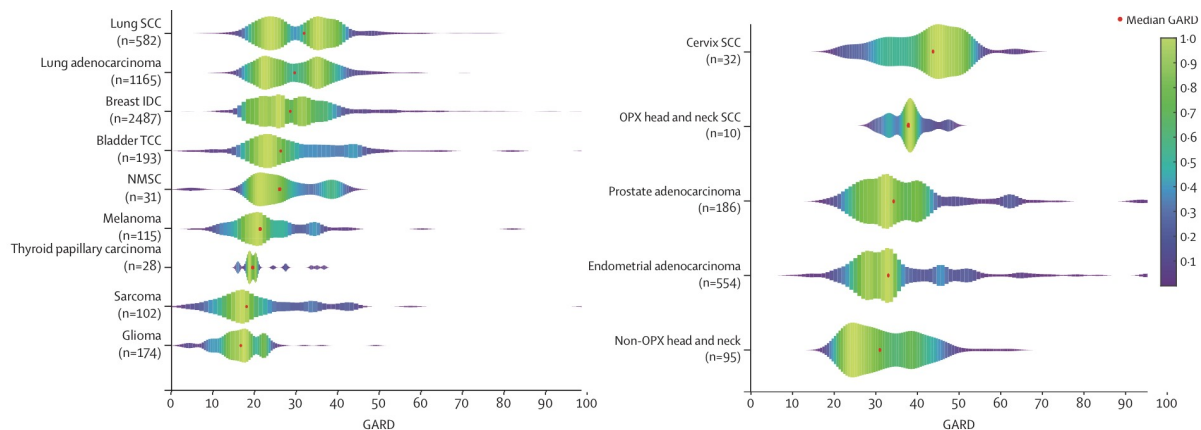


Figure 2.10: GARD score distribution and density within putative 60 Gy (left) and 70 Gy (right) dose groups, by disease site. We demonstrate wide heterogeneity in radiation effect between patients with the same tumour type, which offers a parsimonious explanation for our clinical failures. The red dot represents the median GARD value for each disease site at assigned dose levels. Colours in the plot correlate with sample density. Figure reproduced from Scott et al. [211], with permission.

Chapter 3

Inferring tumour proliferative organisation from measures of diversity and spatial heterogeneity

3.1 Introduction

The cancer stem cell hypothesis (CSCH) states that tumours are composed of a hierarchy of cells, and at the top of this hierarchy, there is a subpopulation of ‘stem cells’, or tumour initiating cells (TICs) able to self-renew and give rise to the entire diversity of cells within the tumour. The CSCH has provided us with a framework by which to understand many different aspects of cancer progression including: functional heterogeneity despite genetically identical states [147, 223, 239, 243], resistance to chemotherapy [42, 51, 244] and radiotherapy [23, 53, 56], recurrence [58] and metastasis [183]. Regardless of this utility, the CSCH has been the subject of continual debate and modification in an attempt to maintain compatibility with experimental observations [91, 181].

While the specifics of the CSCH are still a matter of debate, that the cells with the traits ascribed to TICs are clinically relevant, is not. While this importance is known, our ability to measure, clinically, any of the dynamical aspects within the cancer stem cell model is severely lacking. In fact, even efforts to measure these *in vivo* (in mice) are limited to carefully conducted live imaging in genetically engineered mice [198], or in limited genetic labelling and subsequent lineage tracing [61] – and to date, researchers are limited to defining whether or not a tumour is hierarchically organised, and are not yet able to measure specific parameters governing the hierarchy. While *in vitro* work is better able to define these parameters, very little work has been done to quantify them as tedious single cell lineage tracing [131] is required. These experimental difficulties speak to the need for more theoretical work in this area – specifically to provide metrics for quantifying parameters of the stem cell hierarchy using clinically available data. This is of particular

importance as evidence is mounting as to the relevance of the stem cell hierarchy in determining response to treatments including both radiation [233] and chemotherapy [42, 106]. Further, we now know that certain microenvironmental factors such as hypoxia [45, 54], acidosis [104], growth factors [59], and even stromal cell co-operation/co-option [139, 238] can perturb this system.

3.1.1 Previous work

Several cellular automaton (CA) models of hierarchically organised tumour growth have been used previously to study tumour evolution and the resultant heterogeneity. Poleszczuk et al. [188] modelled a hierarchically organised tumour growing in a homogeneous domain where each dividing stem cell ‘mutates’ with some probability by adopting a small change in a parameter, through random drift, determining motility, symmetric division probability, transit amplifying cell (TAC) replicative potential, or death. These new parameter values are then passed to the cell’s offspring in a heritable manner. The authors identified distinct parameter regimes for which qualitatively different tumour growth patterns – including tumours with long dormancy periods and aggressive growth – were observed, and these populations often existed in the same tumour. In addition, random sampling, through radially oriented virtual ‘core biopsies’, identified less than 10% of the overall heterogeneity of the main tumour bulk. Results from another CA model of tumour growth by Kostadinov et al. [128], this time with neutral mutation (no phenotypic change conferred by mutation), suggest that spatially segregated biopsies taken from a tumour can provide false evidence for higher mutation rates. To address this, the authors suggest considering bulk tumour shape characteristics, not just molecular signatures from biopsy.

Beyond modelling intra-cellular heterogeneity, several groups have modelled how bulk tumour characteristics depend on proliferative organisation. Sottoriva et al. [223] develop a three-dimensional hybrid CA model of cancer growth and progression and compare hierarchically organised (TIC-driven) to ‘flat’, TIC only, tumours. They found that TIC-driven tumours exhibit a different spatial morphology than ‘flat’ tumours, a finding replicated in a contemporaneously published, qualitatively similar, model from Enderling et al. [66]. Both groups observed a naturally invasive edge that emerged in the hierarchically organised tumour, but not the flat one. Further, Sottoriva et al. observe, in a different work using the same model system, significantly different rates of ‘phenotypical evolution’ between the hierarchically and non-hierarchically organised tumours [224]. Given a fixed (phenotypic) mutation rate, the authors find a significantly increased rate of generation of novel phenotypes in the flat tumours, but after normalisation for generation rate, the ‘meaningful’ heterogeneity is larger in the hierarchically organised tumours.

Considering the evolutionary forces that shape TIC-driven tumours, Sproufsske et al. [228] developed a suite of models to study how selection drives TIC proportion. In particular, they used a stage-structured ODE model to study the role of TACs in a tumour, and the evolution of their number during tumour growth. Their findings, that the proportion of TICs within a tumour is far below that suggested in a null, ‘cell-centric’ (i.e. no interaction), model of tumour evolution, suggests that there is selection acting within the tumour beyond the cell level. As discussed above, the microenvironment has previously been implicated in selection, or promotion of ‘stemness’; however, the concept that there is ‘multi-level selection’, or interplay between the TIC and non-TIC cells, is novel.

3.1.2 Aim of work

Multiple mathematical models, taking different forms and considering heterogeneity in different elements, have predicted that hierarchically organised tumours differ significantly in their evolution as compared to ones without this hierarchy. These models have included spatial proliferation constraints, microenvironmental heterogeneity and selective pressures, and the noted differences include shape, heterogeneity, rate of evolution and growth dynamics. In none of these models, except Sproufsske et al. [228], in which the main question centred on TAC numbers, were these differences studied across probabilities of TIC symmetric division, which is a key parameter governing the hierarchy, and one that is exceedingly difficult to measure or perturb *in vitro* or *in vivo*. In this chapter, we study the effect of changing symmetric division probability in both a spatial and non-spatial model of TIC-driven neutral tumour evolution. As the evolution of TIC proportion has been studied, we focus instead on observed spatial and clonal group (i.e. genetically identical groups) size patterns to work to discern useful measures to aid in the clinical determination of TIC symmetric division probability. Having a tool to estimate this probability in the clinic could be very useful in aiding our understanding of the effect of therapies on this crucial parameter as well as by which to stratify tumours (or patients) for consideration of different types of therapy. In this way, we hope to provide translatable measures to aid in the understanding tumour biology: to use mathematical modelling to ‘see the invisible’.

As we aim to translate these findings to clinical samples, considering biologically observable analogues within our model system is of great import. While not in standard clinical practice, the shape of clonal groups bearing specific mutations can be observed in histologic samples, albeit only for a very small number of mutations at a given time. Further, the new gene editing tool CRISPR-Cas9 [43] and DNA barcoding [109] have, together, allowed experimentalists to design methods with which to observe distributions

of colony size with high resolution. Examples of these two data types can be seen in Figure 3.1.

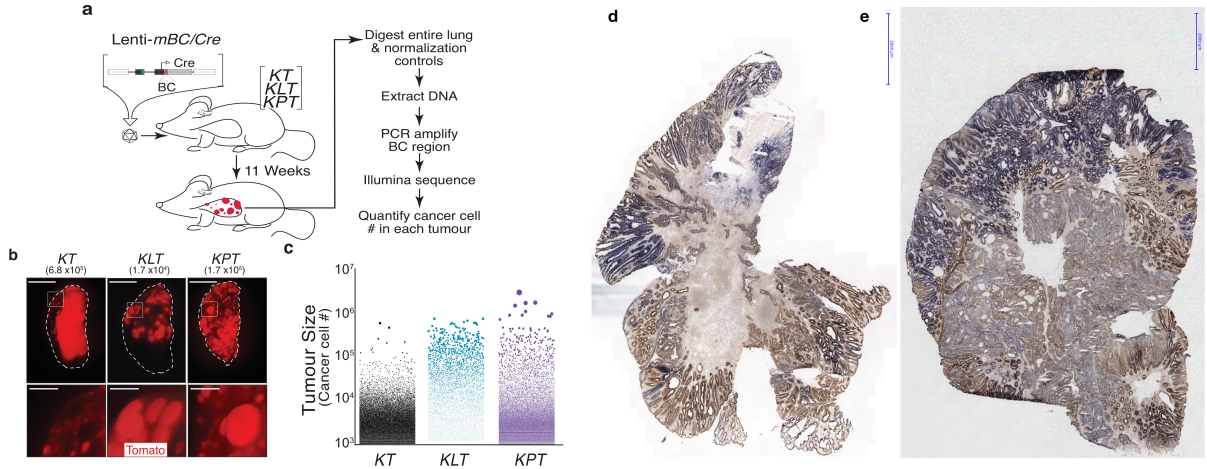


Figure 3.1: Colony size distribution is observable in mouse models (a-c) and colony shape is observable in clinical samples for some neutral mutations (d,e). Genetically engineered mouse models (a) provide a new experimental set up whereby investigators can instantiate many tumours with different drivers using gene editing to test relative fitness using Tomato red fluorescent protein (b) as measured by DNA barcodes and association with clone size distribution (c). *In vivo* examples adapted from Rogers et al. [201]. Colorectal adenoma samples (d,e) are stained for the neutral mutation in cytochrome oxidase, and can reveal the shape of a clone. Image generously provided by Trevor Graham, unpublished.

3.1.3 Structure of chapter

The remainder of this chapter is structured as follows. In Section 3.2 we build a stochastic, non-spatial, cell-based model which couples a hierarchical model of cell proliferation together with neutral evolution. We explore the effects of spatial constraints on cell proliferation and the TIC hierarchy on clonal diversity in Section 3.3.1, by comparing across symmetric division probabilities and between the spatial and non-spatial models. In Section 3.3.2 we analyse the effects of space and TIC symmetric division probability on clone size distribution and then, in the case of the spatial model, clonal shape in Section 3.3.3. With an eye to a translatable measure, in Section 3.3.4, we introduce a method of ‘biopsying’ our spatial model and explore the effect of varying biopsy strategy on our ability to discern tumour diversity. We conclude, in Section 3.4 with a discussion of our results and future directions.

3.2 Model development

3.2.1 Non-spatial TIC-driven model of neutral evolution

Before we consider a full, spatial model, it is of value to consider a simpler model. Therefore, we introduce a non-spatial stochastic model of neutral evolution to see the effects of the proliferative hierarchy before increasing the complexity to include spatial constraints. Neutral evolution has been modelled before in non-spatial models, most notably by Kimura [125] in molecular evolution, but also by others, for example Bjerknes [27] studying neutral evolution in human colorectal cancer and Williams et al. [246] in human cancer as a whole. The effect of the proliferative hierarchy on neutral evolutionary dynamics, however, is less well studied.

Proliferative hierarchy

To begin, we model two types of cells within a proliferative hierarchy: TICs and TACs. We allow TICs to divide either symmetrically with probability α , creating two TIC daughters, or asymmetrically, creating one TIC and one TAC, with probability $1 - \alpha$. In this work, we do not consider symmetric differentiation of a TIC to form two TACs. It has been shown previously [200] that the overall dynamics of TIC-driven tumours is equivalent with or without TIC symmetric differentiation, and in this study of neutral evolution it would cause undue complexity due to lineage extinction, hence we leave this consideration for future work. We model TAC division as always symmetric, creating two identical TAC daughters, but only allow this to progress for β rounds of division. This parameter, β , represents the replicative potential of TACs, and has been posited to biologically represent telomere length [187]. Previous theoretical work by Morton et al. [167] has shown that tumour growth kinetics in spatially constrained geometries are strongly affected by TAC proliferative potential, β . In particular, they showed substantial changes in growth kinetics when $\beta > 5$, whereby simulated tumours experienced growth delays so significant as to not be considered realistic. As this has been previously studied, and our goal in this chapter is to determine our ability to infer TIC symmetric division probability from surrogate measures of tumour growth and heterogeneity, we will hold TAC replicative potential constant, and we choose $\beta = 4$. Of note, this assumption has been made previously by several groups to ease analysis of the hierarchical model [223, 228]. This mode of growth and differentiation is schematised in Figure 3.2.

Neutral evolution

To understand the effects of neutral evolution, we extend our model of tumour growth under a proliferative hierarchy to include mutation, which we allow to occur at TIC

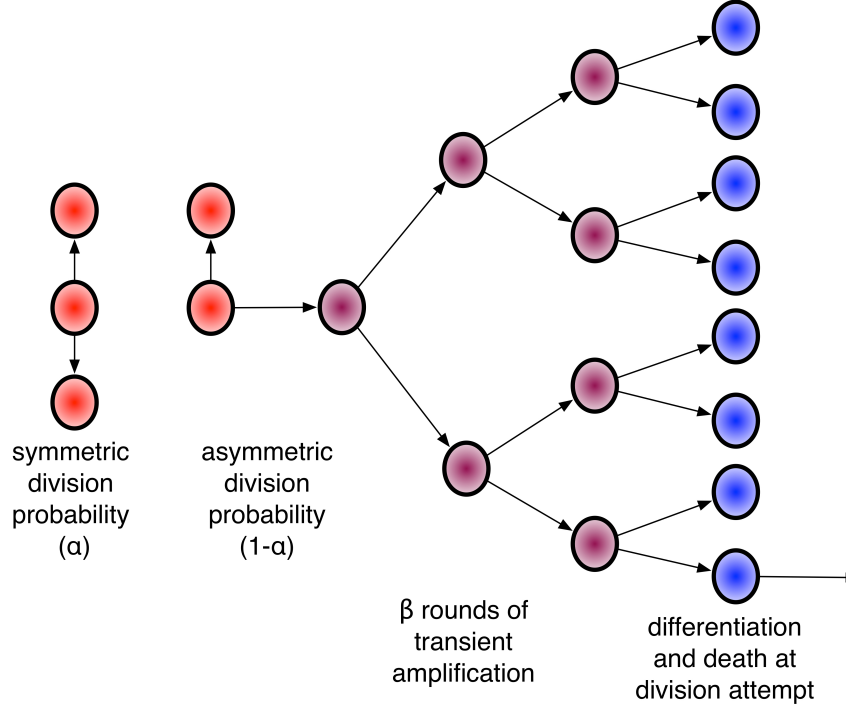


Figure 3.2: A schematic representing the hierarchical model of tumour cell proliferation and differentiation. In this formulation, each tumour initiating cell (TIC; red) can undergo two types of division, either symmetric with probability α or asymmetric, with probability $1 - \alpha$. Each subsequently generated transit amplifying cell (TAC; purple) can then undergo β rounds of amplification before differentiating into a terminally differentiated cell (TD; blue) which will die at its next division attempt.

division events, a similar assumption as has been made by others, including Sottoriva and colleagues. [224]. At each cell division event, there is a possibility that one or more mutations occur. To determine this number, we independently draw a random number of new mutations to be bestowed upon each daughter from a Poisson distribution with expected value λ . In this case the probability of having k mutations at a given event is given by

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}. \quad (3.1)$$

Given the extremely large number of possible mutations in a real cancer, we prescribe that every mutation arising in this CA model is unique, an assumption usually ascribed to Kimura and termed the ‘infinite sites assumption’ [124]. Invoking this assumption removes the possibility of convergent genetic evolution, which would complicate analysis. All mutations acquired are neutral: they confer no advantage, disadvantage or any other phenotypic change and therefore serve only as a method by which to track clonal lineages. A schematic of this evolutionary process, and labelling scheme, is shown in Figure 3.3.

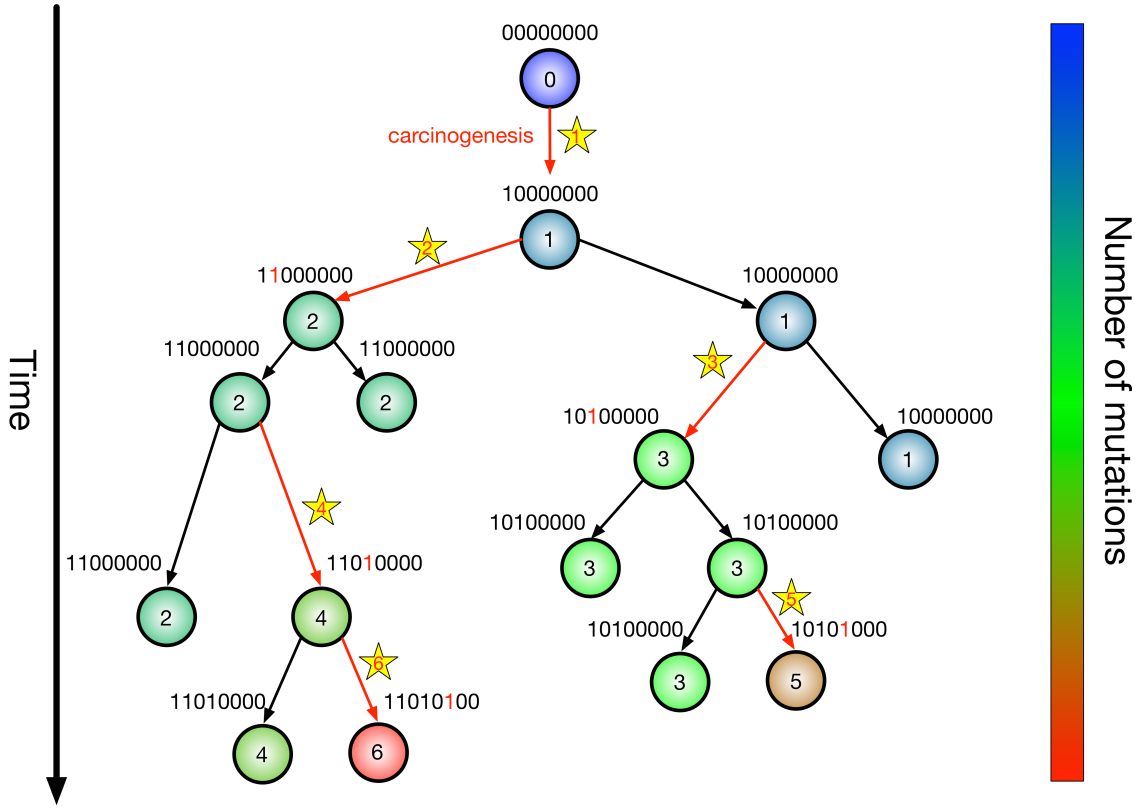


Figure 3.3: Schematic for mutation in our model of neutral evolution. At each division event (branching) after the first (carcinogenesis, labelled with a starred 1), a random number of mutations drawn from a Poisson distribution with expected value λ is conferred on a daughter (subsequent starred events). The n th mutational event is given a unique flag, which is inherited by its offspring unless they too mutate. Each unique mutation can then be considered as a novel mutant allele (red) appearing in the population. In this chapter, we consider only the integer flags for each cell. Each simulation begins immediately after the first mutation occurs, hence the first cell is labelled ‘1’.

For computational efficiency, we record a unique flag only for the most recent mutation accumulated within a cell, which is passed down to its progeny. We also record each individual mutation event in the form of an ordered pair so that the complete ‘genomes’ (bit strings) can be reconstructed for future use. As they are the only cells capable of forming tumours on their own, and infinite replication, we, like Sottoriva [223] and others [188, 228], consider new mutations to accrue only in TICs.

Non-spatial model implementation

To implement a first version of our hierarchical model with neutral evolution, we consider the case in which there are no spatial constraints. We initialize our simulation with a single TIC. We then implement discrete updates, which we will term ‘time steps’. As we are not studying temporal dynamics here, we do not prescribe any proper dimension to this time. At each discrete time step we choose a cell uniformly at random from the

population to either divide and possibly mutate, or die. Our simulation is considered complete when the total population of cells reaches a prescribed number.

If the cell is a TIC, we first draw a random number, r , from $U[0, 1]$ and compare it to the probability of symmetric division, α . Then, if $r < \alpha$, the cell divides symmetrically, we draw the number of new mutations accumulated in each daughter independently from $\text{Poisson}(\lambda)$. A new TIC daughter is then generated and given a mutational ‘identity’ which is the mathematical sum of the parent cell’s identity and the number of new mutations. The parental cell is also updated by changing its identity based on the number of new mutations it accrues, if any. If the cell is not determined to divide symmetrically, it divides asymmetrically and a TAC daughter is created with age 0 and the parental TIC is updated for new mutations as above, see Figure 3.3.

If the chosen cell is instead a TAC, we first check that its age is less than the allowed TAC age, β . If the age is equal to the allowed TAC age, then the cell dies. If the age is less than the allowed TAC age, then the cell ages and divides, and a new daughter with the same identification is created, whereupon the loop continues (see Figure 3.5 for schematic).

3.2.2 Spatial model

As we are interested in the effect of spatial proliferative constraints and spatial patterns that emerge from the proliferative hierarchy, as well as from neutral evolution, we will extend the previously described non-spatial CA model by embedding it on a two-dimensional lattice.

Embedding the model on a lattice

Beyond rules governing division mode as described above, in our spatial model we also consider each cell division to be modulated by contact inhibition [12]. Therefore, each cell is only allowed to divide if there is available space within a prescribed neighbourhood, for which we use the Moore neighbourhood (see Figure 3.4). If a cell is contact inhibited, we consider it to be in a quiescent state, a state which is reversed when space becomes available. At each time step, each cell has an opportunity to divide. Cells are chosen from the list of all cells uniformly at random to determine if and when they divide, to avoid order bias. No matter what type of cell is chosen, the first check is for available space in the neighbourhood.

Cell-type specific rules

TIC: If space is available, and the cell is a TIC, then the type of division is determined by choosing a random number, r , from $U[0, 1]$. If $r < \alpha$, then the TIC will divide

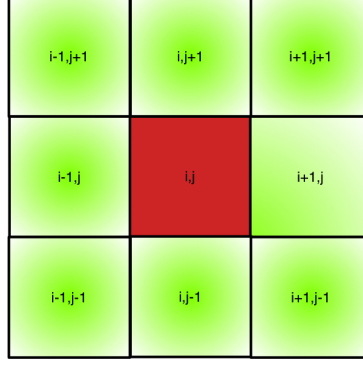


Figure 3.4: Definition of lattice position neighbourhood for cell division. Cells at position (i, j) ascertain whether they have space for division by considering their Moore neighbourhood, which consists of the 8 spaces in their immediate vicinity: $\{(i+1, j+1), (i+1, j), (i+1, j-1), (i, j+1), (i, j), (i, j-1), (i-1, j+1), (i-1, j), (i-1, j-1)\}$.

symmetrically, creating another TIC which will be placed uniformly at random in one of the empty neighbouring spots in the previously defined Moore neighbourhood. Further, the daughter and/or the chosen TIC, will acquire a number of mutations, as described in Section 3.2.1. If the random number is greater than or equal to α , then the TIC divides asymmetrically. When dividing asymmetrically, a new TAC is created and placed in a neighbouring, empty, lattice position at random. The new TAC daughter is created with the same mutation ID as the parent, and age = 0. Further, the TIC which divided asymmetrically will be updated with some number of mutations, as above.

TAC: If the chosen cell is instead a TAC, then the check after available space is a check of the cell's (proliferative) age. If the TAC age is equal to the proliferative potential, then the TAC dies. If the TAC age is less than the proliferative potential, then the TAC creates a new TAC daughter and places it in an empty space in the Moore neighbourhood at random. Both the chosen TAC and the new daughter share the same mutation ID and their age is updated to be one more than the age of the originally chosen TAC.

Full implementation: The full CA flow-chart, represented in Figure 3.5, schematises the entire process of cell fate decisions that each cell undergoes at each time step in the spatial model. In the top panel, the rule set followed by the TICs is represented to include differentiation and mutation. In the bottom panel, the TAC rule set is defined to include death by terminal differentiation and TAC aging. This schema also represents the non-spatial model implementation with the exception of the spatial checks, which are denoted in red.

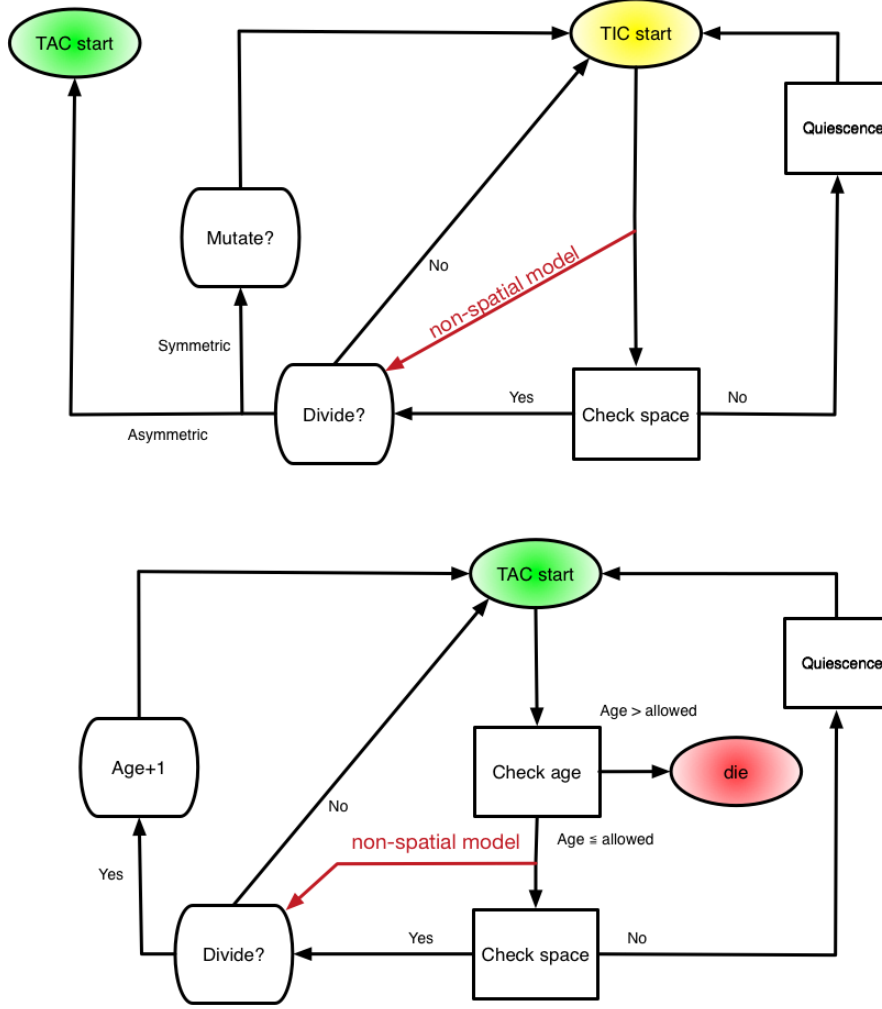


Figure 3.5: Schematic of discrete time updating algorithm for the hierarchical CA. At each CA update, each cell in the domain undergoes a series of fate decisions based on intrinsic cell parameters and neighbourhood occupation. Top: we represent the fate decisions made by a TIC (yellow) as it divides symmetrically or asymmetrically or stays quiescent. Bottom: we represent the fate decisions of a TAC as it divides and ages. For the non-spatial model, the ‘check space’ step is skipped, as indicated by the red arrows.

3.3 Results

3.3.1 Ascertaining the effect of space on tumour diversity

It has been shown previously *in silico* that the choice of rules governing proliferation, in particular whether or not a tumour consists of cells which obey a TIC-cell hierarchy, strongly affects tumour growth rate [65, 167]. It has also been shown that the spatial pattern of mutations differs between these two tumour types [66, 223, 224]. In particular, it has been shown that TIC-driven tumours have a ‘patch-like’ clonal architecture, with small ‘islands’ of clones displaying the same (epi)genetic states, while non-TIC-driven tumours display a radial patterning of clones, as can be seen in Figure 3.6. As previously

observed by Enderling and colleagues [65, 167], we note here a very large difference in population growth rate, which we further quantify in Appendix B, Figure B.1. While the observations of growth kinetics are not novel, we will further quantitatively explore the utility of the tumour clonal shape/architecture observations later in this chapter.

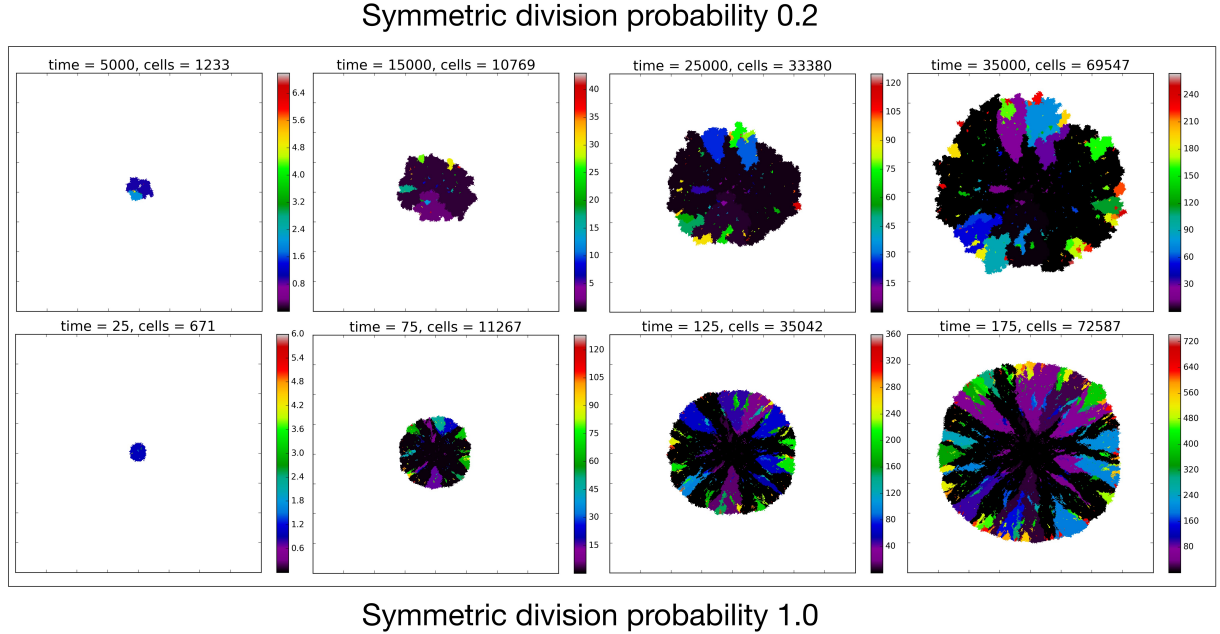


Figure 3.6: Temporal evolution of the spatial model reveals observable morphologic differences between TIC-driven and non-TIC-driven tumours, as observed by others. We plot representative results of a single example simulation for two tumours, each simulated on a square lattice of size 400×400 . Top: a tumour simulated with symmetric division probability 0.2 and $\beta = 4$. We notice, as have others, a ‘patchy’ clonal architecture, and non-uniform edge. Bottom: a tumor simulated with symmetric division probability of 1.0, which has no proliferative hierarchy. We note smooth edges, radial patterns of clonal architecture and relatively faster population growth, reaching $\approx 70,000$ cells in less than 200 time steps. To reach a similar size, the tumour with symmetric division probability of 0.2 took 35,000 time steps.

In order to make these qualitative observations quantitative and enable us to objectively determine differences in the symmetric division probability, we introduce the Shannon index, a commonly utilized metric of overall diversity within a population. Originating from information theory, it was originally created to understand how different one string of text was from another [214]. It has been widely utilized in many other fields of biology, as it is a convenient metric to describe diversity [148]. For a population of size N comprising n_i individuals of type i , for $i = 1 \dots R$, the Shannon diversity index is defined as:

$$\text{Shannon index} = - \sum_{i=1}^R \frac{n_i}{N} \log \frac{n_i}{N}. \quad (3.2)$$

While a very non-specific diversity metric, the Shannon index has shown predictive capacity in cancer progression [52, 149, 151].

To begin to explore the effect of spatially constrained proliferation on clonal evolution in our model, we will measure diversity in comparable scenarios both with and without spatial constraints. To this end, we will simulate a range of symmetric division probabilities holding TAC lifespan and mutation probability constant, and use the Shannon diversity index to compare across the different symmetric division probabilities and between spatial and non-spatial models.

We begin by briefly observing the results of the spatial model. Figure 3.7 shows example resultant tumours after the domain is filled for symmetric division probabilities 0.2, 0.6 and 1.0, from left to right, with the associated mutation map as a colour bar on the right and Shannon indices for 10 averaged simulations reported above. We find that, trivially, the higher the symmetric division probability, the more mutations are accrued in a tumour of equivalent size. Comparing the average Shannon indices with the same parameter sets, we find little difference between the two TIC-driven families of simulations (SI means ≈ 3.47 and 3.42), but each TIC-driven family differs in a statistically significant way from the non-hierarchical family (SI mean ≈ 3.85 , $p \approx 3.4 \times 10^{-7}$ and $p \approx 5.2 \times 10^{-9}$ using a Wilcoxon rank sum test comparing 0.2 and 0.6 to 1.0 respectively). This suggests that, while we may not be able to determine the symmetric division probability of an *in silico* tumour from the Shannon index, we can at least discriminate between TIC-driven and non TIC-driven. To more accurately determine the changes as we vary TIC symmetric division probability within each model, and the difference between a spatially constrained tumour and one that is not, we will next perform a side by side comparison.

Beyond changes in diversity measures between TIC and all-TIC simulations, we also notice, as reported previously [66, 223], that the TIC-driven tumours have a patchy pattern as compared to the ‘flat’, all-TIC tumour ($\alpha = 1$), which exhibits a radial pattern. This change can not, of course, be studied in the non-spatial model, so we will leave it until later in this chapter to explore on its own.

Non-spatial model diversity

To ascertain the effect of spatial constraints on proliferation, we begin with the simpler model, and measure diversity, using the Shannon index, in our non-spatial model. We perform 50 stochastic simulations for each of a range of symmetric division probabilities (0.2, 0.4, 0.6, 0.8, 1.0) in the non-spatial model described in Section 3.2.1. For each simulation, we measure the Shannon index for the resulting tumour when it has reached a prescribed size (in this case 250,000 cells), and plot these results in Figure 3.8 (left).

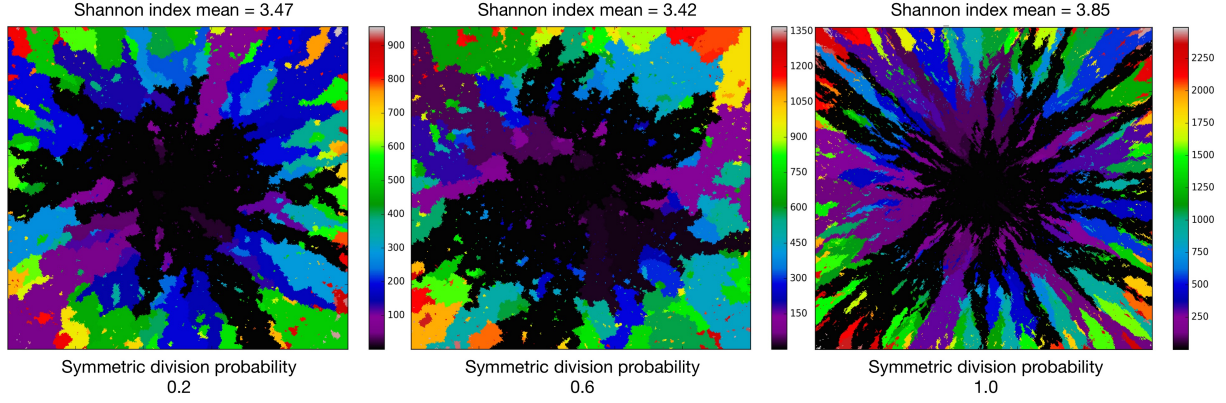


Figure 3.7: Example simulations of the spatial CA for different probabilities of symmetric division. For three different symmetric division probabilities (left: 0.2, middle: 0.6, right: 1.0), we plot an example output of the simulation when the domain is filled (due to different growth kinetics, these plots represent different times) and report the average Shannon index over ten stochastic realisations. There is no statistically significant difference between the two hierarchical families of simulations ($\alpha = 0.2$ and $\alpha = 0.6$, $p > 0.05$ by Wilcoxon rank sum), but the difference between each of the TIC-driven simulations and the all-TIC simulations is ($p \approx 3.4 \times 10^{-7}$ and $p \approx 5.2 \times 10^{-9}$ using a Wilcoxon rank sum comparing 0.2 and 0.6 to 1.0, respectively.)

We find that the Shannon index monotonically decreases with increasing TIC symmetric division probability.

Spatial model diversity

To ascertain if the initial differences we noted in the spatial model hold true in a larger family of simulations, and how the addition of spatial constraints changes our results from the previous section, we perform a series of stochastic simulations of our spatial model with the same parameters as in the non-spatial model. In this series, we choose symmetric division probabilities of 0.2, 0.4, 0.6, 0.8 and 1.0, the latter case being comprised entirely of TICs, and run simulations until the domain is filled, in this case with 250,000 cells.

Model diversity comparison

Comparing the Shannon index across identical simulations between the spatial and non-spatial model reveals significant differences. In Figure 3.8, we see that, for the non-spatial model, the diversity decreases as symmetric division increases, which is not the case for the spatial model. We also note that the mean diversity indices are higher in the spatial model for all symmetric division probabilities as are the variances.

In the non-spatial model we observe a monotonic decrease in the mean diversity and variance of diversity with increasing symmetric division probability. All distributions are significantly different (all p -values $< 1 \times 10^{-6}$) as compared by Wilcoxon rank sum except 0.8 and 1.0 ($p = 0.11$). Therefore, from this diversity measure alone, we can estimate (at

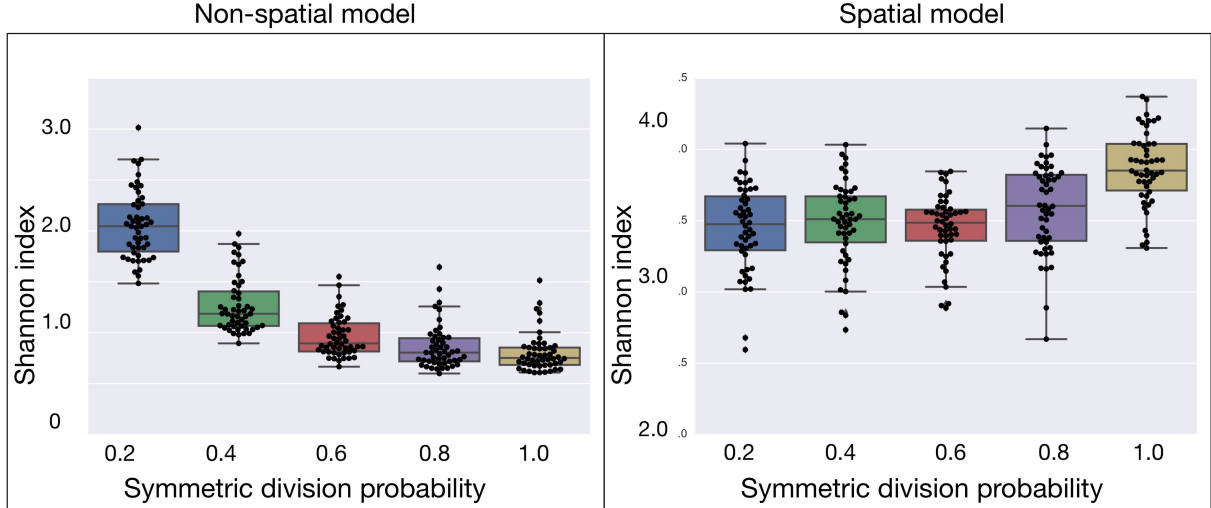


Figure 3.8: Spatial constraints on proliferation affect the Shannon index across a range of symmetric division probabilities. We plot the Shannon index in a box-whisker/swarm plot for 50 simulations for each of the non-spatial (left) and spatial (right) models across a range of symmetric division probabilities, α . The box-whisker plots represent the mean of the data with a horizontal line within a coloured box, the edges of which represent the interquartile range, and the whiskers represent the non-outlier extrema of the data. The swarm plot shows all the data points explicitly. In both instances, all tumours are simulated with TAC lifetime, $\beta = 4$, mutation expectation, $\lambda = 0.01$, and are ended when the tumours reach a size of 250,000 cells.

least relative) symmetric division probability in the non-spatial model, assuming mutation rate and tumour size are known.

In the spatial model, we find that there is no statistical difference between the distributions of Shannon index of symmetric division 0.2, 0.4 and 0.6, but between two of them and 0.8 there is (Wilcoxon rank sum test $p \approx 0.02$, 0.17 between 0.2 and 0.6, respectively). Further, every distribution is significantly different from the all-TIC plot (Wilcoxon rank sum test for each vs. all-TIC respectively: $p \approx 5.35 \times 10^{-10}$, $p \approx 7.30 \times 10^{-9}$, $p \approx 4.81 \times 10^{-11}$ and $p \approx 9.61 \times 10^{-6}$). Means and standard deviation for each symmetric division probability are: 3.44 ± 0.30 ; 3.48 ± 0.28 ; 3.45 ± 0.23 ; 3.58 ± 0.31 ; and 3.86 ± 0.26 . From this metric, at least, we seem to be able to reliably determine if a population is TIC-driven or not, but we are not able to reliably say anything about symmetric division probability.

Effect size: While the differences mentioned above suggest statistical differences between the individual cases, these numbers can be misleading. In particular, as the statistical tests were performed post-hoc, and the sample size was specified independently, it is important to explore more general measures of the differences. For a thorough discussion of the dangers of over-reliance on p-values in particular in simulation-based studies, we refer the reader to the work of White et al. [245]. To bolster our confidence in our

measures, we present a summary plot of the effect size of the change in symmetric division probability between the two models in Figure 3.9. For the spatial model we find a regression slope of 0.423, $R^2 = 0.137$, and $p = 6.57 \times 10^{-8}$. For the non-spatial model we find a regression slope of -1.461 , $R^2 = 0.621$, and $p = 3.24 \times 10^{-54}$.

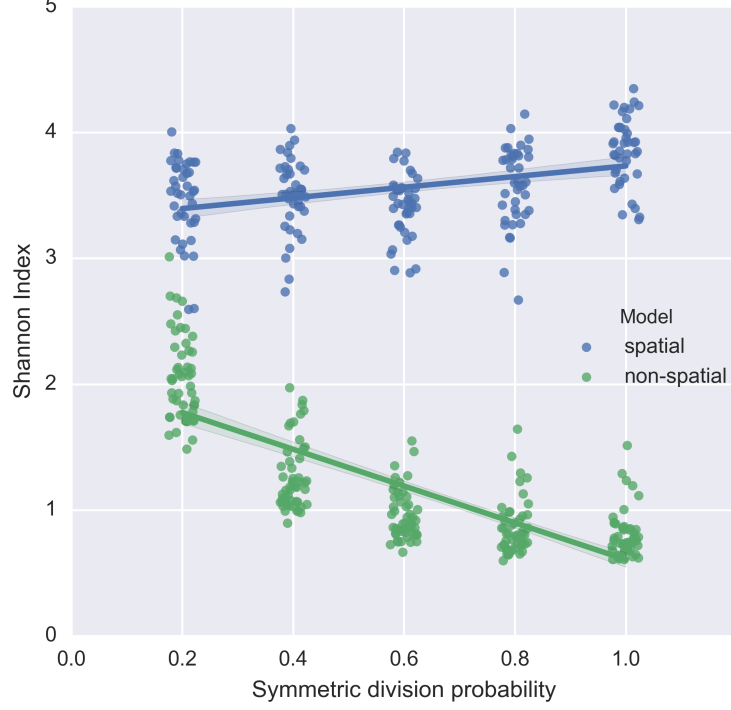


Figure 3.9: Effect size of changes in symmetric division probability on diversity between the spatial and non-spatial model. We plot the effect size for the data shown in Figure 3.8.

3.3.2 The effect of space and proliferative heterogeneity on colony size distribution

In the previous section we found that measurements of the Shannon index were sufficient to allow us to differentiate between tumours of different TIC symmetric division probability in our non-spatial model. When we tested this same measure on our spatially constrained CA, however, we were only able to determine if a tumour was hierarchically organised or not. To attempt to remedy this, we will perform measures of clonal group size on the same families of simulations, with the aim of discriminating between symmetric division probability in the spatial model by comparing the moments of these distributions.

Non-spatial model colony size

As above, we begin with the non-spatial model, and we plot, in the top panel of Figure 3.10, the first three moments of the clonal group size distribution for the simulations of our non-spatial model. We find that the moments changes monotonically with symmetric division probability and are significantly different for all families of simulations with the notable exception of the standard deviation of clone size in the lowest symmetric division probability.

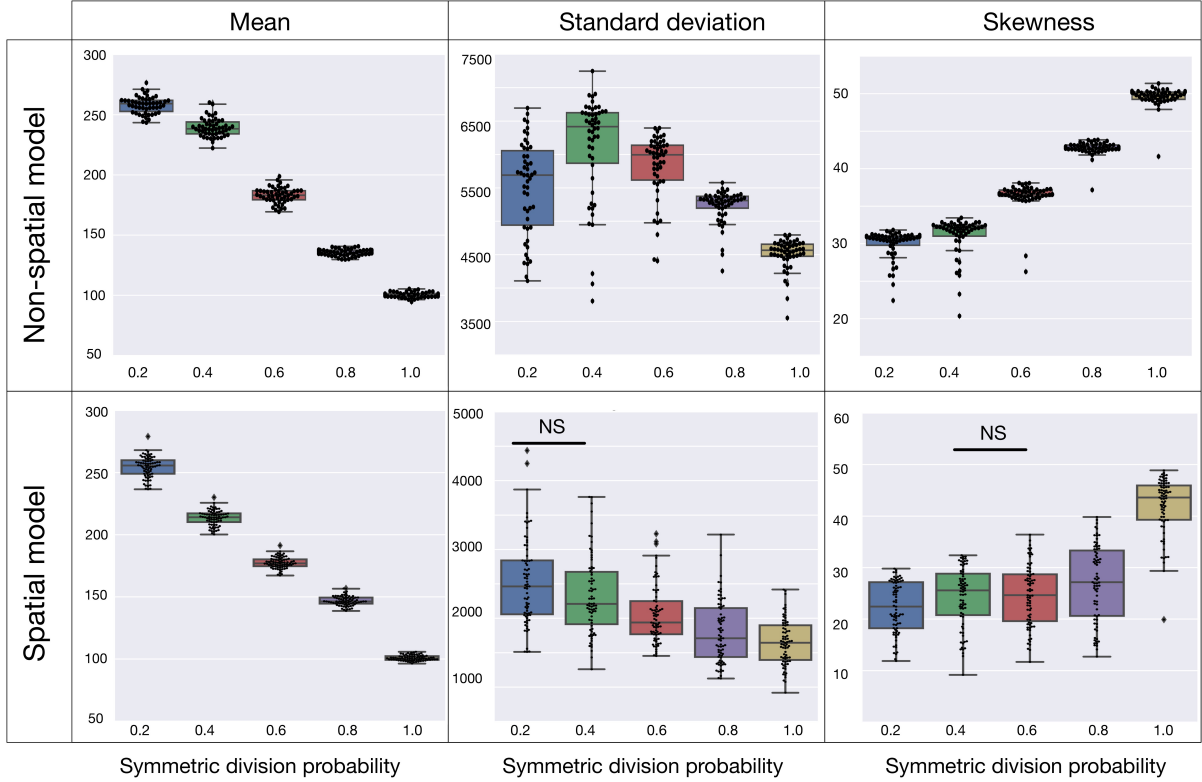


Figure 3.10: Distributions of moments of colony size distribution for each family of simulations for both spatial and non-spatial models. For each family of simulations of differing symmetric division probability, we plot each moment of the colony size distributions for the non-spatial (upper) and spatial (lower) models. We find that as the symmetric division probability increases, the mean and standard deviation of colony sizes decreases significantly. We also observe a larger skewnesses with increasing symmetric division with a large discrepancy between the non-hierarchical vs. all hierarchical families of simulations. All distributions are significantly different ($p < 0.05$ by Wilcoxon rank sum except the standard deviation of symmetric divisions 0.2 and 0.4 and the skewness of 0.4 and 0.6).

Spatial model

To investigate these differences in the spatial model, we plot, in Figure 3.10 (lower panel), the distribution of each moment of colony size for the individual simulations from each family. From these plots, we see the same qualitative relationships we observed in the

non-spatial model results but with less discernable differences, with many distributions in standard deviation and skewness having overlapping error. This suggests that in the spatial model, discerning the symmetric division probability from these distributions would not be possible. The notable exception here is that the skewness of the non-hierarchical ($\alpha = 1$) family of simulations is significantly different from all others, which could be used to tell, at least, whether or not a simulation is hierarchically organised, which in itself could be useful.

As above, comparing the individual families of simulations to one another is only part of the story. To this end, we compute the effect size of the change of symmetric division probability for each of the moments of colony size for both models. For the mean colony size in the spatial model we find a regression slope of -187.04 , $R^2 = 0.986$, and $p \approx 0$. For the standard deviation of colony size in the spatial model we find a regression slope of -1149.70 , $R^2 = 0.309$, and $p = 7.36 \times 10^{-32}$. For the skewness of colony size in the spatial model we find a regression slope of 21.10 , $R^2 = 0.411$, and $p = 7.99 \times 10^{-45}$. For the mean colony size in the non-spatial model we find a regression slope of -206.72 , $R^2 = 0.974$, $p \approx 0$. For the standard deviation of colony size in the non-spatial model we find a regression slope of -1435.51 , $R^2 = 0.262$, and $p = 4.32 \times 10^{-18}$. For the skewness of colony size in the non-spatial model we find a regression slope of 25.21 , $R^2 = 0.893$, $p \approx 0$. These results are plotted in Figure 3.11.

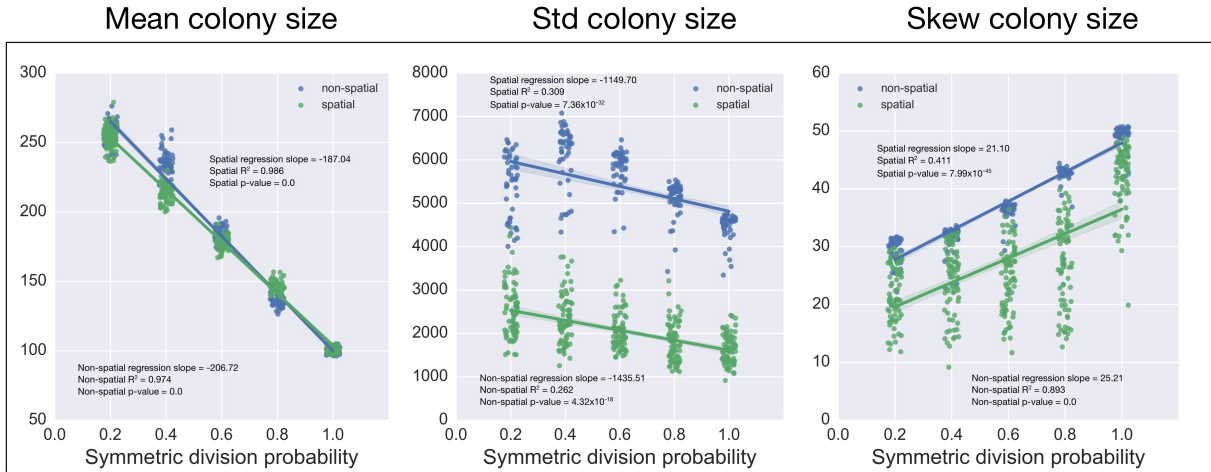


Figure 3.11: Effect size of changes in symmetric division probability on the moments of colony size distribution between the spatial and non-spatial model. We plot the effect size for the data shown in Figure 3.10.

We have described several measurable differences in the moments of the colony size distribution from the families of simulations with varying symmetric division probability. In particular, the mean clonal group size monotonically decreases with increasing symmetric division probability, and the skewnesses of the TIC-driven simulations (regardless

of symmetric division probability) are clustered and lower than the skewness of the non-hierarchical family. While these differences are measurable, their reliability and utility is somewhat questionable. In particular, as many of these differences are relative, it would be difficult to determine into which class a given tumour would fall without another for comparison. Further, specific information about colony distribution *in vivo* is difficult to come by, though recent methodologies employing DNA barcoding [137, 201], genetic labelling studies in mice [61], and multi-region sequencing [108, 221] are becoming more accessible.

3.3.3 Inferring symmetric division rate from spatial measures of clonal architecture

While we can make reliable statements about symmetric division probability from diversity in the non-spatial model, this eluded us in the spatial model. Furthermore, while the moments of colony size do seem to present a promising avenue as a biomarker of symmetric division probability, the lack of access to these data in clinical samples makes it less likely to be translatable in the near term. We have not yet, however, studied any aspects of the spatial model related to colony shape. With an eye to a translatable measure, then, we consider a feature that could be gleaned from a single pathologic specimen. As has previously been observed [66, 223], there is a patchy clonal architecture in the TIC-driven tumours as compared to the more radial architecture of the all-TIC, non-hierarchical ones. With this in mind, we ask: can we determine quantitative differences between the spatial patterns themselves that would allow us to differentiate between the underlying symmetric division probabilities? The ability to do this would lend significant evidence toward the debate as it would allow for a comprehensive understanding of the distribution of ‘stemness’ across many individual tumours, allowing us to determine whether or not they are hierarchically organised. As of now, without single cell tracking (a technique that is very challenging to perform *in vivo* [61, 205], and impossible in patients), this can not be comprehensively determined.

Circularity

As the most striking visual difference between the TIC-driven and non-hierarchical simulations is the radial pattern that the non-hierarchical simulation output has, it would make sense that a metric loosely able to measure something like circularity could help. Most circularity metrics depend on ratios of perimeter (or perimeter²) and area. A more advanced metric [256], however, has considered the higher moments of distance to centroid of the points within a group.

Motivated by these observed differences, we calculate the moments of the distribution of distances to centroid for each member of each colony. We first compute the centroid of each colony, as the ordered pair defined by the mean of all x coordinates and the mean of all y coordinates, of each member of the colony. Once the centroid is determined, the distance from each member to the centroid is calculated. The mean, standard deviation and skewness of these distributions were then computed.

When first exploring this simple metric, any signal which would help discriminate between colonies in terms of circularity could easily be drowned out by size effects. In particular, in the non-hierarchical simulation, we find that the thinnest and longest colonies (those with the largest mean distances from centroid) are often the largest (see Figure A.2), which over-represents those larger moments and reduces our ability to judge shape.

To circumvent this, we take one step further and normalise our distance to centroid moments by the square root of the area of colony (in a square lattice the area is simply a count of the members of the colony). These normalised moments are plotted as box-whisker plots in Figure 3.12. Of note, we find that the ‘flat’ simulation ($\alpha = 1$) can clearly be differentiated from the hierarchical simulations ($\alpha < 1$) by both the distribution of the means and standard deviation, but the individual hierarchical simulations are not able to be discriminated using this method.

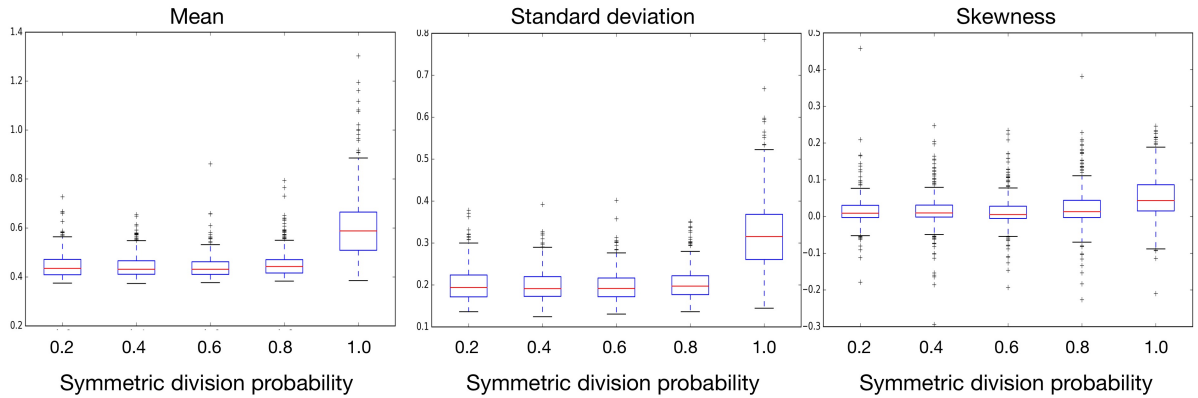


Figure 3.12: Distribution of distance to clonal group centroid as a function of symmetric division probability. For each symmetric division probability, we show a box whisker plot of the distribution of the moments of normalised distance to centroid for the entire simulation. Left: means; middle: standard deviation; right: skewness. In each case, the distance to centroid for each point measured was divided by the square root of the number of cells in the colony.

Ellipse eccentricity

As a second attempt to differentiate between the differing TIC symmetric division probabilities using the shape of the colonies, we look at the distribution of eccentricity (major:minor axis ratio) for ellipses fit to each colony. To fit these ellipses we follow the

procedure suggested by Fitzgibbon et al. [74] and find the centre, axis of rotation and major and minor axes of the ellipse that minimizes the least squares distance of all points within a given clonal group.

To compare the simulation outputs, we then consider the 50 simulations from the same range of symmetric division as before. We consider colonies of size greater than 25 in this comparison, though the results are robust to a number of different observation thresholds (results not shown). We then plot the distribution of the eccentricity for each family of simulations in Figure 3.13. Similar to our findings when we studied circularity and diversity, we find that it is possible to discriminate between the hierarchical and non-hierarchical families of simulations, but not to make any statement about relative TIC symmetric division probability.

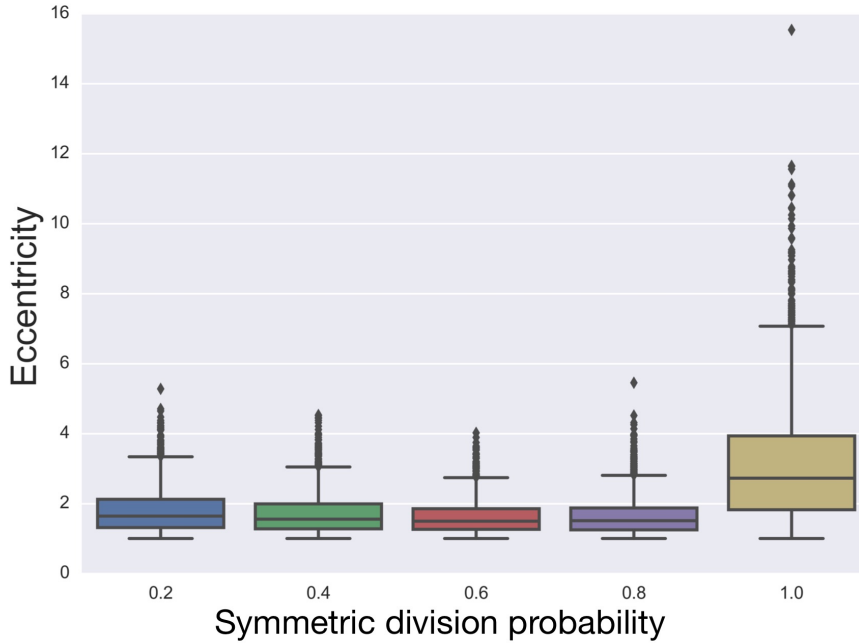


Figure 3.13: Comparison of distribution of colony ellipse eccentricities for a range of symmetric division simulations. For 50 simulations from each of a range of symmetric division probability (0.2, 0.4, 0.6, 0.8 and 1.0) we perform a direct least squares fitting of each colony to an ellipse. We then plot the distribution of the eccentricities of these fit ellipses. As in the previous circularity measure, we can discern the TIC-driven ($\alpha < 1$) to non-hierarchical ($\alpha = 1$) simulations, but can not differentiate between the different symmetric division probabilities of the TIC-driven tumour simulations.

Shape measure comparison

For both measures of clonal group shape, the circularity metric and the distribution of ellipse eccentricity we observe the same result: whilst we can discriminate between hierarchically organised and ‘flat’ proliferative architectures, we can not discriminate

between specific TIC symmetric division rates. Thus far we have considered our analyses based on perfect information regarding our system. In reality, both experimentally and in patients, we need to deal with partial or imperfect information. Therefore we next ask if we can determine whether or not a simulated tumour is, or is not hierarchically organised from ‘imperfect information’: in this case from the limited information typically available in the clinic from tissue obtained through biopsy.

3.3.4 Inferring symmetric division probability from imperfect information

Our analysis thus far has been predicated on ‘perfect’ knowledge of the tumour architecture and cellular make-up. We have learned that different rule sets governing proliferation can create different spatial patterns, and have shown that, using specific measures of colony size distribution, we can discriminate between the underlying proliferative rule sets, at least to the level of TIC-driven versus non-TIC-driven. In reality, this sort of ‘perfect’ information is unattainable, and we must rely on the imperfect information provided by biopsies. The standard of care for cancer patients only requires a single biopsy to confirm the underlying histology, and from this we make a diagnosis. The now accepted reality of tumour heterogeneity is therefore entirely ignored in this paradigm. Even for large scale sequencing endeavours, like the cancer genome atlas (TCGA), most tumours are characterized by a single biopsy, even in the face of overwhelming evidence that biopsying different parts of the tumour can yield very different genotypic and phenotypic groups of cells [222].

Biopsying the spatial model

In clinical reality, we cannot access maps of entire tumours, but instead are forced to attempt to understand the diversity of a tumour from small samples, or biopsies. Here, we begin to understand what information can be gleaned from biopsies alone. This is a situation not unlike the challenge faced by ecologists when attempting to determine overall species diversity in a large ecosystem from small samples [103].

To understand what, if any, information regarding the tumour’s overall composition, or underlying biological rule sets, can be gleaned from the ‘imperfect’ information inherent in a biopsy, we create an algorithm to spatially sample our simulated tumours. A similarly inspired computational approach to understanding the risk of carcinogenesis from biopsies was taken by Dhawan et al. [52]. In this work, the authors developed a CA model of mutation and progression from pre-malignant to malignant tissue to study biomarker development and biopsy strategies. They found that more biopsies were better than one, that several biopsies through time were better than the same number all at a single

‘latest’ time point and that measures of clonal diversity were superior to measures looking for specific alterations. Another approach to the question of imperfect information was taken by Poleczuk et al. [188], using a model of differing phenotypes. They found diminishing returns with increasing biopsies when the measured output was a fraction of total phenotypes sampled. In their model, TICs could mutate a number of phenotypic parameters which contributed to changes in growth, while here, we consider the case of neutral mutation. Further, the sampling was done in a radial fashion from different angles, whereas we, like Dhawan et al. [52], will consider instead in-plane, ‘punch’ biopsies. Further, as we discussed in Section 3.1, Kostadinov et al. [128] suggest that spatially segregated biopsies taken from a tumour can provide false evidence for higher mutation rates. With these results and limitations in mind, we will formulate a method to biopsy our simulated tumours.

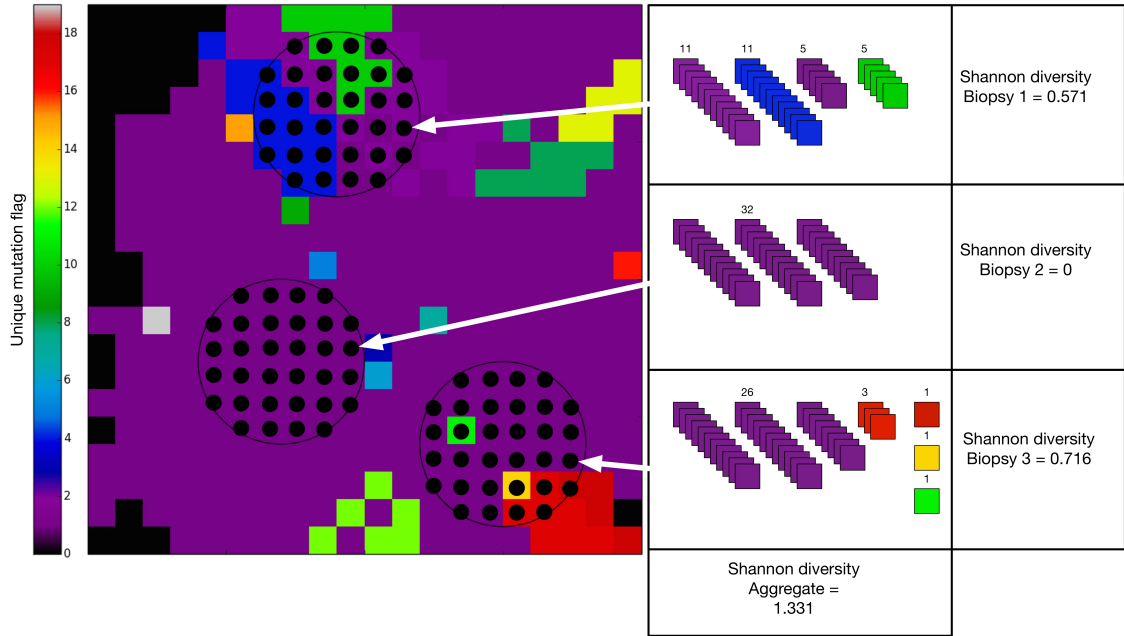


Figure 3.14: Biopsying the CA model. We biopsy uniformly at random and do not allow overlap (sample without replacement). We can calculate a diversity measure either within, or in the aggregate of, samples. Shannon indices are reported as absolute values here.

To simulate a tumour biopsy, we choose a sample of a given radius, r , measured in cell diameters, from our computational domain. For computational efficiency, we first prescribe a neighbourhood function which defines the (approximately) circular domain surrounding a given point of the chosen radius, r . After choosing the point, either randomly or at a defined location, we then record each automaton element within the prescribed domain. From this sample, we can then calculate the Shannon index from the subset of ‘cells’ and compare it to the whole, or to one another. This method is illustrated in Figure 3.14.

As information about spatial organisation is not available from these limited samples, we ask: can we infer the overall diversity of a tumour from biopsy alone? And, if so, how many biopsies do we need and of what size?

3.3.5 The influence of biopsy size on diversity estimation

Trivially, as the size of the biopsy approaches the size of the tumour, we approach the real diversity, but this is not feasible biologically. Here, we briefly investigate how the changing size of biopsies affects the resulting estimate of diversity. To begin, we will compare two sets of biopsies, taken from the same locations, of different sizes. In particular, we choose 40 non-overlapping sites from within the domain sampled uniformly at random which are at least 30 cell diameters from the edge. We then compare the resulting diversity as sampled by biopsies of radius 10 and then radius 30 from the same sites. We plot, in Figure 3.15, the actual biopsies in the top panels, and the distribution of individual biopsy Shannon indices (blue) along with associated kernel density estimation in the bottom panels for the biopsies of radius 10 (left) and 30 (right). We find, as we expect, that the larger biopsies give us, on average, a better approximation of the overall diversity. This effect seems to have a diminishing return with larger and larger sizes however (see Figure A.1), which we will explore in more detail later.

3.3.6 Comparison of diversity prediction from biopsy in TIC-driven and non-TIC-driven tumours

Beyond questions of optimal biopsy strategy (number and size), we return to the question regarding whether or not we can discriminate between CA states generated by hierarchical vs. non-hierarchical proliferative rule sets. To study this, we compare the deviation from the ‘true’, tissue-scale diversity of a set of biopsies from each of two simulations: one with symmetric division probability 0.2 and one non-hierarchical (symmetric division probability 1.0). In Figure 3.16 we plot the randomly chosen biopsy positions (the same positions for both simulated tumours) for the simulated tumour with TIC symmetric division probability $\alpha = 0.2$ on the left, and with TIC symmetric division probability $\alpha = 1.0$ on the right in the top panels. Below each of these we plot a histogram of the individual biopsies and the deviation from the ‘true’ Shannon index as well as an overlying kernel density estimate, and report the mean and standard deviation. In red, we also plot the Shannon index of the aggregate of all of the 50 biopsies.

While the average individual biopsy (mean of the distribution) is significantly worse when sampling the hierarchical CA, when we compare the Shannon index of the aggregated samples, they are quite similar. In this example, we took a large number of biopsies,

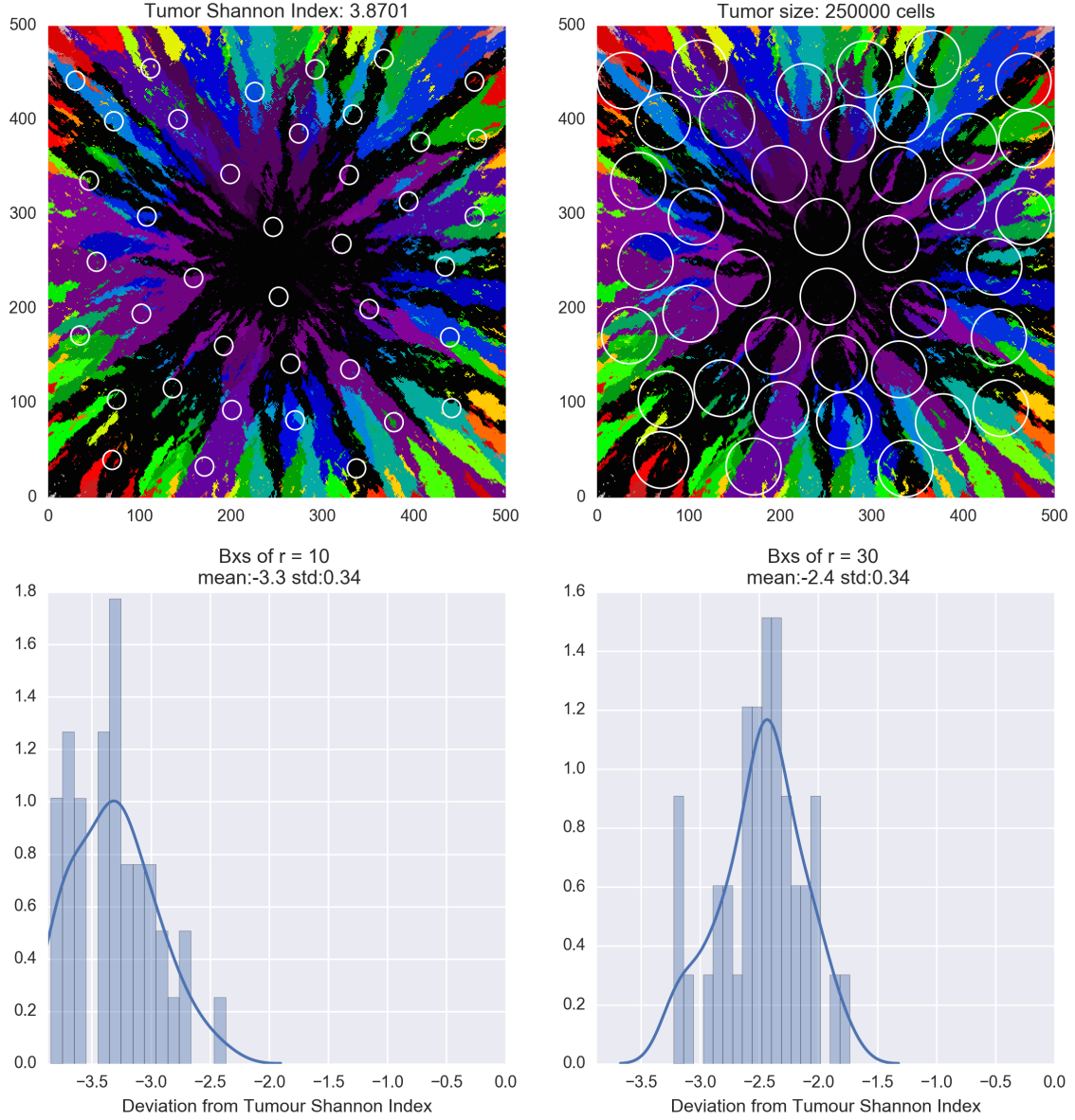


Figure 3.15: Increasing biopsy size, with fixed position, decreases average deviation from tumour Shannon index. We plot the distribution of the difference between the Shannon index of the tumour and 40 individual biopsies of differing radii ($r = 10$ left, $r = 30$ right) along with the associated kernel density. We find that increasing the size of the biopsy decreases the deviation from the tumour Shannon index on average (mean decreases from -3.3 to -2.4 .)

which leads us to ask how quickly we could make up for (necessarily) small biopsies by taking, and aggregating, more than one? To answer this, we take 250 sets of a given, increasing, number of biopsies of a constant size from the same positions from each of two simulations ($\alpha = 0.2$ and 1.0), and aggregate them. We then plot the distribution of the normalised deviation from the tissue-scale Shannon index for each family of biopsies as a paired box-whisker plot in Figure 3.17. We observe that, given a single, or even two biopsies, the ability to predict the tissue-scale diversity is worse, on average, in the hierarchical ($\alpha = 0.2$) simulation. However, unexpectedly, we find that as we increase

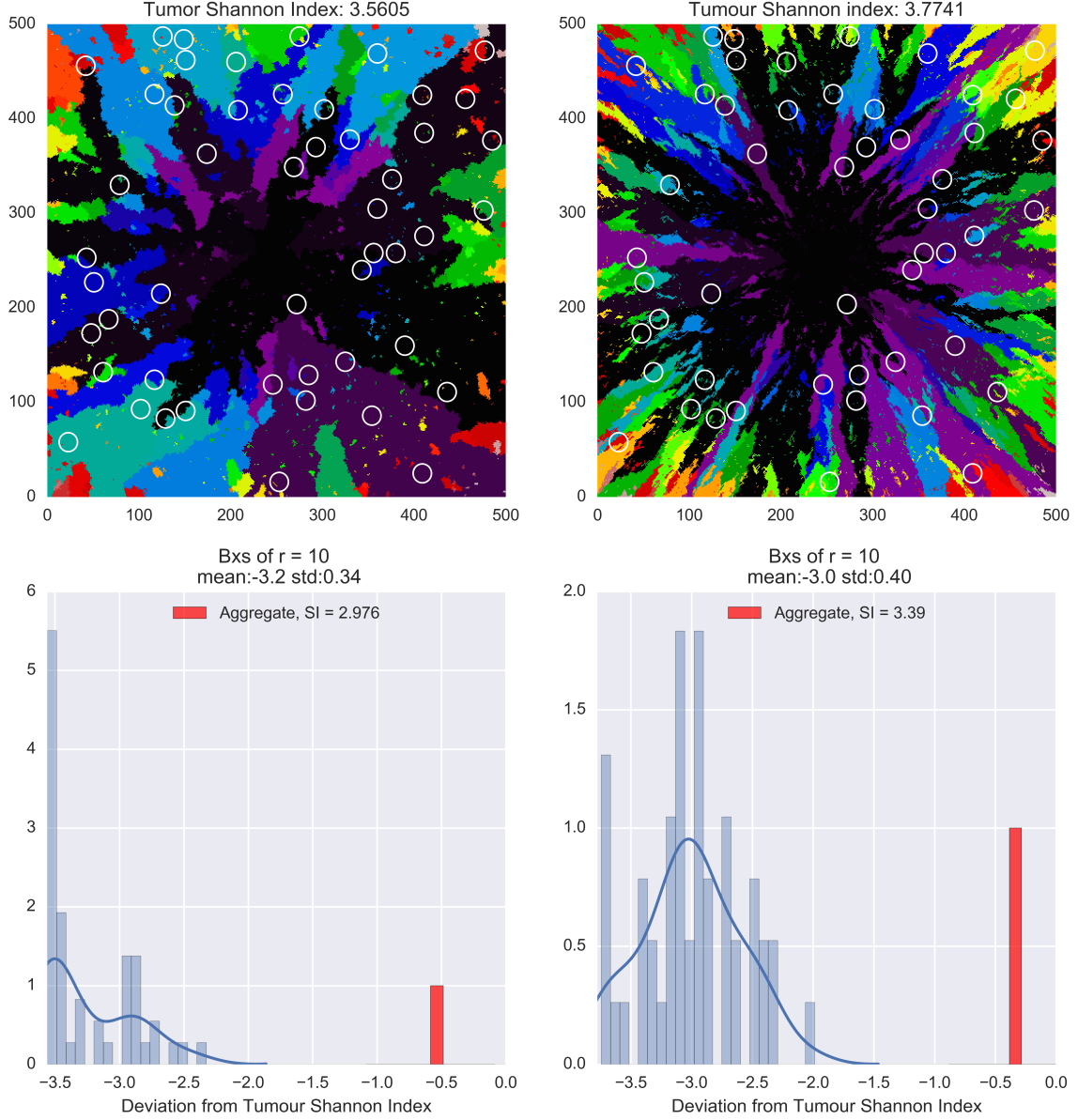


Figure 3.16: Individual biopsies taken from non-hierarchical tumours are better representations of true diversity than those from hierarchical tumours. We take biopsies from 50 non-overlapping positions drawn uniformly at random from the computational domain of two simulations with differing symmetric division probabilities (left: $\alpha = 0.2$, right: $\alpha = 1.0$). We find, on average, and in aggregate, that the biopsies taken from the non-hierarchical simulation better represent the tissue-scale diversity.

the number of biopsies, the predictive ability of the aggregated Shannon index for the hierarchical simulation begins to surpass that of the non-hierarchical one.

Number vs. size of biopsies: diminishing returns

As we have seen already, there is a trade-off between the number of biopsies taken and their size, and others have noted that there are diminishing returns in the number of

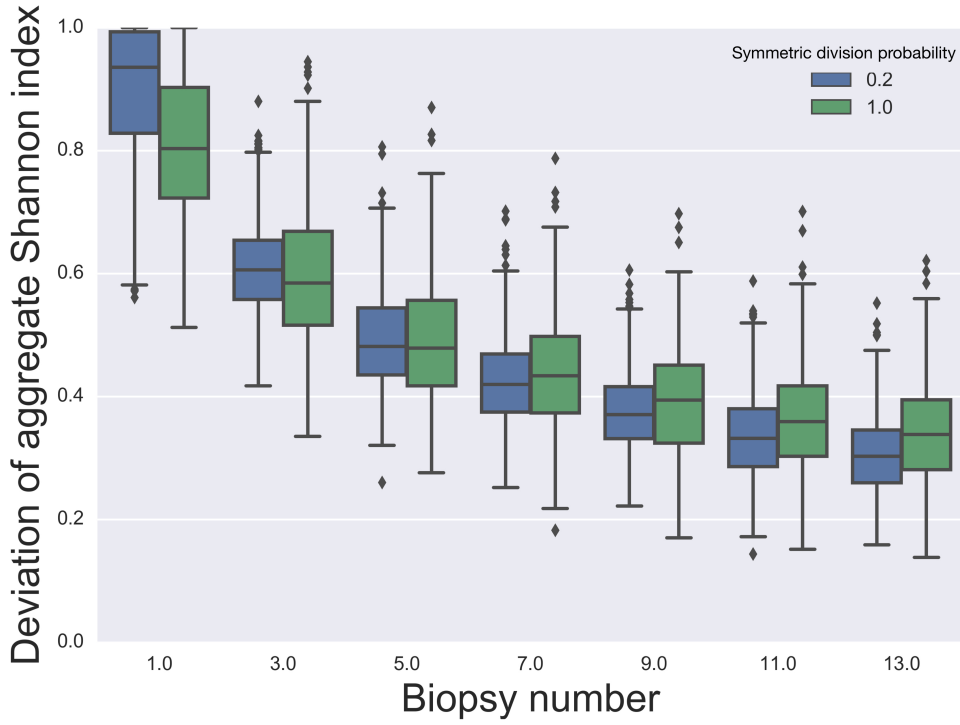


Figure 3.17: The accuracy of diversity prediction between hierarchical and non-hierarchical tumours changes with increasing number of biopsies. We plot the distribution of the difference between aggregated Shannon index and the tissue-scale Shannon index in a box-whisker plot. Each box represents the distribution of 500 sets of a number of biopsies of uniform size (here, 10) taken uniformly at random from a 500×500 computational domain of a hierarchical (symmetric division $\alpha = 0.2$, blue) and non-hierarchical (symmetric division probability $\alpha = 1.0$, green) simulation. As the number of biopsies increases past five, we are better able to predict the hierarchical simulation’s diversity.

biopsies taken [54]. To understand how these trade-offs interact with our ability to determine the true diversity of a tumour through limited sampling, we study this more closely by varying both the number of biopsies taken and the size of the biopsies, and compute iso-surfaces of total size of aggregate tissue sampled for ease of visualisation. In Figure 3.18 we plot results from a representative simulation with TIC symmetric division probability $\alpha = 0.2$ (left), and TIC symmetric division probability $\alpha = 1.0$ (right). The x -axis represents biopsy area, and the y -axis represents the number of biopsies. The cyan lines represent iso-surfaces approximating equivalent total cell numbers (cells per biopsy $\times$ number of biopsies). The colourbar represents the normalised deviation from the ‘true’ Shannon diversity, which is measured as the mean of the aggregated Shannon index from 20 sets of the denoted number and size of non-overlapping biopsies, normalised by the ‘true’ diversity of the simulation. We find both plots to be asymmetric in that there is an advantage to taking more biopsies that is greater than taking larger biopsies. Further, we find that while a single biopsy is equivalent, or possibly better, in predicting true

diversity in the $\alpha = 0.2$ simulation (as we saw in Figure 3.17), as the number of biopsies increases, this changes rapidly to favouring the $\alpha = 1.0$ simulation.

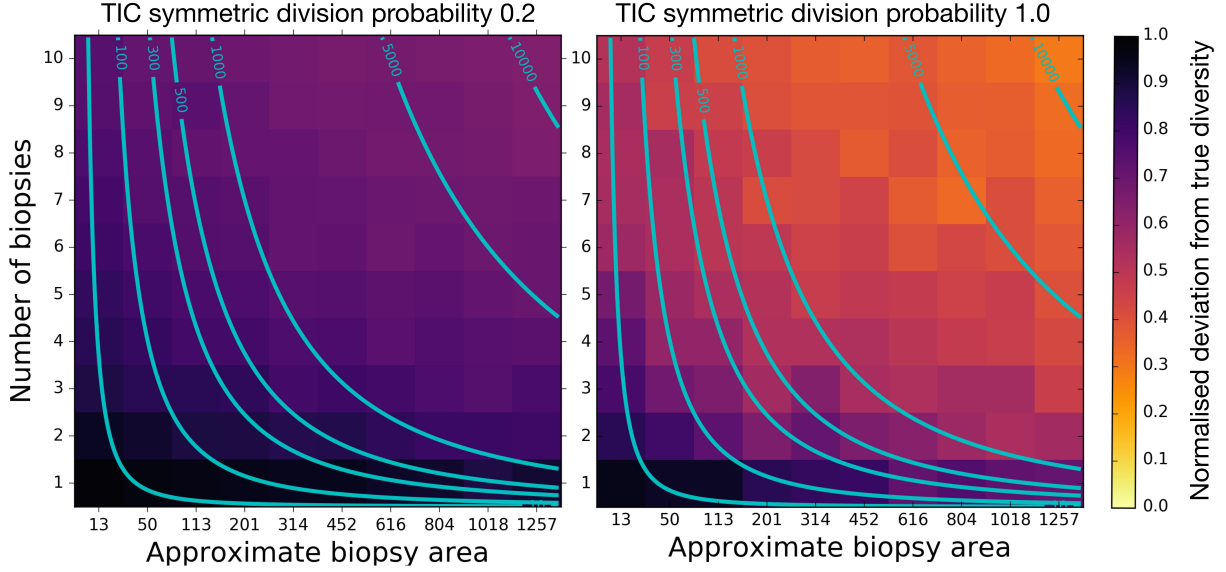


Figure 3.18: Comparison of trade-off between biopsy size and number for different symmetric division probabilities. We plot a heat map of normalised deviation from ‘true’ diversity for two simulations: one with TIC symmetric division probability $\alpha = 0.2$ (left); and for TIC symmetric division probability $\alpha = 1.0$ (right). We find, as before, that for low numbers of small biopsies, the deviation from ‘true’ diversity is nearly equivalent, but as the size and number increase, the deviation from ‘true’ diversity is lower for the non-hierarchical simulation, $\alpha = 1.0$. Further, we see that the plot is asymmetric for both simulations, favouring more biopsies compared to larger ones.

3.4 Discussion

In this chapter we have explored a model of tumour growth with neutral mutation and an underlying hierarchical cell proliferation structure. We have considered how spatial constraints on proliferation and a varying symmetric division probability affect the resultant diversity of the tumours as well as several measures of spatial organisation. We have purposefully reduced the complexity of the model as much as possible, specifically not including heterogeneous space or mutations that confer any fitness advantage in order to be able to tease apart effects from just space and underlying cellular proliferative heterogeneity. We have focused on understanding the differences in overall diversity, as measured by the Shannon index, resulting from different parameters of the underlying cellular growth rule set, as well as spatial metrics of colony size and shape – with the overarching aim of finding translatable measures to help infer TIC symmetric division rate from clinically accessible information.

We are by no means the first to use colony (or clone) size as a proxy for determining the underlying evolutionary dynamics of stem cells. For example, Driessens et al. [61] used an *in vivo* cell-labelling experiments in both pre-cancerous (papilloma) and invasive squamous cell cancers in mice to parameterise a branching process model of clonal expansion. The authors find differing, characteristic colony size distributions between the pre-cancerous and cancerous lesions, which stem from their assumptions on symmetric division probability in the underlying branching process. These predictions are tested against data taken from multiple 2-d sections of mouse epithelium with labelled cancer cells, and are found to agree. While direct comparison between the results of Driessens et al. and ours is not possible because of structural model differences, the qualitative agreement about high versus low symmetric division rate and changing colony size distributions holds.

Beyond describing these differences, we have worked to quantify the information that can be gleaned from a variety of ‘imperfect’, or incomplete, sets of information - analogous to a tumour biopsy. We have found, unsurprisingly, that larger biopsies, or more biopsies aggregated together, provide better estimates of overall diversity. However, we find that there are diminishing returns to increasing both size and number of biopsies. This is important because increasing both becomes quickly clinically, biologically and financially infeasible, not to mention dangerous and painful for the patient. Defining an optimal size and number of biopsies could, for a given parameter set, be derived - but given the current limitations (mostly clinical) on number and size of biopsies, another route is desirable.

While a strong assumption, our simplification that only TICs accumulate mutations is not new [66, 223]. We realize however, that this could, in certain cases considered, change the quantitative results. This being said, using the purely hierarchical formalism that we have adopted means that in the long time limit, TACs will eventually die off. The only cells comprising a long-lasting proportion of anything sampled will therefore be TICs, or their relatively recent (evolutionary time) TAC progeny. These recent progeny will carry their parental TIC genomes plus only small differences based on the background mutation rate and TAC lifespan. While it has been shown that TICs also cycle [127], there is a biological separation of time scales that suggests considering TICs alone is a reasonable assumption, which is also made in fate mapping studies [143, 219]. While this formalism is hotly debated, we adopt it here, and extending our analysis to consider TAC mutation will be considered later.

We have made progress towards finding a measure which can be used to predict the underlying tumour proliferative organisation in both tumours which are, and are not, constrained by space. We have provided insight into the relative value of multiple biopsies in tumours which are hierarchically organised compared to those which are not. Further,

we have identified the Shannon index as a way to discriminate between different TIC symmetric division rates in non-spatially constrained tumours, like leukemias. Disappointingly, we have not yet found a way to discriminate between TIC symmetric division probabilities for a spatially constrained tumour. Remedying this will be the focus of the next chapter.

Chapter 4

Inferring tumour proliferative organisation from phylogenetic tree measures

4.1 Introduction

In Chapter 3 we introduced spatial and non-spatial stochastic models of clonal evolution in a hierarchically organized tumour. We found that the Shannon index and spatial measures of clone size and shape provide a reliable biomarker for TIC symmetric division probability, α , in the non-spatial model. In contrast, for the spatial model the Shannon index allows us only to discriminate between hierarchically organized tumours ($\alpha < 1$) and those comprised of TICs only ($\alpha = 1$); even with perfect information of tumour composition, our ability to determine the value of α is limited. While this is a step forward, as determining symmetric division probability of a tumour from clinically available data is yet lacking, we have not yet determined a method of doing so for spatially constrained tumours. Further, the spatial measures considered, clone size and shape, are not clinically accessible except for single, or possibly several, mutations. To address these issues, in this chapter we turn to another property of the genetic heterogeneity that is present in tumours and measurable from biopsies: the mutational ‘life history’ that is encoded in the branching structure of their phylogeny. We develop a method to reconstruct a tumour’s phylogenetic tree and then use our stochastic tumour growth model to assess tree-based measures that can be used to discriminate between hierarchically organised tumours with different TIC symmetric division probabilities.

4.1.1 Phylogenetic reconstruction in human disease

There is growing interest in using reconstructed phylogenies to determine many aspects of tumour evolution [89], including the origin within the primary of individual metastatic

tumours [90, 170], the effect of chemotherapy on primary and metastatic sites [71, 169], and the genetic heterogeneity present within a primary tumour [222]. This application to cancer was preceded by similar work in ecology based on morphological dimorphisms and speciation events and, more recently, the field of phylodynamics [95], which combines both population dynamics and phylogenetics in the study of infectious diseases [105, 176]; see Figure 4.1 for examples.

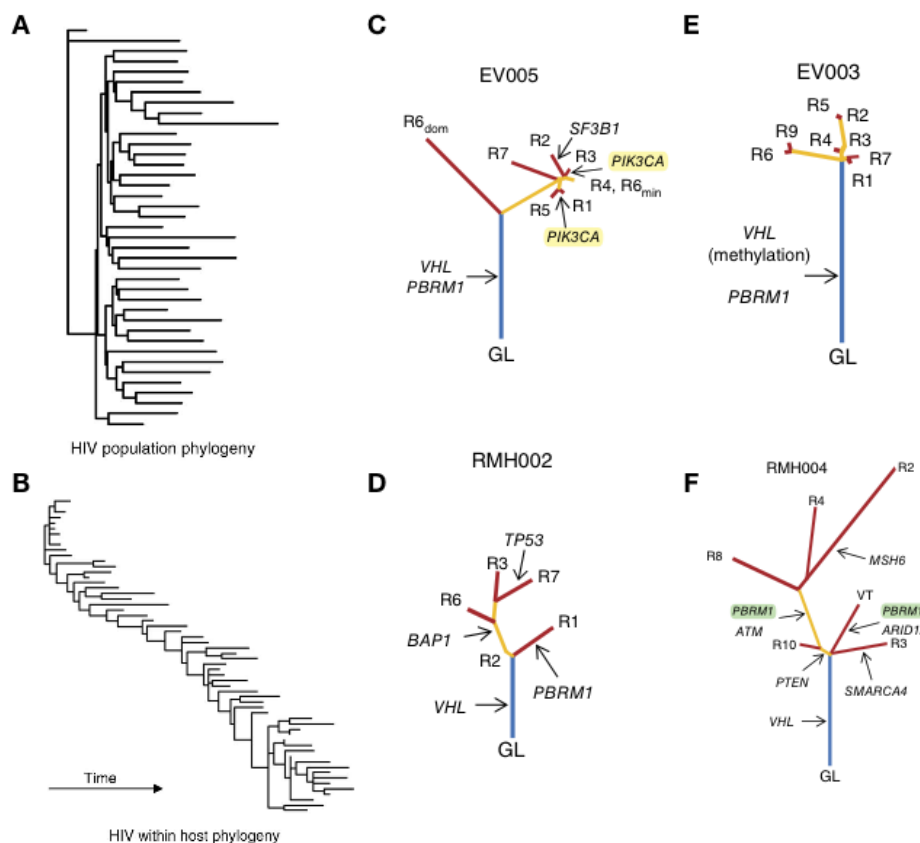


Figure 4.1: Phylogenetic reconstruction in the study of evolution in infectious diseases and cancers. Striking differences are seen between reconstructed phylogenies of HIV *between* many patients (A) and *within* a single patient (B). While fewer data are available in cancer, it is evident that tumours also display significantly differing tree topologies. (C-F) We can see four quite different tree topologies revealed by multi-region sequencing of a single tumour type, clear cell carcinoma of the kidney. Figures adapted from [95] (A,B) and [89] (C-F).

Of particular relevance to the present chapter is work from Leventhal et al. [136] based on the concept that the phylogenetic tree contains a ‘fingerprint’ of the underlying biological process. They further suggest that this ‘fingerprint’ can be used to determine the evolutionary process driving the population in question, a concept which has been explored for ecology and biogeography by Mooers and Heard [164], but to date not for cancer. The authors consider a susceptible-infected-resistant (SIR) model of human immunodeficiency virus (HIV) transmission with mutation on a network where individual

humans are represented by vertices and contact between them by edges, and investigate whether the topology of the underlying contact network structure can be inferred from the resulting disease phylogenies. To answer this question, the authors choose several families of graphs, including the ‘small-world’ network [241], the Erdős-Renyi [68] random graph and the ‘scale-free’ network [24], and the complete graph. Holding the parameters of the SIR model itself constant, the authors simulated a range of epidemics and measured the phylogenetic trees of the resulting *in silico* viruses. They found that using statistical measures of the resulting phylogenetic trees, in particular a normalized version of the Sackin index [203], they could predict the underlying network structure, or at least discriminate between the qualitatively different topologies tested. We implement a similar approach here, to discriminate between simulations with different symmetric division probabilities.

4.1.2 Aim of work

In this chapter we aim to extend our work from Chapter 3. Specifically, we hope to use an approach similar to Leventhal et al. [136] to discriminate between the underlying symmetric division probabilities in the spatial model, which eluded us in Chapter 3. To do this, we develop an algorithm to build the phylogenetic tree structures from the simulation life histories, and then seek a measure that allows us to discriminate between the families of simulations with differing symmetric division probability.

4.1.3 Chapter structure

The remainder of this chapter is structured as follows. In Section 4.2 we present an algorithm to construct both the genomes (represented as binary strings) and the branched phylogenetic structure from a stochastic realization of our tumour growth model. We then introduce a suite of statistical measures of phylogenetic tree topology which we apply to simulation outputs for comparison. In Section 4.3.3 we consider how the temporal dynamics of these measures change over the course of tumour growth to assess whether they are robust to tumour size changes, and then to changes in mutation frequency in Section 4.3.4 In Section 4.3.6 we attempt to discriminate between simulations using information gleaned from our biopsy method presented in Chapter 3, and then conclude with a discussion of clinical utility of these measures in Section 4.4.

4.2 Methods

While experimentalists and clinicians must infer phylogenies from incomplete data, construction of the ‘true’ phylogeny is possible in our model as we can capture the entire

life history of the simulated tumour. Thus, we can test whether phylogenetic tree-based measures are able to discriminate TIC symmetric division probability in the case where the ‘ground truth’ is known. At each time step we record the spatial location of each individual cell with its mutational flag. Additionally, we record every mutational event as it occurs as a simple list of parent-child pairs, which gives us an evolutionary ‘life history’. We must then recursively construct the phylogenetic tree from this life history.

4.2.1 Phylogenetic tree reconstruction algorithm

To create the complete tree data structure that we need to begin our quantitative analyses we need only the information encoding the mutational events from our stochastic simulation. To this end, we create a list of unique parent-child pairs using the life history of mutational events. We then apply an iterative process in which each child is added as a subnode below the parent (from the unique parent-child pair). This process is continued until all parent-child pairs are added to the structure, and the tree is complete. The function to create these trees was written in the python programming language, specifically for the purpose of this chapter, and is available on request.

4.2.2 Qualitative comparison of reconstructed trees

To investigate the differences in phylogenies for simulations with different underlying rule sets, we first construct and visualize the phylogenies constructed from three example simulations with differing TIC symmetric division probabilities in Figure 4.2. It is clear by inspection that the number of mutations increases with symmetric division probability (more branches). However, the topology of the trees is not as easy to parse visually. For ease of visualization the trees depicted in Figure 4.2 have been pruned of all terminal nodes (also called leaves) with no children of their own. While this transformation does effect the quantitative results, it does not qualitatively affect the resultant phylogenetic tree statistic ranks (see Figure B.2). Regardless, we will use the full trees going forward in our measures.

4.2.3 Candidate tree-based measures for model comparison

Visual inspection of Figure 4.2 suggests that simulations with different TIC symmetric division probabilities generate phylogenetic trees that are qualitatively distinct. However, to make meaningful conclusions we must determine a method of quantitative comparison. Here we present several measures useful in summarising and simplifying the comparison of phylogenetic trees. The most commonly studied property of phylogenetic tree shape is its balance, which is defined as the degree to which internal nodes (branch points) have

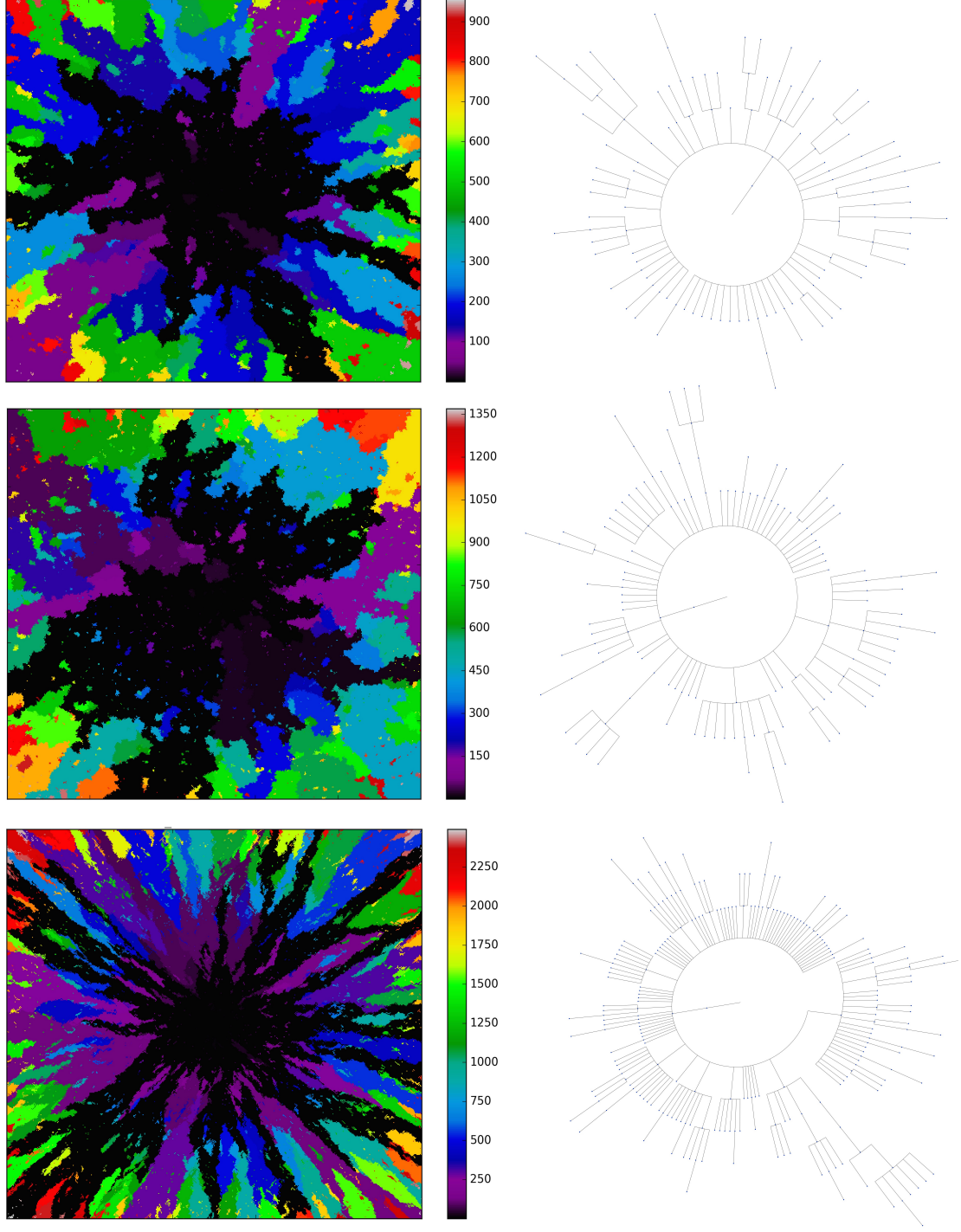


Figure 4.2: Three example simulations with increasing symmetric division probability (0.2, 0.6 and 1.0 from top to bottom) and their associated phylogenetic trees. Each plot is the result of a single stochastic realisation of our spatial CA model. Each simulation is initiated with a single TIC and complete when the domain is full, in this case 250,000 cells. Allowed TAC divisions, $\beta = 4$ and mutation probability, $\lambda = 0.01$. Visualised trees (right) have been pruned of all leaves for ease of visualisation, which does not qualitatively affect measure rank (see B.2).

the same number of children as one another. Balance (or imbalance) indices depend only on the branching topology of trees, and not on other factors like branch length or other morphological or biological features of the terminal branches (often called leaves). Since the first description of a balance index by Sackin [203], many others have been proposed with slightly differing properties [160]. One of the first papers to present a systematic comparison of a suite of balance indices (often denoted with the letter ‘B’) and indices of imbalance (denoted with ‘I’) was by Shao and Sokal [215], who reported striking differences between the studies’ measures. Their central message was that different measures on trees can give insight into different aspects of the underlying processes governing the interactions, and one should thus consider several measures for any given tree or family of trees. In this chapter we will consider several tree topology-based measures.

Before describing the measures, it is worthwhile to briefly define the terms which are used to describe trees, and the two basic underlying stochastic models which have been proposed to describe neutral evolution and the resulting topologies. As discussed in Chapter 1, phylogenetic trees are mathematical objects which describe the evolutionary relationship between individuals with different physical traits from one another, or in the case of our model, different mutational combinations (genotypes). In our model, each simulation begins with a cell with mutation flag 1, or a genotype with the first allele mutated (1000...), termed the ‘root’, and evolution progresses stochastically, by adding individual mutations at subsequent alleles and increasing the mutation flag, as described in Figure 3.3. At each mutational event, an evolutionary branch point is created, which is termed a node in phylogenetic tree terminology. If this node gives rise to no other children during the simulation, it is termed a terminal node, or leaf. There are two common classically referenced models which bear mention here as well, as many tree topology-measures are normalized against them. The first, described by Yule in 1924 [252], sometimes termed the ‘equal rate Markov’ model, begins with a single root and proceeds by replacing, uniformly at random, a given leaf with a node with two children of its own. The process continues until the desired number of leaves exist. The other main model, termed the ‘Proportional to Distinguishable Arrangements’ or uniform model, was described by Rosen [202]. This model, which is truly a model of tree growth rather than an explicitly evolutionary process, begins as does the Yule model (and indeed ours) with a single node labelled 1. At each update step, a new leaf is added to the tree to any point, either internal node or extant leaf. These models will serve as normalisation factors in several of the measures we will present in the following.

Sackin index

The Sackin index was the first statistic used to understand the balance of a phylogenetic tree [203, 215]. To compute this statistic, one sums the number of ancestors (N_i) for each of the n terminal nodes of the tree. The classical representation for this metric is:

$$I_s^n = \sum_{i=1}^n N_i. \quad (4.1)$$

It is clear that this index increases with tree size. Indeed the expectation of the measure, $E[I_s^n]$, under the Yule growth model grows as $2n \log n$ [252]. Therefore, one cannot perform a meaningful comparison of Sackin indices of trees generated from tumours simulated with the same parameter values, but to different sizes.

Normalized Sackin index

To address this, several normalisations to the Sackin index have been proposed, two of which we explore here. In particular, one can normalise the Sackin index of a phylogenetic tree to the expectation value of a similarly sized tree, under the assumption of growth by the Yule model. This is defined as

$$I_{Yule}^n = \frac{I_s^n - 2n \sum_{j=2}^{n+1} j^{-1}}{n}, \quad (4.2)$$

where n is the total number of terminal nodes. One can alternatively normalise using the Proportional to Distinguishable Arrangements (PDA) model [8, 9, 202] which is simply the Sackin index scaled by $n^{3/2}$.

The B1 statistic

The B1 statistic, originally described by Shao and Sokal [215], considers the balance of a tree. To calculate the measure, one uses all i internal nodes of the tree with the exception of the root (the founding cell). Then, for each non-root internal node, the maximum number of nodes traversed along the longest possible path to a terminal node is counted and termed M_j . Therefore, for each i there is a single value M_j . The B1 statistic is then defined as

$$B1 = \sum_i \frac{1}{M_j} \quad \forall i \neq root. \quad (4.3)$$

$\bar{N}$

$\bar{N}$ reports the average number of nodes above a terminal node. To compute this, we sum the path from each terminal node to the root, and divide by the number of terminal nodes. An alternative definition is the Sackin index ‘normalised’ by the number of terminal nodes. For a more complete review and comparison of the measures presented here, and others, see Blum et al. [29] and Shao and Sokal [215].

Example tree measures

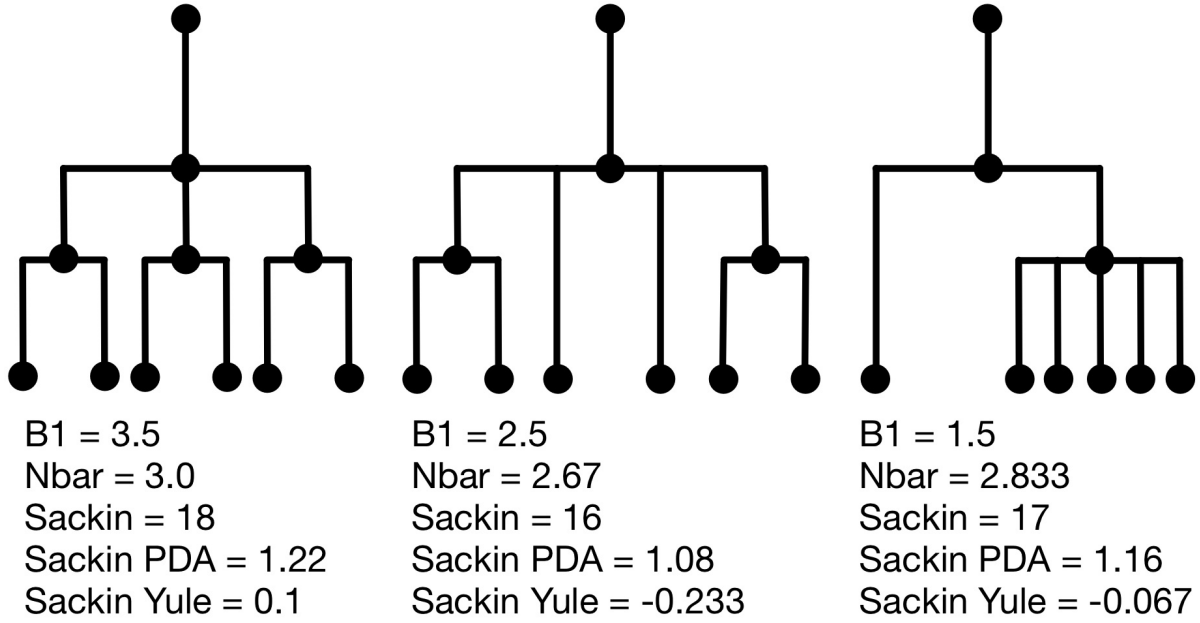


Figure 4.3: Example phylogenetic trees and their measures. From left to right the trees contain 4, 3 and 2 internal nodes respectively, but the same number (6) of terminal nodes.

Examples of how these measures change on several example trees with equal numbers of leaves (but different numbers of internal nodes) are presented in Figure 4.3. In these examples, we compute each of the presented measures for comparison. From left to right, the trees contain 4, 3 and 2 internal nodes respectively, but the same number (6) of leaves. We note that the measures do not all follow the same pattern. For an exhaustive description of all possible trees with 6 leaves, and the correlation of a larger family of associated measures, see Shao and Sokal [215].

4.3 Results

4.3.1 Measuring trees from simulation

As our primary goal in this chapter is to differentiate between simulated tumours with different TIC symmetric division probabilities, we focus on changes in tree measures in

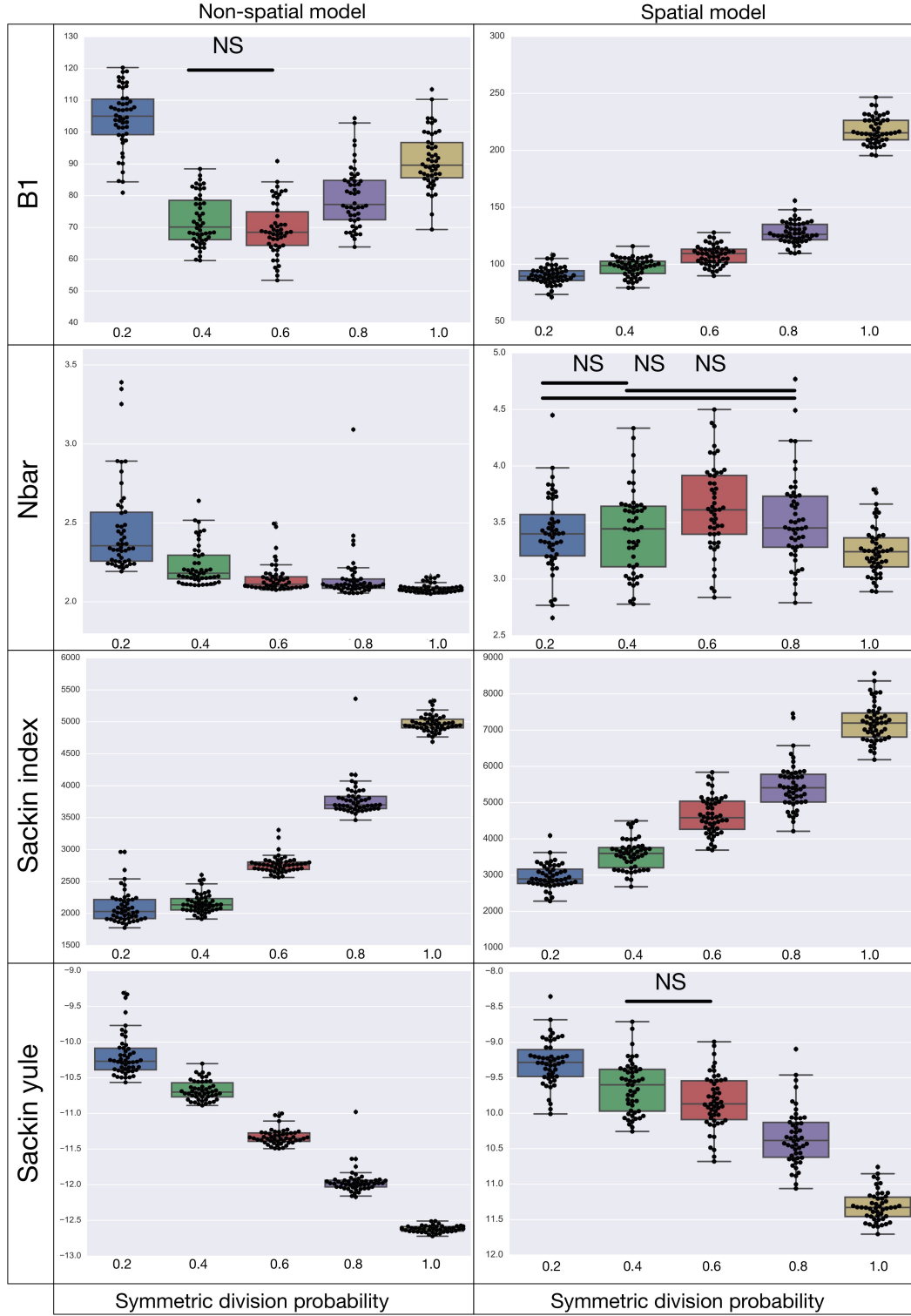


Figure 4.4: A summary of four tree indices measured over a range of symmetric division probability. We plot the distribution of each of four measures of tree balance for the final resultant trees from 50 simulations against symmetric division probability for the non-spatial model (left) and the spatial model (right). All simulations were run with $\beta = 4$ and $\lambda = 0.01$ until a tumour size of 250,000 cells. In each plot we display a box-whisker plot as well as the individual results as points. NS = non-significant by the Wilcoxon rank sum test.

the spatial model. Before doing that however, it is additionally worthwhile to compare the results from the non-spatial model to elucidate changes in tree structure and statistics that are due to spatially constrained proliferation.

To compare the model tree measures, we first perform 50 stochastic simulations of each of our spatial and non-spatial CAs (as presented in Chapter 3) using a range of TIC symmetric division probabilities (0.2, 0.4, 0.6, 0.8 and 1.0), holding mutation probability and TAC lifetime constant ($\lambda = 0.01$ and $\beta = 4$). For each simulation, we construct the resulting phylogenetic tree at tumour size 250,000 cells, as described in Section 4.2. We then measure the value of each summary index presented in Section 4.2.3 for all 50 simulations at the final time point and plot the distribution in a box-whisker plot, which is shown in Figure 4.4 with each data point overlaid in a swarm. Differences between distributions are determined using the Wilcoxon rank sum test.

Effect size: To better understand the impact of a changing symmetric division probability, rather than just use differences between families of simulations, we compute the regression slope, R^2 and p-value of the regression line for each case. For the B1 statistic in the spatial model we find a regression slope of 142.64, $R^2 = 0.72$, $p = 1.74 \times 10^{-71}$ and in the non-spatial model a regression slope of -13.2 , $R^2 = 0.056$, $p = 1.7 \times 10^{-5}$. For the Sackin index in the spatial model we find a regression slope of 5178.61, $R^2 = 0.871$, $p \approx 0$ and in the non-spatial model a regression slope of 3690.92, $R^2 = 0.8673$, $p \approx 0$. For the Yule normalised Sackin index in the spatial model we find a regression slope of -2.380 , $R^2 = 0.743$, $p = 3.25 \times 10^{-75}$ and in the non-spatial model a regression slope of -3.118 , $R^2 = 0.948$, $p \approx 0$. For the $\bar{N}$ statistic in the spatial model we find a regression slope of -0.111 , $R^2 = 0.0075$, $p = 0.172$ and in the non-spatial model a regression slope of -0.457 , $R^2 = 0.303$, $p = 3.33 \times 10^{-21}$. These values are plotted in Figure 4.5.

4.3.2 Variation of tree-based measures with symmetric division probability

Non-spatial model: The results of the non-spatial model are presented in Figure 4.4 (left). Each of the Sackin index based measures (Sackin index, Sackin Yule and $\bar{N}$) are monotonic in symmetric division probability. The Yule normalised Sackin index appears to decrease linearly, while the Sackin index itself and the $\bar{N}$ statistic are increasing superlinearly and decreasing sublinearly, respectively. The B1 statistic, however, reaches a minimum at symmetric division probability 0.6. Using the Wilcoxon rank sum, all distributions are significantly different ($p < 0.05$), except in the B1 statistic between symmetric division probabilities of 0.4 and 0.6, however, the interquartile ranges overlap for all of 0.4, 0.6 and 0.8 for the B1 statistic as well as most of the $\bar{N}$ measures. For an

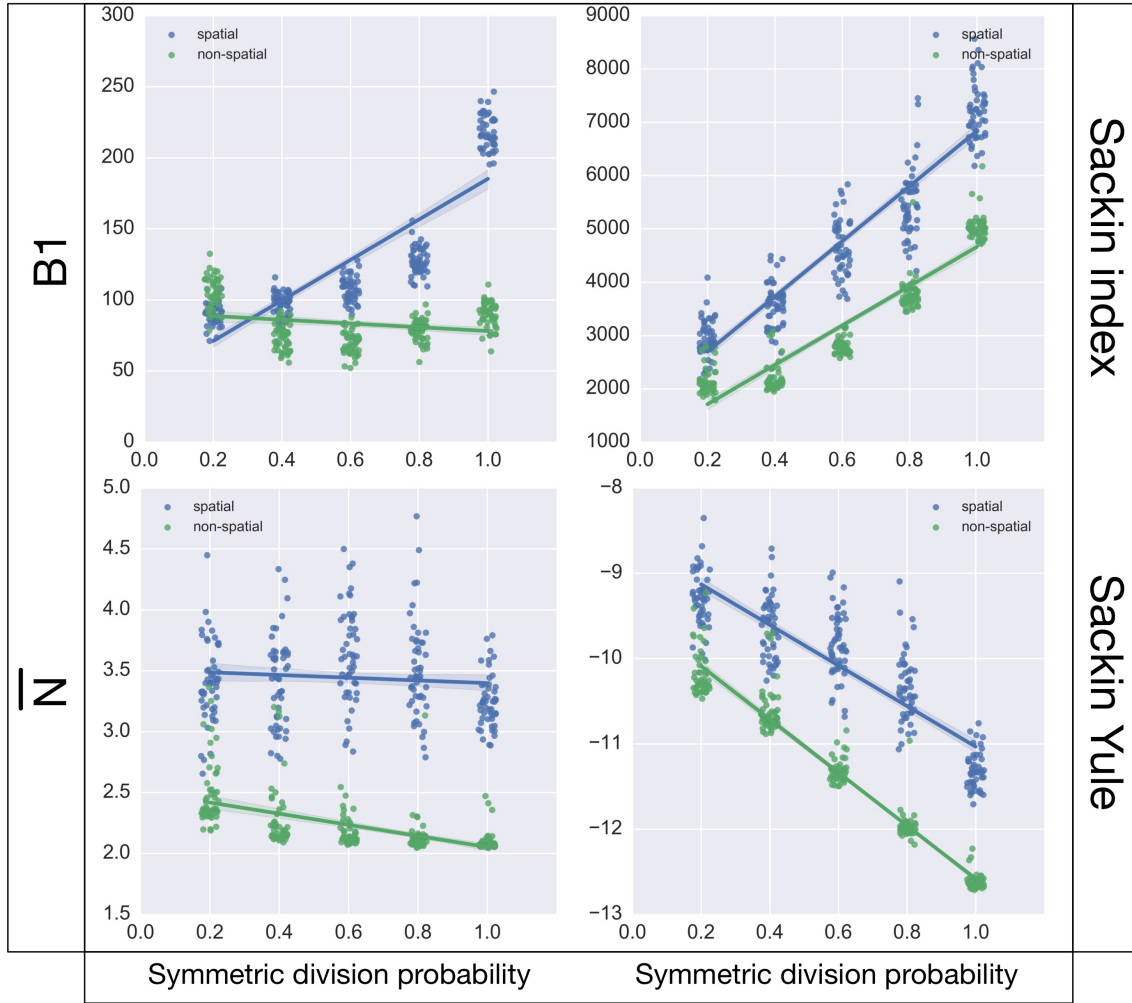


Figure 4.5: Effect size of symmetric division for four tree indices. We plot the effect size for the data shown in Figure 4.4.

exhaustive list of descriptive statistics and pairwise comparisons of these distributions, see Tables B.1 and B.2. It is clear for the non-spatial model that the Yule normalised Sackin index is the most useful for determining symmetric division probability.

Spatial model: The results of the spatial model are presented in Figure 4.4 (right). We find that all of the indices are able to discern between the different symmetric division probabilities except for $\bar{N}$. Of the normalised indices, the B1 statistic has the least overlap in error between symmetric division probabilities, but all measure distributions are significantly different by the Wilcoxon rank sum test ($p < 0.05$) except 0.4 and 0.6 in the Sackin index normalised by the Yule model ($p = 0.08$). This allows an opportunity to infer the symmetric division probability for tumours of similar size - in contrast to the results in measuring only the Shannon diversity index in Chapter 3. For an exhaustive list of descriptive statistics and pairwise comparisons of these distributions, see Tables B.3 and B.4.

4.3.3 Dynamics of tree-based measures during tumour growth

As discussed in Section 4.2.3, most of the measures considered here are strongly dependent on the total number of nodes in the tree. With all other parameters held constant, simply allowing a tumour to grow larger would increase the number of total mutations, and therefore the number of total nodes. To ensure that the differences we have noted are robust to changing tumour size, we will consider how these measures change during the growth of a tumour.

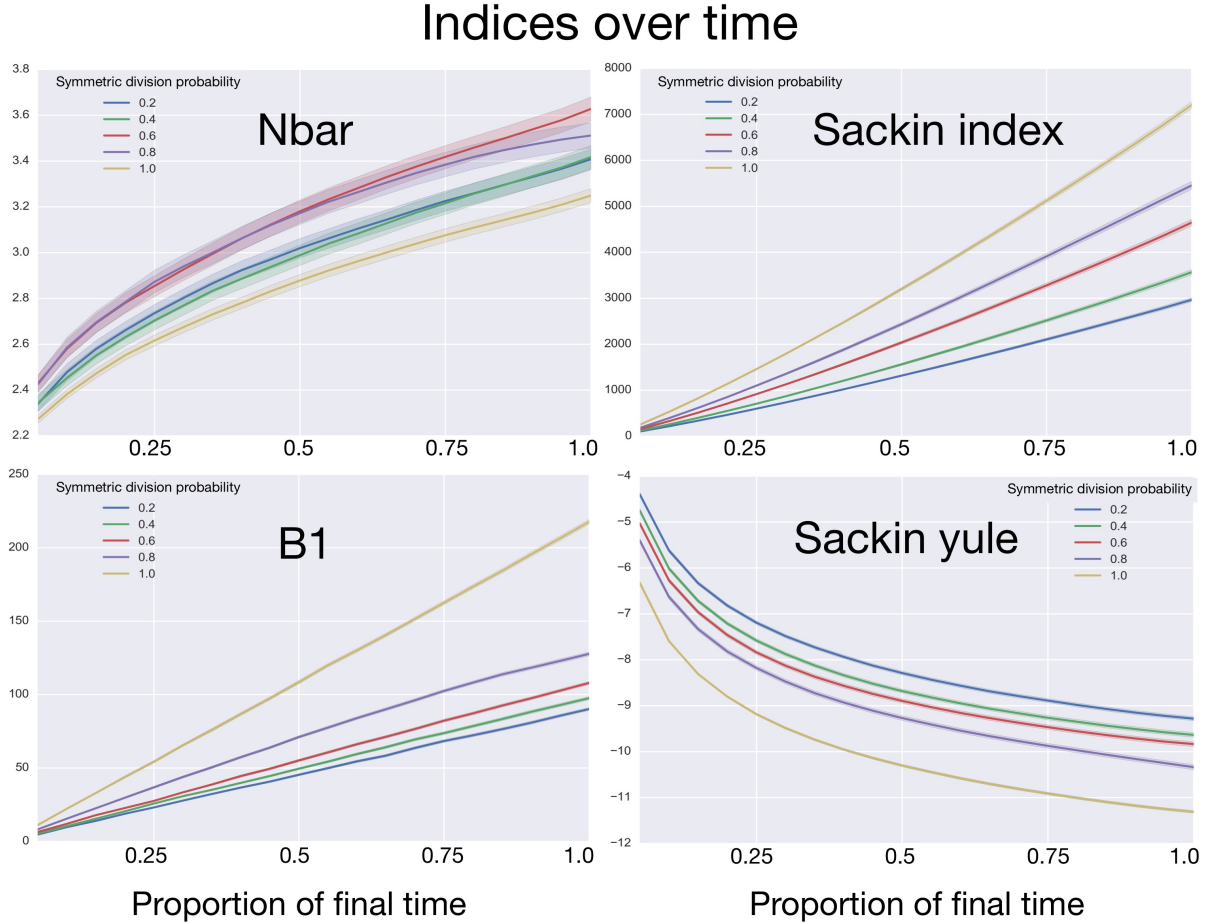


Figure 4.6: Comparing phylogenetic tree measures across symmetric division probability through tumour growth. We plot the average and standard deviation (error bars) of four phylogenetic tree measures for each of the 50 simulations for a range of symmetric division probabilities over the course of tumour growth. Rank is maintained across symmetric division probabilities for each of the 3 tree measures with which we could discriminate between symmetric division probabilities. As before, $\bar{N}$ is not predictive and changes rank throughout tumour growth. All tumours are grown to eventual confluence at 250,000 cells. In all simulations $\beta = 4$ and $\lambda = 0.01$.

To determine how these measures vary over the life of a growing tumour then, we measure the index over the course of each simulation at increasing tumour sizes. To

accomplish this, we use the life history to reconstruct the tree at 20 equally spaced time points during the lifespan of each of the 50 simulations for each symmetric division probability in the spatial model. The time to fill the domain for each of the symmetric division probabilities is quite different as the dynamics of tumours driven by differing symmetric division probabilities are different (see Figure B.1). So, while we break the life history into equally spaced time intervals, as the total times in each family of simulations are different, when we compare across symmetric division probabilities we need to consider this ‘time’ to be a surrogate for tumour size instead of explicitly comparing times. Comparing across tumour size is of greater utility clinically, however, as the age of a given tumour is rarely known, while size can be readily approximated.

After reconstruction, we then create a ‘time’ trace for each statistic. We plot these time traces in Figure 4.6, where each family of 50 simulations (for a given symmetric division probability) is represented by a single trace with the standard deviation represented by the (tight) coloured error bars. We find that for each of the statistics, except $\bar{N}$, the relationships between the symmetric division probabilities are maintained over time, suggesting that, if we know the tumour size, and true phylogeny, we can estimate the relative symmetric division probability between two samples from these measures. This statement must be somewhat qualified by the fact that mutation probability was also held constant for these simulations. While estimating mutation probability is not trivial, significant advances have been made into measuring the speed of the ‘evolutionary clock’ of tumours: essentially a proxy for mutation probability [47]. Further, we found that the rank order of each discriminatory measure holds throughout tumour growth, indeed becoming more discriminatory as the tumours grow larger, with the exception of $\bar{N}$. As the tumours simulated in this study are unrealistically small given the computational constraints, this information gives us hope that in tumours of realistic size, these measures would be even more useful. This becomes particularly important as the statistics that we have calculated come from the ‘true trees’, that is, trees comprised of all mutational events. In reality, trees would be inferred from the imperfect information gleaned from biopsies, and also would suffer from our significant lack of ability to identify mutations with low penetrance in clinical samples.

4.3.4 Dependence of tree-based measures on mutation probability

As the tree measures depend heavily on the number of mutations within a given tumour, and therefore the number of branches within a given tree, we next ask how these measures behave when we vary mutation probability (λ) and symmetric division probability

simultaneously. To answer this, we perform 10 stochastic simulations for each combination of the symmetric division probabilities considered previously and 5 different values for λ varying over two orders of magnitude (0.001, 0.005, 0.01, 0.05, 0.1). We then use the previously described method to reconstruct the resulting phylogenies and calculate the measures previously discussed. In particular, we ask how the Sackin index, the B1 statistic and the normalized Sackin index perform over this range of λ to better understand the applicability of these measures in determining differences in symmetric division probability.

We plot the results of this parameter investigation in Figure 4.7. In each heat map, we plot the mean of the 10 simulations for each parameter combination with symmetric division probability varied along the horizontal axis and mutation probability along the vertical. The indices which are not normalized by branch number, namely the Sackin index and B1 statistic, increase monotonically with mutation probability and symmetric division probability in all cases. The Sackin index normalised by the PDA model, however, varies somewhat unexpectedly and has a global minimum at symmetric division probability of 1.0 and mutation probability 0.01. This measure is monotonic in symmetric division probability except at the highest mutation probability where it becomes somewhat more difficult to determine the differences. As earlier, in Section 4.3.1, the B1 statistic appears to be the most stable, and only breaks down slightly in its ability to distinguish between the families of simulations at the lowest mutation probability ($\lambda = 0.001$) and the middle range of symmetric division probability (symmetric division probabilities = 0.4 – 0.8), as can be seen in Figure 4.7. While this is a relatively small number of simulations, it provides some understanding of the ranges in which these measures are useful. To further investigate this, a more thorough investigation of the measures including statistics on their distributions in each area of parameter space could be enlightening, but we leave that for future work.

4.3.5 Changes in tree measures with spatial constraints

In Chapter 3, we found that the addition of spatial constraints strongly affected clone size and diversity measures. The changes we observed there, however, were usually simply to blunt the trends (and subsequently remove the ability to discern differences) that were much more clear in the non-spatial model. Here however, we see much more striking, qualitative and quantitative differences in some measures between the two models, specifically the B1 statistic and $\bar{N}$. In particular, the addition of spatial constraints induces a clearer trend in tree balance as measured by the B1 statistic, with the measure increasing monotonically with symmetric division probability for the spatial model, but not in the non-spatial model. The same change is true for the $\bar{N}$ statistic, but in reverse. While

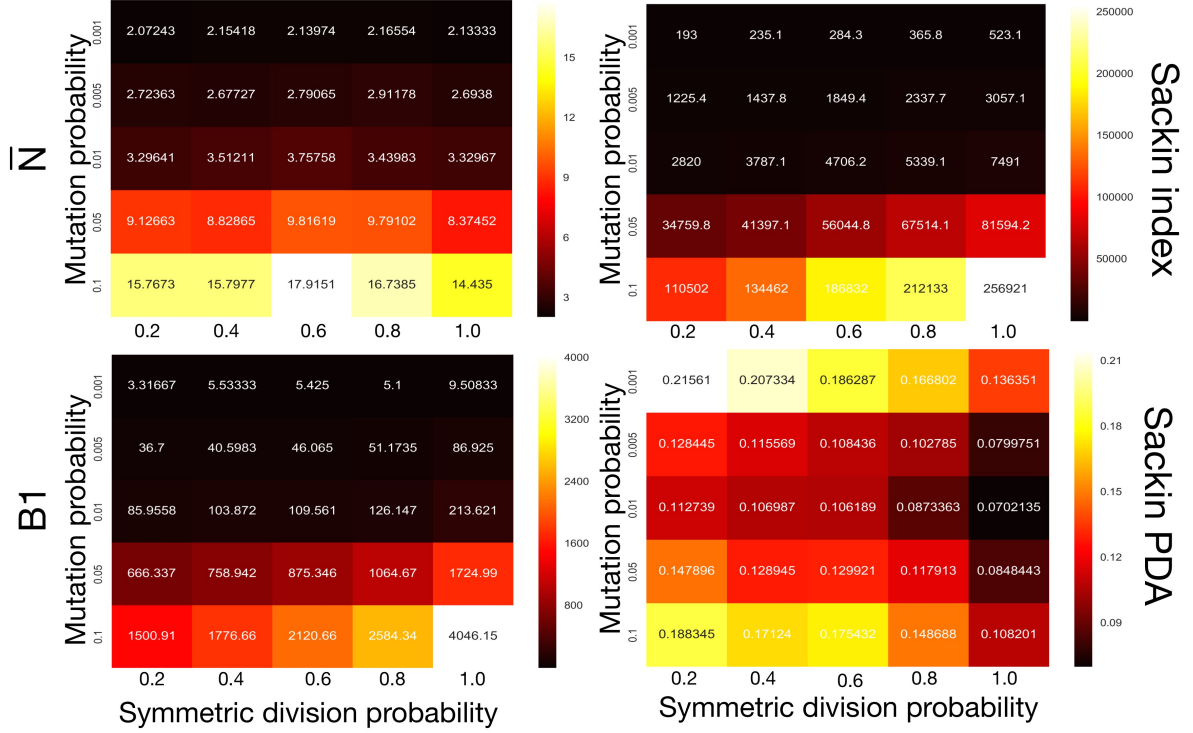


Figure 4.7: Comparing phylogenetic tree measures across symmetric division probability and mutation probability. We plot the average of each of four phylogenetic tree measures at the end of each of 10 simulations for a range of symmetric division probabilities and mutation probabilities. We vary mutational probability over two orders of magnitude ($0.1 - 0.001$), and simulate all tested symmetric division probabilities. Rank is maintained across symmetric division probabilities for each of the three four measures with which we could discriminate between symmetric division probabilities with changing mutation probability, allowing for differentiation between parameters. As before, the $\bar{N}$ statistic is not predictive. As expected, for the non-normalized indices, Sackin and B1, the measures change monotonically with both symmetric division and mutation probability. For the PDA normalized Sackin index, however, there is a global minimum for $\lambda = 0.01$ and $\alpha = 1$.

monotonic in the non-spatial model, this characteristic is lost when spatial constraints are added. These differences will be important to understand as we move closer to being able to implement these tests in the clinic as it appears that different measures should be considered for liquid tumours like leukemias (i.e. not spatially constrained) as compared to solid tumours (i.e. spatially constrained). We leave an explicit exploration of these changes for future work.

4.3.6 Utility of imperfect information in estimating tree measures

As a first test of the effect of this imperfect information on our ability to predict tree topology, we randomly sample a single simulation of our stochastic spatial model. For this example, we choose a simulation with symmetric division probability 1.0, $\beta = 4$, $\lambda = 0.01$

and domain size 500×500 . Using the biopsy method introduced in Chapter 3, we then aggregate each cell from an increasing number of (uniformly at random) non-overlapping biopsies of increasing size. Specifically, we compare the instances of 5 biopsies of radius 30 cells, 10 biopsies of radius 30 cells and 15 biopsies of radius 50 cells. In this case, we allow all mutations, regardless of their frequency in the samples, to be counted. To go from the unique mutational flag stored by the CA model to a ‘genome’, we developed an algorithm which takes as data the mutational life history of the simulation and the sampled cells, and provides a list of bitstrings representing the genomes of the cells in the list (see B.5 for pseudocode). Using the list of sampled ‘genomes’ for each condition, we then utilised ‘Phylogenetic Analysis Using Parsimony (*and other methods)’ (PAUP*) [146] algorithm to infer the phylogeny of our sampled tumour. In several of the cases, a number of equivalent trees were suggested by the algorithm, but for clarity we choose to represent only one in each situation for this current exposition.

As expected, the more cells sampled, the more mutations and relationships were found. This increase in tree complexity, approaching that of the ‘true’ phylogeny, also encodes changing (im)balance measures (Figure 4.8). At the level of detail we find here, however, discerning between simulations of different symmetric division probability is not possible in our simulations with only the imperfect information gleaned from biopsy. How the measures approach the measures of the ‘true’ phylogeny, we will leave to future work.

4.4 Discussion

In Chapter 3 we found that we were able to determine quite reliably the symmetric division probability of simulated tumours grown without spatial constraints by simply using the Shannon diversity index. Our analyses of tumour diversity, clone size distribution and shape were less able to determine this for spatially structured tumours, however. Further, we found that our ability to determine diversity from imperfect information, such as that gleaned from a biopsy, was significantly compromised, making its use appear to be insufficient using our current method. Given the importance of understanding differences in symmetric division probability at diagnosis and before and after therapy, we therefore endeavoured to derive a method that could be translated to the clinic to determine this parameter in solid tumours.

To address this shortcoming, we worked to leverage the growing interest in inference of tumour phylogenies from spatially distinct tumour biopsies, and the recent success of Leventhal [136] and others in teasing apart complex biological rules from phylogenetic information. To this end, we developed an algorithm to construct phylogenetic trees from

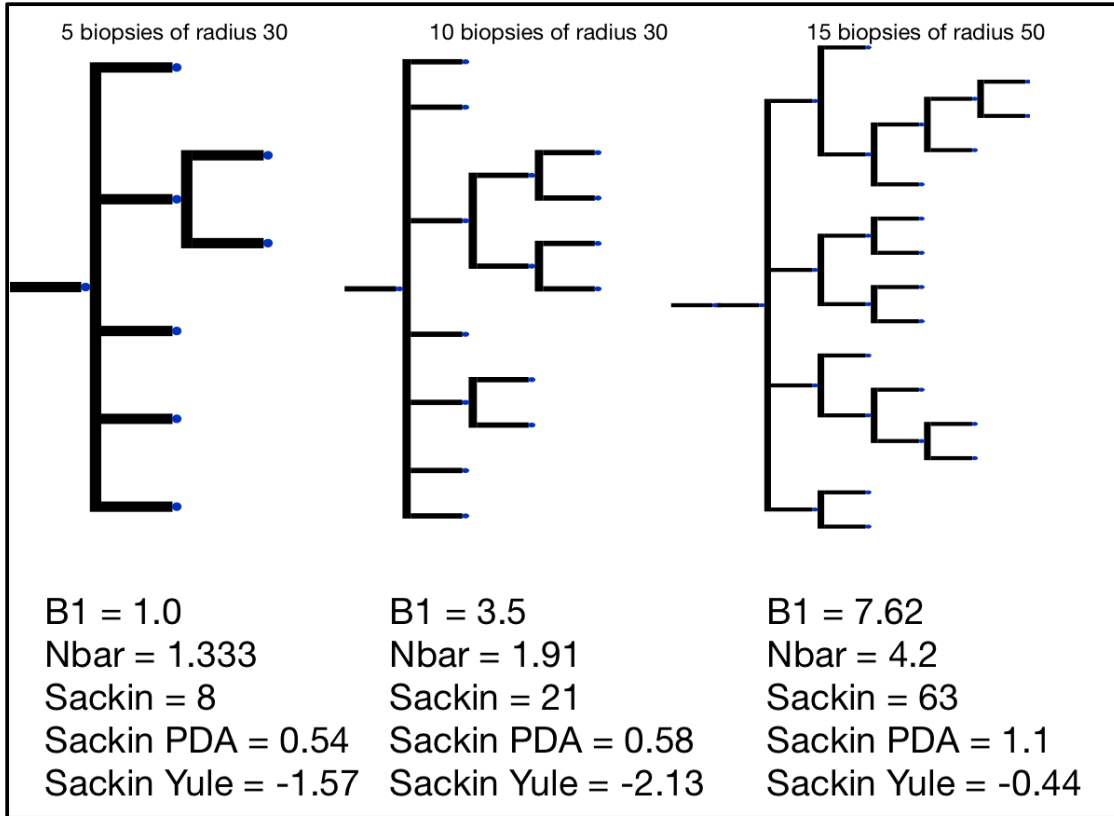


Figure 4.8: With clinically relevant sample numbers and sizes, inferred trees miss the majority of the complexity of the ‘true’ phylogeny. We plot three example trees inferred using PAUP* from an increasing number and size of biopsies from the same *in silico* tumour. Each sample is drawn from a single 500×500 simulation with $\alpha = 1$ and $\lambda = 0.01$.

the simulated tumours we analysed in Chapter 3. The resultant trees were then analysed and compared using a suite of statistical measures of tree (im)balance.

In particular, we compared the classical measures of tree topology – the Sackin index and the B1 statistic – as well as normalized versions of each across several parameters of our spatial and non-spatial models as well as through the process of tumour growth. Not surprisingly, we found that the Sackin index was able to discriminate between the families of simulations as it is directly correlated with branch number (in this case correlating with total number of mutations in the TICs, which also is increased with increasing symmetric division probability). Encouragingly, we also found that the normalised version of this metric was able to discriminate between the different symmetric division probabilities, suggesting a more meaningful (and measurable) topologic difference between the underlying phylogenetic trees resulting from these parameter changes (representing diverse biological traits).

4.4.1 Translating findings to the clinic

While the use of phylogenetic trees is increasing in translational oncology laboratories, there has yet to be a method found by which we can utilise the information clinically. In this chapter, we have generated a large dataset that includes the observed statistical measures of the ‘true’ phylogeny for tumours with a range of symmetric division probabilities. While we have shown that these measures differ significantly from one another, we have not yet provided a method by which we can use the metric of a given tree to directly predict the symmetric division probability of an unknown tumour.

However, the present work at least allows us to understand the rank order of symmetric division rate for two tumours given their measured indices. This could be particularly useful in certain clinical settings. For example, this could allow us to determine how a given therapy affects symmetric division probability by using our calculated measures over serial biopsies, and subsequent phylogenetic reconstruction.

Given the difficulties inherent in multiple biopsies over time in solid tumours, this information is most likely translatable in the setting of liquid tumours like leukemias or myeloma. In these settings, patients often receive multiple bone marrow biopsies through time, and as our ability to perform single cell sequencing (and subsequently to reconstruct phylogenies from many, single cell samples) improves, this will be more feasible.

4.4.2 Conclusions

With an aim towards a translatable method by which to infer the symmetric division probability in solid tumours, we have worked to develop a suite of measures on phylogenetic trees that correlate with symmetric division probability. We have found striking differences between the spatial and non-spatial models and, importantly, have found several measures which are able to discern differences in simulated tumours between symmetric division probabilities in the spatial model. These results are also robust to changes in tumour size, specifically maintaining their rank through growth of the tumour. Changes in mutation probability does affect the results to some degree, but rank is typically maintained.

While there is some overlap amongst the measures when more than one parameter is varied, with information on mutation probability and tumour size, relative symmetric division probability can be estimated. To this point, we have only considered measures of (im)balance, a crude measure of phylogenies which considers only the branching topology. With more information, such as evolutionary branch lengths [126, 164] which are linked to the ‘speed’ of a tumour’s molecular clock [47], some of these limitations could be obviated. Further, to this point, we have only considered neutral evolution. While most

tumour evolution is likely neutral [246], there is certainly evidence for non-neutrality in the form of driver and passenger mutations [155], which would drastically affect the resulting phylogenetic trees [95] – especially with intervening treatment regimens. How non-neutral evolution and treatment affect our measures, however, we leave for future work.

Chapter 5

Quantifying microenvironmental heterogeneity in a hybrid cellular automaton model of cancer in a vascularised domain

5.1 Introduction

In Chapters 3 and 4, we concentrated on heterogeneity at the level of cell proliferation, in the form of the cancer stem cell hypothesis (CSCH), and through genetic changes via neutral evolution. As discussed in Chapter 1, however, there is also significant heterogeneity possible beyond the cells themselves, in the tumour microenvironment. This heterogeneity, even without any differences between cells, can cause changes in tumour growth, invasion [82] and treatment response [56]. While there are myriad factors to consider including nutrient delivery, healthy immune cells, extracellular matrix, acid produced by tumour cells and more, we will consider only one here: oxygen. As my clinical focus is radiotherapy, I will focus in this part of the thesis on understanding the relationship between spatial heterogeneity in oxygen distribution, vascular patterning, cellular growth patterns and radiation response. In order to understand these relationships, we will first consider a simplified version of our cellular automaton (CA) model from the previous part of the thesis without proliferative heterogeneity or neutral evolution. We will extend the model, however, to include several complexities. Specifically, we will consider *en face* vessels, oxygen transport and uptake, and healthy surrounding tissue. Of note, most of the results and methods in this, and the following chapter, have been published in the manuscript entitled: ‘Spatial metrics of tumour vascular organisation predict radiation efficacy in a computational model’ published in January of 2016 in *PLoS Computational Biology* [212]. Further, all code and raw results are included in a .zip file within that publication.

5.1.1 Previous work

While many tumour models focus on avascular tumour growth, or ignore the tumour's need for oxygen, the idea of modelling tumour growth in vascularised domains is not a new one. Indeed, many methods have been used to address specific aspects of vascularised tumour growth and resultant microscopic oxygen dynamics. As monitoring these microscopic oxygen dynamics and levels in living systems (including patients) is extremely difficult, a number of efforts have been made to use mathematical modelling to better understand the dynamics. Continuum models include non-spatial approximations using ordinary differential equation (ODE) systems to study the effect of vessel maturation on tumour growth [19, 20], as well as spatially resolved descriptions using coupled partial differential equations (PDEs) to approximate relationships between tumour cells and vasculature [31, 138, 178, 232] to understand how a growing tumour evolves in space and time.

Alarcón et al. approached this problem by building a two-dimensional hybrid discrete-continuous CA model of cancer cell growth. The authors paid careful attention to the biophysics of blood flow in branching vessels in the plane of the growing tumour [4]. They found that the organisation of the vascular trees could strongly affect the shape of the growing tumour as well as the speed of its advance and effects of chemotherapy [7]. Macklin and colleagues utilised a detailed biophysical model of flow and hydrostatic pressure to understand tumour progression [145] and found that only vessels at the surface of a tumour were effective in delivering oxygen.

Powathil et al. focussed on the problem of measurement in real systems and used a coupled system of reaction-diffusion PDEs to understand what the optimal strategy would be for physically sampling a tumour [190] with oxygen probes. They utilised microscopy images from glioma xenograft specimens that were stained for the location of areas of hypoxia and blood vessels to initiate their model and then simulated measurements of the domain with oxygen probes. They report an optimal strategy for measurement, and also noted several characteristic oxygen profiles, each of which responded to simulated radiotherapy differently.

More recently, Grogan et al. have readdressed the complex issue of modelling tumour vasculature at multiple scales [97], and since the publication of the results of this thesis, extended their work to three dimensions with radiotherapy [98]. With an eye to standardising these types of models, and in support of the community, Grogan et al. have also released an open source software package for Chaste [162, 186] specifically designed for microvessel modelling [99].

5.1.2 Aim of work

Up to the point of publication of these results, however, there had yet to be concise, translatable, method of understanding cellular oxygen distributions from clinically available data. To address this, we will develop and parameterise an HCA model to explicitly study the relationship between vascular architecture and cell-oxygen distributions.

5.1.3 Structure of chapter

In this chapter, we will first introduce our modelling framework in Section 5.2 and discuss in detail the assumptions underlying each component of the hybrid CA model as well as our method of model implementation, numerical solution scheme, parameterisation and non-dimensionalisation. Then, in Section 5.3, we investigate the effects of varying vascular density and patterning in healthy (non-cancerous) tissue. We then proceed, in Section 5.4, to model tumour cell growth and invasion in vascularised domains and finally, in Section 5.5, make some initial observations of tumour growth and cellular oxygenation.

5.2 Hybrid cellular automaton model of tumour and normal tissue

To understand the effect of a heterogeneous microenvironment on tumour growth, we will utilise a model similar to the one we built in Chapter 3 but with several changes. As our question here concerns microenvironmental heterogeneity, we will first reduce the complexity of the earlier model by removing the cellular heterogeneity. We will then extend the model to consider the microenvironment by the inclusion of blood vessels as discrete sources of oxygen. The addition of vessels supplying oxygen also necessitates inclusion of a field of healthy cells surrounding our growing tumour to preclude supra-physical quantities of oxygen.

To this end, we will modify our spatial individual cell-based model into a hybrid CA [13, 17], to which we will refer henceforth as the HCA. In this framework, cells are described by individual objects whose states are updated over discrete time steps, and which occupy sites on a two-dimensional square lattice representing a slice of tissue. As we are now concerned with events linked to specific time and length scales, such as oxygen diffusion, we must further consider parameterisation. To this end, the size of the lattice spacing is chosen such that each automaton element is approximately the size of a single cell. A second lattice, identical in size and dimension, is created on which we approximate the freely diffusing, continuous oxygen concentration. The relatively coarse spatial scale is balanced by a finer time scale in our numerical scheme, as discussed in Section 5.2.3.

While the cell states and oxygen distribution are updated on separate lattices, each influences the other, the feedback between cells and the microenvironment being captured through a reaction-diffusion PDE governing oxygen transport and consumption. While the HCA model we aim to create in this chapter is designed to be generic, so that the most general conclusions can be drawn from it, we base parameter values on a concrete biological entity, the primary brain tumour glioblastoma.

5.2.1 Modelling the continuous oxygen field

The continuous component of our model describes the distribution and consumption of nutrients. While it is clear that many nutrients are of biological importance, in this model we focus on oxygen as it is central not only to cell metabolism, but also to the DNA-damaging effects of photon radiation therapy [100]. We consider *en face* blood vessels, which are each modelled as point sources of oxygen, occupying one lattice site, and are placed throughout the lattice at the start of a given simulation in a density and pattern appropriate for the question being studied. While in reality, vessels can be different sizes, they tend to be normally distributed around a mean size, and a variety of measures, including number of vessels, total vessel area and vascular fraction have been used interchangeably [80, 140] in previous studies. As we are interested in steady-state characteristics, in all simulations we neglect vascular remodelling and angiogenesis, processes which are triggered by hypoxia and mediated by hypoxia inducible factor-1 α (HIF-1 α) and vascular endothelial growth factor (VEGF), among other molecules, and have been well studied [13, 31, 145, 180, 185, 229]. Each vessel is assumed to carry an amount of oxygen equal to that carried in the arterial blood [129]. This oxygen is then assumed to diffuse into the surrounding tissue and be consumed by normal and tumour cells.

The spatiotemporal evolution of the oxygen field is described by the reaction-diffusion PDE

$$\frac{\partial c(\mathbf{x}, t)}{\partial t} = D_c \nabla^2 c(\mathbf{x}, t) - f_c(\mathbf{x}, t), \quad (5.1)$$

where $c(\mathbf{x}, t)$ is the concentration of oxygen at a given time $t \geq 0$ and position $\mathbf{x} \in (0, N\Delta)^2 \setminus V$, where Δ is the length of a typical cell, N is the number of lattice sites on a side of the $N \times N$ domain and $V = \{x_1, x_2, \dots, x_v\}$ denotes the set of points occupied by blood vessels, which are v in number. We define D_c to be the diffusion coefficient of oxygen, which for simplicity we assume to be constant (providing linear, isotropic diffusion), and the function $f_c(\mathbf{x}, t)$ denotes the cellular uptake of oxygen, whose form is given by

$$f_c(\mathbf{x}, t) = \begin{cases} \mu_i r(c(\mathbf{x}, t)) & \text{if there is a cell of type } i \text{ at } \mathbf{x} \text{ at time } t, \\ 0 & \text{otherwise,} \end{cases} \quad (5.2)$$

where $i \in \{H, T\}$ and the labels H and T are used to refer to healthy and tumour cells, respectively. Here μ_i is defined as the cell type-specific oxygen consumption rate constant. This is taken to modulate $r(c)$, the oxygen-dependent consumption rate, which we assume to have Michaelis-Menten form

$$r(c) = r_c \left(\frac{c}{c + K_m} \right), \quad (5.3)$$

where r_c and K_m denote the maximal uptake rate and effective Michaelis-Menten constant, respectively. We supplement equation (5.1) with the following initial and boundary conditions. We begin with a spatially uniform oxygen distribution $c(\mathbf{x}, 0) = c_0$, a positive constant, with the entire domain, $(0, N\Delta)^2 \setminus V$, occupied by normal cells. In the case of a cancer simulation, we occupy the central lattice site with a single tumour cell. Vessels are placed throughout the domain at a prescribed density and pattern, depending on the situation being simulated. To approximate the constant value of the oxygen concentration in the arterial blood we impose $c(\mathbf{x}, t) = c_{\max}$ at each point $x \in V$, where there is a vessel (for a more detailed approach to this problem, see Patel et al. [184]). We further impose no-flux boundary conditions at the edge of the domain,

$$\mathbf{n} \cdot \nabla c = 0, \quad (5.4)$$

where $\mathbf{n}$ is the unit normal outward vector to the boundary.

Vascular dynamics

While vascular patency is the norm in healthy tissue, as tumours grow the vasculature can experience significant spatio-temporal changes. There continues to be significant theoretical work to describe angiogenesis, the process of new vessel formation [13, 180, 229], as well as the biophysical processes governing vascular occlusion [21, 112]. In this chapter, we endeavour to understand the effect of heterogeneous oxygen concentrations, similar to that studied by Alarcón and colleagues [4] on tumour growth and progression. Unlike Alarcón, who studied a regular vascular architecture with changing flow dynamics, we will prescribe spatially varying (regular and irregular) vascular architectures and study the resultant spatially heterogeneous oxygen concentrations and its effect on cellular growth. Therefore, we consider only the simplest scenario, in which there is no feedback from the environment to the vessels.

5.2.2 Individual cell behaviour

Cell fate decisions in our model are determined by a number of microenvironmental and cell state-specific thresholds and values. The order in which cells are chosen is computed in a random fashion so as to avoid any order bias. To determine the position of cells within our domain, we tessellate the continuous domain on which the oxygen PDE is defined into squares of size $\Delta \times \Delta$ to arrive at a regular lattice occupied by both cells and vessels. All tumour and healthy cells are assumed to be the same size and shape (a single lattice site), and divide to produce two identical daughters. The probabilities and thresholds for these stochastic cell fate decisions are listed in Table 5.1 and Figure 5.1, respectively, and will be discussed individually in the coming sections.

Oxygen consumption

With the advent of single-cell techniques for measuring oxygen consumption rates, quantitative data concerning the differential oxygen consumption rates of different types of tumour cells are now available. It is widely accepted that, in general, tumour cells are more metabolically active than healthy cells because of the Warburg shift [83], and we thus use an approximate modification of double normal consumption, that is $\mu_T = 2\mu_H$ [110], where $\mu_H = 1$.

Proliferation

Cell proliferation is a process governed by a complicated set of subcellular processes, and has been modelled previously in great detail [5, 6]. The cell cycle has many levels of complexity including checkpoints, specific temporal sequences and many biophysical processes that alter the cell shape and response to microenvironmental cues and stresses at different times in the cycle [236]. As we are interested in a phenomenon occurring at the much larger, tissue scale, we will encapsulate this complexity into a simple rule set and threshold values on microenvironmental (oxygen) conditions and neighbourhood occupation.

For cells to progress through the cell cycle and divide, we require their local oxygen concentration to exceed a fixed threshold value c_p , and that the neighbourhood (in this case utilizing a Moore neighbourhood, as shown in Figure 5.3, left), contains at least one empty space. If these constraints are met, then we allow the cell to divide with type-specific probability p_j (where $j \in \{H, T\}$) at each time step, and place the resultant daughter in a randomly chosen empty neighbouring space. If no empty neighbouring space exists, then the cell becomes quiescent (see below).

Cell death

The main mechanism of tumour cell death we consider in this chapter will be through starvation from lack of oxygen. Healthy cells can also die through lack of oxygen or if they are selected for replacement by dividing tumour cells, a situation which would normally be ascribed to a tumour cell's greater fitness, allowing it to out-compete healthy cells for space or nutrients locally, or possibly secondary to an acid-mediated event [83, 184]. Death from other mechanisms, such as radiation, will be incorporated in Chapter 6. Cells in extreme hypoxic conditions are often found to undergo autophagy (directly translated as 'self-eating'), a state in which they become resistant to nutrient starvation [248], and cells are known to die over different time scales and by different mechanisms (apoptosis versus necrosis) depending on the magnitude and duration of the hypoxic insult. While these differences have been shown to affect tumour growth [34], as this is not the main aim of this model, we will simplify this scenario so as to avoid confounding our results. We will therefore assign a rate, p_d , for cell death at each cellular automaton update defined in Section 5.2.3, when under extreme hypoxia (i.e. $c < c_{ap}$ where $c_{ap} \in (0, c_{max}]$ is a fixed, model-specific positive constant). We will leave exploration of a more detailed model with different types of cell death for future work.

Quiescence

When cells sense that there is not enough oxygen to divide, or experience contact inhibition, they undergo a state of quiescence during which there is no division. We model this process by considering two factors: the local oxygen concentration and contact inhibition. We consider an oxygen threshold c_p , where $c_p \in [c_{ap}, c_{max}]$ is a fixed, model-specific positive constant, below which cellular division is not possible. We further consider contact inhibition which also changes a cell to a quiescent state if there are no empty neighbouring lattice sites (again using a Moore neighbourhood). In the case of a tumour cell, we require that at least one neighbouring space be empty or inhabited by a healthy cell, a situation in which we assume that the tumour cell will replace the healthy cell if chosen for division (see Section 5.2.2). When a cell senses that sufficient oxygen and that at least one neighbouring site has become available, its state of quiescence is reversed.

Normal tissue approximation

The invasion of tumour cells into healthy tissue has been widely studied [16, 81, 82, 88, 184]. While this process is not the focus of our study, we nevertheless must consider the healthy tissue surrounding our growing tumour, as it plays an important role in

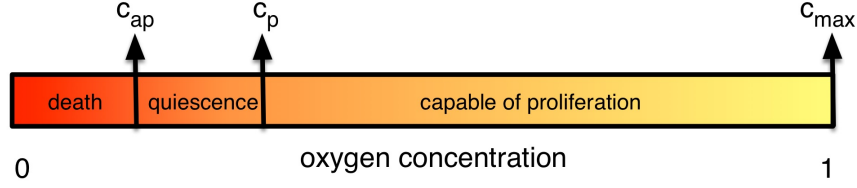


Figure 5.1: Summary of oxygen-based cell fate thresholds. At each cellular automaton update, each cell in the domain undergoes, in a random fashion, a series of fate decisions based on the local oxygen concentration. When $0 \leq c < c_{ap}$, cells die at rate p_d . When $c_{ap} \leq c \leq c_p$, cells are quiescent. When $c_p < c \leq c_{max}$, cells are capable of proliferation at rate p_j , if all other constraints are met.

modulating the local oxygen concentration. Previous investigations have modelled tumour spheroid growth [28, 41, 96, 234], or assumed a constant influx of oxygen from the boundary of the tissue [88, 223], but as we endeavour to understand how heterogeneous vascularisation affects the growth of a tumour, we must also incorporate normal tissue effects [4]. To do this, we assume that the entire domain is initially filled with healthy cells, which can divide if they have sufficient space, which consume oxygen at a rate of $\mu_{HR}(c)$ and are subject to hypoxic cell death as described in Section 5.2.2. In addition, normal cells can die through competition with cancer cells (see Section 5.2.2).

Cell competition

The ability to invade into normal tissue is one of the ‘Hallmarks of Cancer’ [102]. Several mechanisms for this have been proposed and modelled, including degradation of extracellular matrix by secreted proteases [12, 16] and mediated by excreted lactic acid [81, 82]. As this specific interaction is not our focus, we make the simplifying assumption that healthy cells do not spatially inhibit tumour cell proliferation, and that a tumour cell can freely proliferate and place a daughter cell onto a lattice location where a healthy cell is located, causing the healthy cell to die, either through an acid-mediated or simple fitness based competition mechanism [172]. Healthy cells, however, are spatially inhibited by tumour cells and other healthy cells.

5.2.3 Cellular automaton rules and updates

Now that we have described each of the constituent parts of the model, we will describe how they are coupled to form the full HCA. We consider a two-dimensional lattice of size $N \times N$, with each lattice element identified by coordinates $(i\Delta, j\Delta)$ where $i, j \in \{1, 2, \dots, N\}$ and Δ is as defined in Section 5.2.1.

As schematised in Figure 5.2, at each automaton time step every cell in the domain is subject to a series of decisions based on the current automaton state and local oxygen

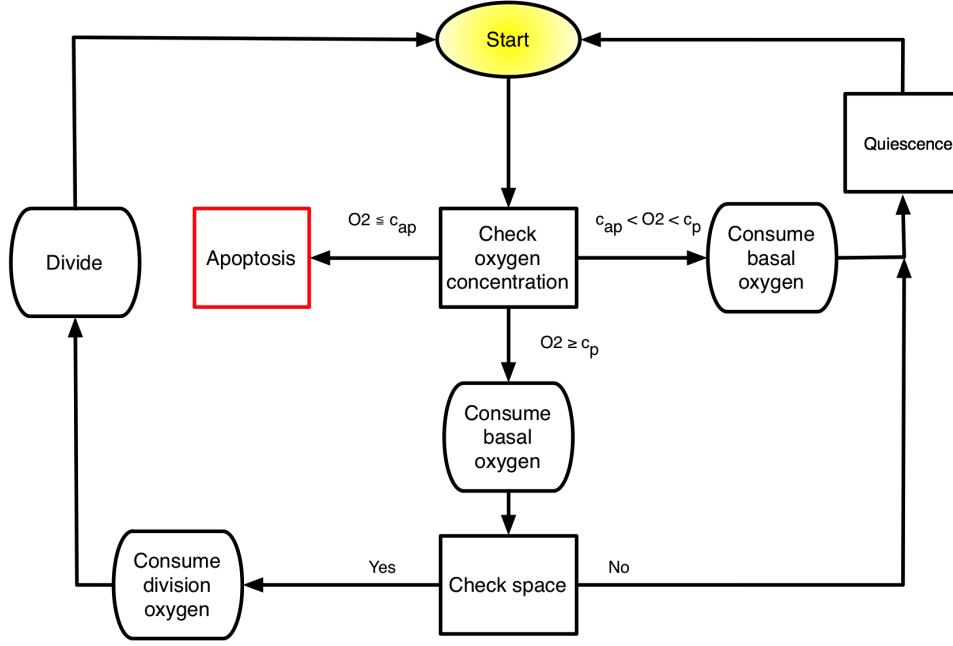


Figure 5.2: Schematic of the discrete time updating algorithm for the HCA. At each cellular automaton update, each cell in the domain undergoes a series of fate decisions based on intrinsic cell parameters and microenvironmental cues.

concentration, $c(\mathbf{x}, t)$. To minimise order effects, cells are chosen uniformly at random, without replacement, to undergo cell fate determination. The specific checks that each cell undergoes, regardless of its phenotype and location, are:

1. compare $c(\mathbf{x}, t)$ to c_{ap} and c_p ;
2. consume oxygen at rate $\mu_i r_c$, die, or remain quiescent;
3. check number of free neighbouring lattice sites;
4. determine proliferative behaviour: if proliferative constraints are met (space and oxygen), then consume additional oxygen for proliferation [77] and place daughter cell in randomly chosen empty neighbouring lattice site, otherwise become quiescent.

Parameters

As in any model of a complex system, ours incorporates a number of parameters to tailor it to the system in question: a dynamically growing tumour dependent on a spatially heterogeneous oxygen supply. Driven by Einstein's mandate, often called Einstein's razor,

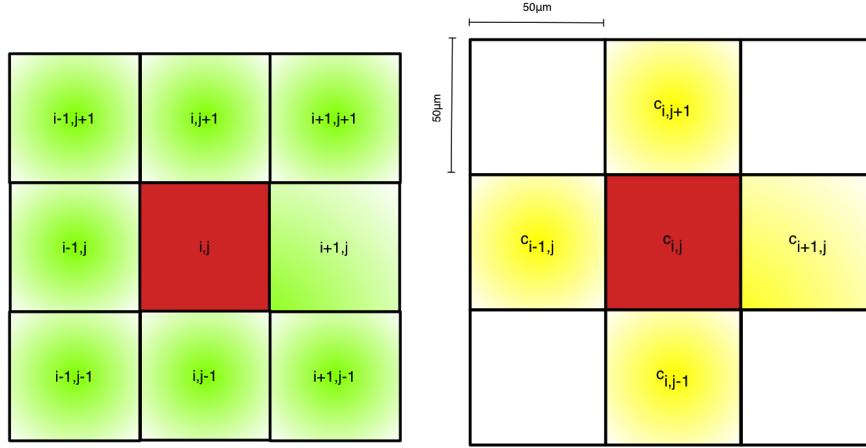


Figure 5.3: Definition of lattice position neighbourhood for cell division (left) and oxygen diffusion (right). (Left) Cells ascertain whether they have space for division by considering their Moore neighbourhood, in green. (Right) The oxygen concentration at position $(i\Delta, j\Delta)$ is calculated using the explicit scheme described in Section 5.2.3, considering the von Neumann neighbourhood indicated in yellow.

to make our model as simple as possible, but not simpler¹, we seek the smallest number of model parameters possible that will allow us to study the question at hand; which here is: how does a heterogeneous microenvironment (oxygen distribution in our case) affect tumour growth and oxygenation? We also recall that, as the values of these parameters are estimated from measurements from many different systems, under disparate conditions, they serve merely as guidelines to help us constrain the values so as to gain a qualitative understanding of the system in question, rather than to provide accurate quantitative predictions.

Cell-cell interaction

We assume that, in addition to having a sufficient level of oxygen for division, cells must also have space available to place their progeny (Figure 5.2). This commonly made simplification is intended to capture the essence of each cell's need for available space. We specifically require, for tumour cells, that in the lattice sites surrounding a cell, there be at least one lattice site that is not occupied by another tumour cell for division to occur. For healthy cells, we require that at least one lattice site be empty for healthy cell division to occur (see Figure 5.3, left).

¹This quote is actually a distillation of his real words: “It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.”

Cell-microenvironment interaction

As we endeavour to understand the effects of a heterogeneous oxygen distribution on healthy tissues and tumours, we must consider a number of parameters that govern a cell's behaviour in response to these environmental cues. Specifically, we consider the hypoxia threshold, c_{ap} , below which cells will die, and be removed from the system. As cell fate decisions are made on a much slower time scale than oxygen diffusion, there is a potential for update bias which would entail cell fate decisions made long after conditions for decisions were met. To reduce this, we will ascribe a probability per automaton update, p_{d} , to this death and removal, and update the HCA more frequently than the cell cycle time, which we will discuss later in this section.

Non-dimensionalisation and parameter estimation

In order to place the model in an appropriate spatial and temporal scale, we non-dimensionalise the system. We rescale time by a typical cancer cell cycle time, $\tau = 16\text{h}$ [36], and lengths by a typical diameter $\Delta = 50\mu\text{m}$ for a glioma cell [46]. We choose to rescale the oxygen concentration to reflect the average oxygen partial pressure in the arterial blood, estimated at about 80mmHg (or equivalently $5.14 \times 10^{-13} \text{ mol cell}^{-1}$) [37], so that the non-dimensional oxygen concentration at a vessel takes the value 1. This value agrees well with an estimate of background tissue oxygen concentration of $c_0 = 1.7 \times 10^{-8} \text{ mol cm}^{-2}$ ($4.25 \times 10^{-13} \text{ mol cell}^{-1}$) taken from Anderson and colleagues [12], which would correspond to a tissue partial pressure of $\approx 66\text{mmHg}$.

As each of these parameters has been estimated from different sources, we will fine-tune the basal oxygen consumption for our specific case using a well-studied tumour spheroid example. We utilize the observation that the diffusion distance of oxygen to support cancer cells is approximately 10 cell diameters [78] and the information from the literature concerning the ratio of cancer to normal oxygen consumption (see Section 5.2.2). To estimate the baseline oxygen consumption rate then, we begin with the value $r_c = 2.3 \times 10^{-16} \text{ mol cell}^{-1} \text{ s}^{-1}$ taken from an *in vitro* study of tumour spheroid growth [77] and then perform a virtual tumour spheroid assay to fine-tune the value for our model system (see procedure in Appendix C.2 and Figure C.1).

Introducing the non-dimensional variables $\tilde{\mathbf{x}} = \mathbf{x}/\Delta$, $\tilde{t} = t/\tau$ and $\tilde{c} = c/c_{\text{max}}$, we define the non-dimensional parameters

$$\tilde{D}_c = \frac{D_c \tau}{\Delta^2}, \quad \tilde{r}_c = \frac{\tau r_c}{c_{\text{max}}}, \quad \tilde{c}_{\text{ap}} = \frac{c_{\text{ap}}}{c_{\text{max}}}, \quad \tilde{K}_m = \frac{K_m}{c_{\text{max}}}. \quad (5.5)$$

For notational convenience, we henceforth refer to the non-dimensional parameters only and drop the tildes for the aforementioned parameters. See Table 5.1 for a full list of dimensional parameter estimates and corresponding non-dimensional values.

Table 5.1: List of dimensional model parameters and their estimates.

Parameter	Meaning	Value (original units)	Value (non-dimensionalised)	Ref
D_c	Oxygen diffusion coefficient	$1.0 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$	$2.304 \times 10^4 \text{ cell cycle}^{-1}$	[174]
r_c	Maximal oxygen consumption rate	$4.6 \times 10^{-16} \text{ mol cell}^{-1} \text{ s}^{-1}$	$51.54 \text{ cell cycle}^{-1}$	[77]
r_p	Proliferative oxygen consumption	$5 \times r_c$		[77]
Δ	Average cell diameter	$50 \text{ } \mu\text{m}$	1 cell diameter	[46]
Δt	Oxygen update time step	0.25 s	1 cell cycle / 230,400	model specific
τ	Average cell cycle time	16 h	1 cell cycle	[36]
c_{ap}	Hypoxic threshold	$5.14 \times 10^{-14} \text{ mol}^{-1} \text{ cell}^{-1}$	0.1	[37]
K_m	Effective Michaelis-Menten constant	$5.14 \times 10^{-15} \text{ mol}^{-1} \text{ cell}^{-1}$	0.01	[157]
c_{max}	Arterial oxygen concentration	$5.14 \times 10^{-13} \text{ mol}^{-1} \text{ cell}^{-1}$	1	[39]
μ_H	Healthy tissue metabolic constant	1		
$\mu_{\text{cancer}}/\mu_H$	Cancer metabolic ratio	2		[110]
p_d	Probability of death when $c < c_{\text{ap}}$	$0.25 \text{ cell cycle}^{-1}$		model specific

Numerical solution

In order to solve equation (5.1) numerically, we discretize space and time by letting $t_k = k\Delta t$, where $k \in \mathbb{N}$, and $x_i = i\Delta$ and $y_j = j\Delta$ where $i, j \in \{1, \dots, N\}$ encode a square lattice of size $N \times N$. We approximate the oxygen concentration at time t_k and position x_i, y_j by $c_{i,j}^k \approx c(x_i, y_j, t_k)$. We use a central difference approximation for the Laplacian and thus approximate equation (5.1) by

$$\frac{c_{i,j}^{k+1} - c_{i,j}^k}{\Delta t} = \frac{D_C}{\Delta x^2} \left[c_{i+1,j}^k + c_{i-1,j}^k + c_{i,j+1}^k + c_{i,j-1}^k - 4c_{i,j}^k \right] - (f_c)_{i,j}^k, \quad (5.6)$$

where $(f_c)_{i,j}^k$ is the cell-specific oxygen consumption rate defined in equation (5.2), $i, j = 2, \dots, N-1$ and (i, j) is not a vessel. We may rearrange equation (5.6) to obtain a solution for $c_{i,j}^{k+1}$, yielding

$$c_{i,j}^{k+1} = c_{i,j}^k \left(1 - 4 \frac{D_C \Delta t}{\Delta x^2} \right) + \frac{D_C \Delta t}{\Delta x^2} \left(c_{i+1,j}^k + c_{i-1,j}^k + c_{i,j+1}^k + c_{i,j-1}^k \right) - \Delta t (f_c)_{i,j}^k, \quad (5.7)$$

where $i, j = 2, \dots, N-1$ and (i, j) is not a vessel.

At each time step, the oxygen concentration in a given lattice site is updated using the values of the surrounding cells in a von Neumann neighbourhood (see Figure 5.3(b)). To impose the zero-flux boundary conditions we modify equation (5.7) in the cases where $i, j \in \{1, N\}$. For example, to calculate the oxygen concentration experienced by a cell at the left-hand boundary (i.e. $i = 1$) we discretise the no-flux boundary condition to obtain $c_{0,j}^k = c_{2,j}^k$, which we substitute into equation (5.7). Note that lattice sites which are occupied by vessels have fixed oxygen values (1 in the non-dimensional system), so we do not update them.

As a check of our numerical scheme, which is based on the forward Euler method, we compare the results of a test case of our diffusion solver against the analytic solution. In Figure 5.4 we plot the results of this analytic solution (for details of the derivation, see Appendix C.1) and the simulation results. To obtain a one-dimensional approximation from the HCA model we begin with a large (100×100) domain without cells with initial oxygen $c = 0$ at all points. We initialize both the right and left edges of the domain with adjacent vessels to approximate vessels running in-plane and allow the oxygen concentration to diffuse using our numerical scheme. We then plot the oxygen concentration profile through the centre of the domain at single oxygen update intervals. The analytic solution of equation (C.5) is compared to the numerical solution of equation (5.7) at each lattice point for time steps $t = 1, 10, 100, 500$ and $1,000$. We find excellent agreement between the two solutions.

Time scales and updates

The difference in time scales between the diffusion of oxygen and the proliferation of cells is managed by updating the continuous part of the model many times per cellular time step. This can become computationally expensive in this explicit forward Euler scheme, and therefore, we seek to minimize this number. However, for stability of the numerical

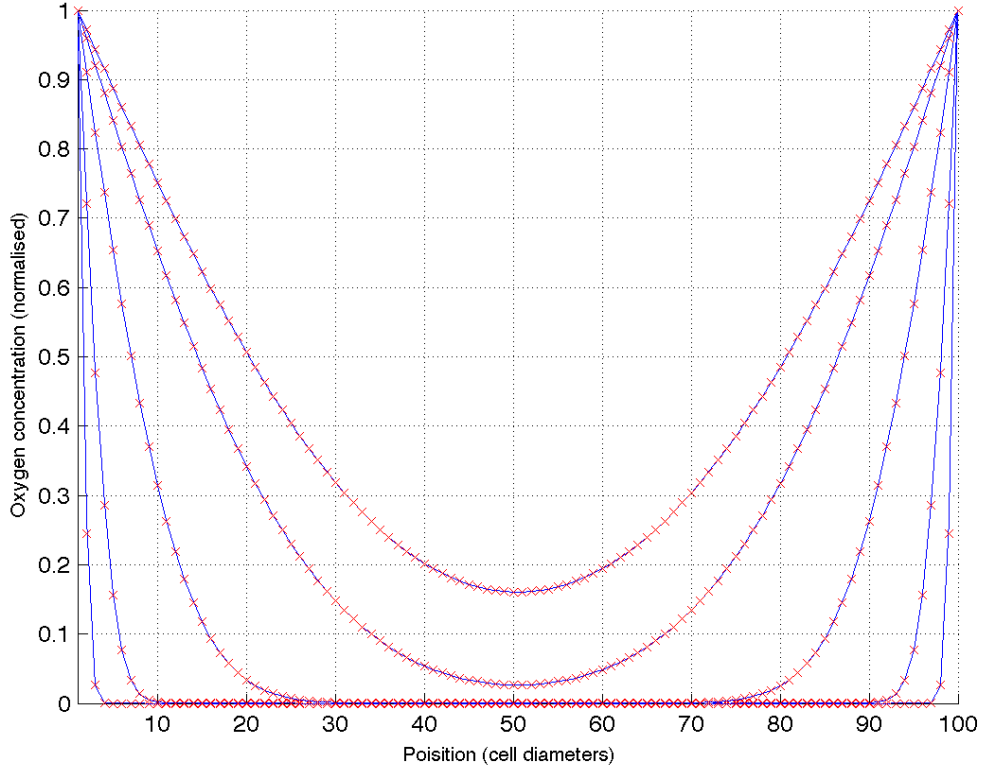


Figure 5.4: Oxygen concentration plotted as a function of time and space in the computational model in comparison to the analytic solution. The oxygen concentration is calculated using the explicit scheme described in Section 5.2.3 with initial oxygen concentration 0 everywhere in the domain, vessels running along the entire left and right hand border (Dirichlet boundary conditions) plotted over 1,000 oxygen update steps. The analytic solutions for $t = 1, 10, 100, 500$ and $1,000$ (blue lines) are compared to the results of the computational model at the same time (red crosses) with excellent agreement. For the analytic solution, we use $n = 10^6$, boundary conditions $c(0, t) = c(L, t) = 1$, initial condition $c(x, t_0) = 0$ and parameters from Table 5.1.

scheme, we require $\Delta t D_c / \Delta x^2 < 0.25$ [163]. We therefore choose $\Delta t D_c / \Delta x^2 = 0.1$. In dimensional parameters, we then calculate

$$\Delta t = \frac{0.1(50 \times 10^{-4} \text{ cm})^2}{1 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}} = 0.25 \text{ s}, \quad (5.8)$$

or equivalently in non-dimensional parameters

$$\Delta t = \frac{0.1(1 \text{ cell diameter})^2}{23,040 \text{ cell cycle}^{-1}} = \frac{1 \text{ cell cycle}}{230,400}. \quad (5.9)$$

While we assume the average cell cycle time to be $\tau = 16$ hours, it is well known that cells in tissues are not synchronized, and also that cell fate decisions such as apoptosis are made on shorter time scales. To model this heterogeneity in cell cycle time and to more accurately match the finer time scale associated with cell death due to microenvironmental

cues [204], we choose to update the cellular portion of our model 100 times per cell cycle and refer to this as the cellular automaton time step, (of duration $\tau/100$) and scale the rates for cell behaviour accordingly (reduced by a factor of 100) so as not to affect the time scale of division, thereby reducing the oxygen calculations to 2,304 updates per cellular automaton time step. This choice of time step is small enough to satisfy the criterion for numerical stability in our forward Euler method, and yields first-order convergence.

5.3 Understanding normal tissue behaviour

5.3.1 Varying vascular density with regular patterning

We begin with the default case of a normal tissue in our model, and find parameter values for which the model recapitulates a ‘normal’ homeostatic scenario. As such, we first model a domain seeded with only healthy cells, and a varying density of vessels with regular spacing, as we would expect from a healthy tissue. Estimates in the literature for ‘normal’ vascular density span several orders of magnitude [80, 134, 242], so we choose to calculate the actual physiological vascular density of this model by exploring regular vessel spacing and healthy cells only.

We choose a regular distribution of vessels such that the distance to each of the nearest vessels is roughly equivalent. To maintain vessel organisational geometry, we choose to seed the same number of vessels (40) in each simulation, and vary the distance between them. Therefore, for each new vessel spacing, the domain size will change to accommodate the full complement of 40 vessels with spacing being studied. In each of these cases then, the pattern of the vascular distribution will remain constant, but the vascular density (number of vessels/domain size) will change. For ease of visualisation, we have chosen to plot the domains as equivalent sizes (see Figure 5.5), and report the changing density.

We find, as expected, that the closer the vascular packing in the domain, the higher the proportion of cells that can be supported and the higher the mean oxygen concentration (Figure 5.5). Further, we find a critical value for vessel density below which cells begin to die, and therefore the proportion of the domain inhabited by living cells (termed carrying capacity henceforth) drops below unity. This value is between 2.7×10^{-3} vessels per area and 3.1×10^{-3} vessels per area, which is well within an order of magnitude of the vascular area to tumour area ratio reported by Zhang et al. [254]. However, this number strongly depends on the diffusion coefficient, D_C , and the maximal oxygen consumption rate, r_c (see section 5.4.1 for a more thorough treatment of this).

In addition to the carrying capacity, we can also investigate the distribution of cell-oxygen concentration. This distribution provides a non-spatial summary metric of an otherwise spatial entity: we can understand how many (or what proportion) of cells

within the domain exist at certain oxygen concentrations, a measure which will become important later when we consider the effect of radiation. In Figure 5.5 we note that, while the mean oxygen concentration and mean cellular oxygen concentration (mean of the cell-oxygen distribution) vary considerably (from a non-dimensional value of 0.16 to 0.45), the second and third moments of the distribution vary little, with the standard deviation staying between 0.08 – 0.11 and the skewness between 2.8 – 3.3. That is, the distribution of cellular oxygen concentration maintains its variance about its means as well as its level of symmetry while the vascular patterning is regular. As we will be interested in the effect of radiation therapy (which itself is dependent in a non-linear fashion on oxygen concentration) on members of these populations, these measures will become of use, which we explore in Chapter 6. We explore whether or not these observations about cell-oxygen distribution moments hold for irregular patterning in the next section.

5.3.2 Varying vascular density with irregular patterning

Now that we understand how healthy tissue is maintained in a domain with regular vasculature, we will explore how the oxygen concentration that the population of healthy cells supported is affected by changes in both density and patterning of vessels in irregularly vascularised domains. While this is a situation that would not likely be seen in the healthy state, understanding the behaviour of the model in this simple situation is valuable before progressing to more complex and biologically plausible states. To this end, we simulate the resulting healthy tissue growth and maintenance in a variety of vascular patterns of the same density (Figure 5.6). We plot the final, stable distribution of healthy cells for a random placement of vessels for a given density ($\Theta = 0.001$, top; $\Theta = 0.0025$, middle; and $\Theta = 0.004$, bottom) in the left and middle columns. In the right column, we plot the average distribution of cells at specific oxygen concentrations over ten simulations with different vascular patterns.

As expected, we see an increase in carrying capacity with increasing vascular density, similar to the regularly vascularised scenario, with mean cellularity (cells/domain size) increasing from 0.3445 to 0.9424 with standard deviations from 0.0208 to 0.0312. Over ten simulations in the irregularly vascularised cases, however, we find striking differences in the distribution of cell-specific oxygen concentration (Figure 5.6, right). While in the regular case (Figure 5.5) we find conserved cell-oxygen distribution shapes (second and third moments), in the irregularly vascularised case we find two changes. First, even in the highest vessel density case there is a large peak at c_{ap} and second, the distribution becomes more and more skewed (to the left) as the vessel density decreases (skewness ranging from 0.2878 to 1.895). We explore the impact of these differences on tumour cells in Section 5.4, and further investigate this effect on radiobiologic efficacy in Chapter 6.

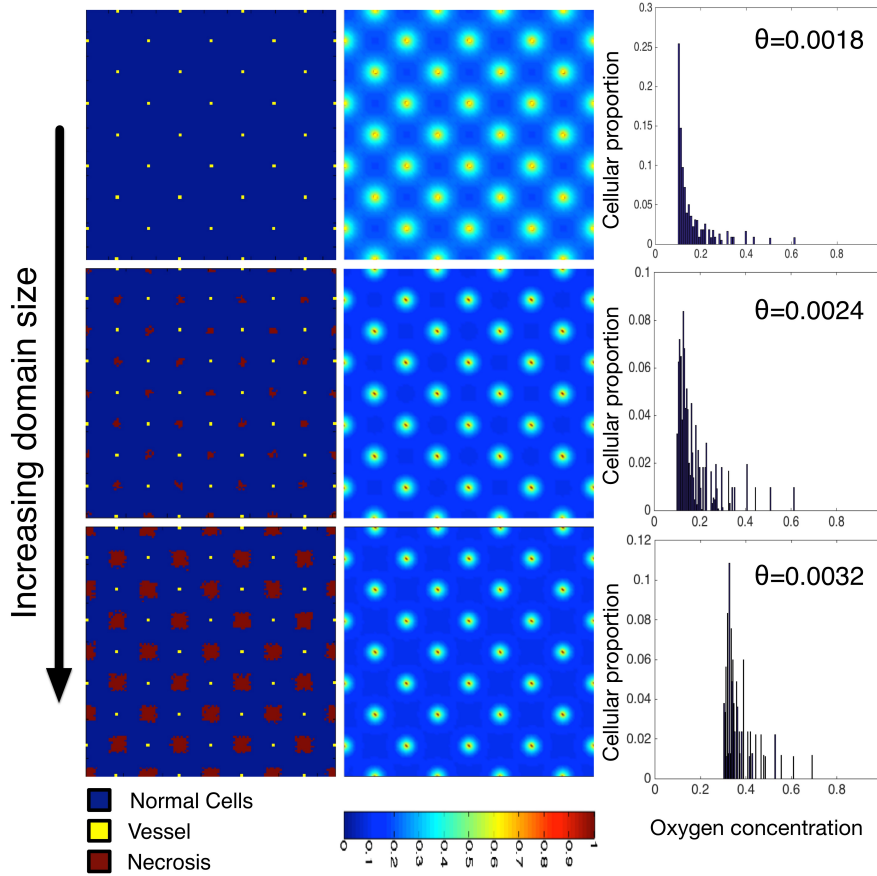


Figure 5.5: Varying vascular density with regular spacing affects the carrying capacity and cell-oxygen distribution in simulated normal tissue. Healthy tissue at dynamic equilibrium is plotted as we vary the vessel density (decreasing density from top to bottom), with cellular distributions (left), associated spatial oxygen concentration (middle) and non-spatial distribution of cells versus oxygen concentration (right). We denote the vessel density (vessels/domain size), Θ , inset. The mean of the cellular-oxygen distribution decreases with vessel density (0.26, 0.18 and 0.17, top to bottom) while the standard deviation and skewness stay approximately constant (std = 0.09, 0.1 and 0.1, skewness = 3.18, 3.23 and 3.32). Domains are visualised as the same size, with the same vessel pattern, for ease of comparison, but increase in size from top to bottom (domains are $N \times N$ where from top to bottom $N = 114, 122, 130$).

5.4 Vascularised tumour growth and invasion

5.4.1 Tumour growth in a regularly vascularised domain

As a tumour initially invades into healthy tissue, the vasculature that it would encounter would be that of the healthy tissue. While we are ignoring the effects of cell crowding and vascular deformation, it is of value to understand how the growth rate and the pattern of tumour spread would change given differing vascular spacing. To this end, we seed a series of regularly patterned vascularised domains with increasing regular vessel spacing, as in Section 5.3.1, with a single cancer cell and measure the tumour growth rate and pattern. Note that, to ensure initial tumour growth, we have placed an extra vessel at

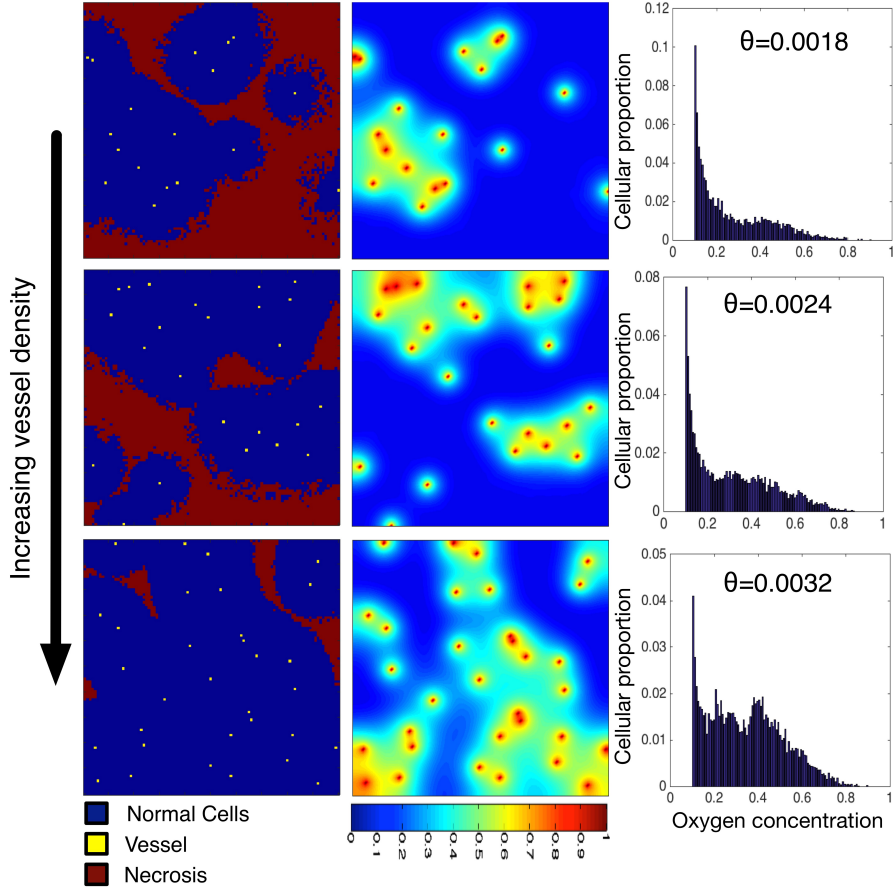


Figure 5.6: Varying vascular density with irregular patterning affects the carrying capacity of normal tissue and the cell-oxygen distribution. Results of normal tissue growth after dynamic equilibrium is reached as we increase the number of randomly seeded vessels from 10 (top) to 25 (middle) to 40 (bottom) on a fixed domain (100×100). Representative examples of a cellular distribution (left) and oxygen concentration (centre) are shown with the average (over ten stochastic realisations) cell-oxygen distribution (right). Unlike the regular vessel examples, the mean of the cellular-oxygen distribution is no longer monotonic with vessel density (0.37, 0.28 and 0.41, top to bottom) and the standard deviation and skewness now vary (std = 0.08, 0.17 and 0.19, skewness = 3.18, 1.19 and 0.28).

the centre of the computational domain with the initial cancer cell as in certain regular spacing patterns, the centre would be an area of necrosis and the simulation would end with no cancer cells.

We plot the results of these simulations in Figure 5.7 along with the individual simulation growth curves. We find four qualitatively different patterns/rates of growth for different vascular densities. First, when there is high vessel density (box 1, green trace, Figure 5.7), the tumour undergoes rapid growth. This approximates contact inhibited tumour growth with proliferation at the edge only [86]. Even in this growth regime, we begin to see necrotic areas in the bulk of the tumour (box 2, blue trace, Figure 5.7). The second qualitatively different regime is when the density decreases to $\Theta = 0.0033$, when the vessel spacing of the vasculature is such that it has lost its ability to maintain

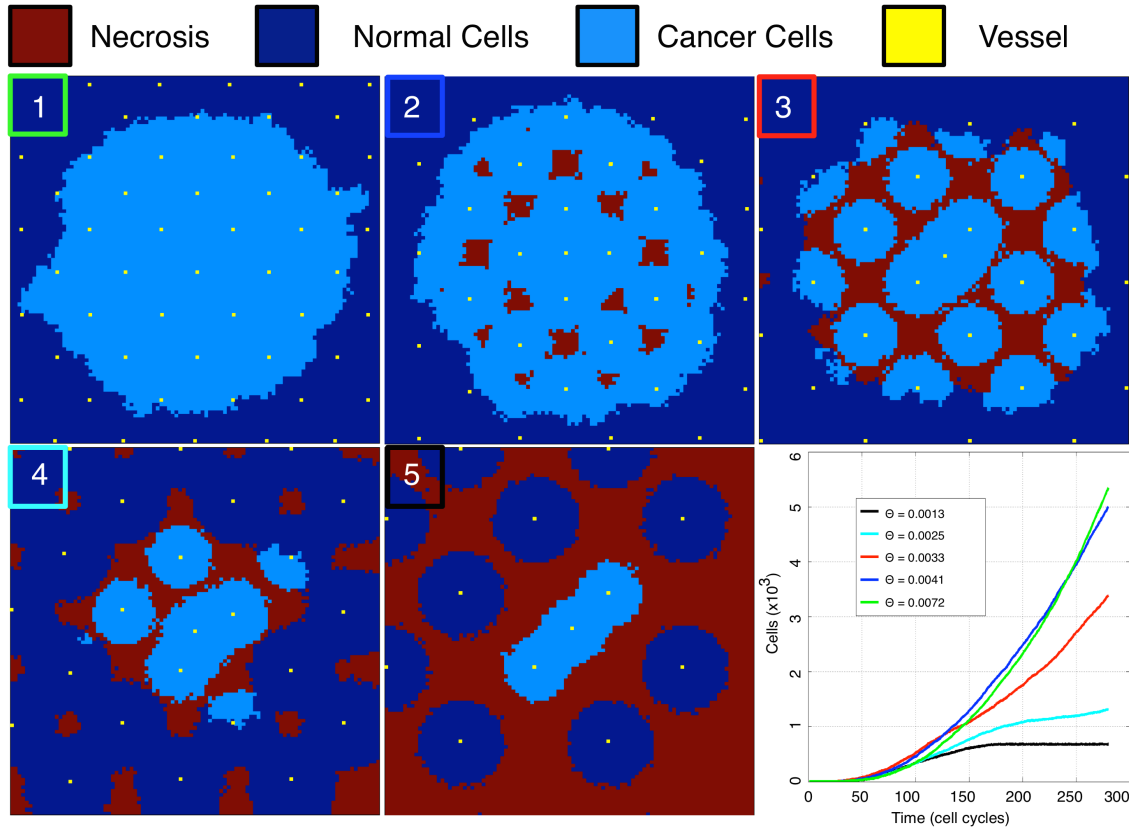


Figure 5.7: Increasing spacing between vessels slows tumour growth and creates areas of necrosis. Tumours are grown in equal sized, regularly vascularised domains with decreasing vessel density from box 1 through 5 (from top left: vessel density, $\Theta = 0.0072$, 0.0041 , 0.0033 , 0.0025 , 0.0013). All plots are at time $t = 280$. Only the smallest vessel density (0.0013 in this figure) entirely constrains growth. Growth over the first 280 time steps is plotted in the lower right.

continuous populations of tumour cells (box 3, red trace, Figure 5.7). In this regime, we find that the tumour's growth rate begins to slow, and further, that the boundary begins to become irregularly shaped as the tumour cells have to persist in regions of hypoxia long enough to place daughters in the new, adjacent, areas of increasing oxygen. Further reduction in vessel density to $\Theta = 0.0025$ results in areas of necrosis forming in the healthy tissue and a sharp reduction in tumour growth rate (box 4, cyan trace, Figure 5.7), as the hypoxic 'valleys' grow deeper and wider. The final scenario is defined by tissue that is so poorly vascularised that the healthy tissue can not maintain contiguous populations, and the cancer can not continue growth beyond the area of initial vessel oxygenation (box 5, black trace, Figure 5.7), and would need to produce its own vessels to grow further.

Parameter sensitivity in limiting cases

In the previous HCA simulations, we varied the vascular spacing and observed the changes in tumour and healthy cell occupancy. This effect is not limited to only the spacing between vessels, but indeed to other critical model parameters, specifically: the chosen levels of oxygen consumption for cancer and healthy cells (μ_H and μ_T); the oxygen diffusion coefficient (D_c); and the value of oxygen chosen below which cell death occurs (c_{ap}).

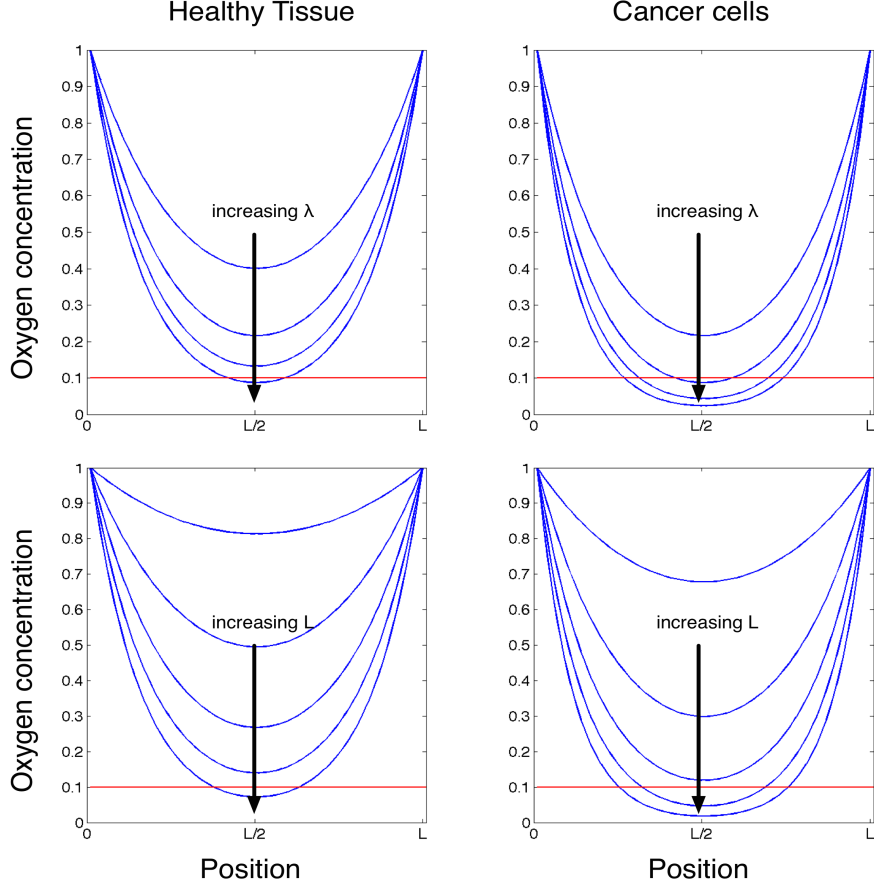


Figure 5.8: Qualitative parameter sensitivity: an analytic representation. We plot the steady state solution from equation (5.11) for a number of non-dimensional parameter combinations for both healthy (left) and tumour (right) cells. As we vary the cellular oxygen consumption, λ (range $\lambda_H = 1 \times 10^{-2} - 4 \times 10^{-2}$ top left and $\lambda_T = 2 \times 10^{-2} - 8 \times 10^{-2}$ top right, with $L = 20$ for both) and inter-vessel distance, L (range $L = 6 - 30$ with $\lambda_H = 2 \times 10^{-2}$ bottom left and $\lambda_T = 4 \times 10^{-2}$ bottom right) we see different proportions of the oxygen curve below c_{ap} (red line, all plots), corresponding to the emergence of necrosis in a full biophysical simulation as we saw in Figure 5.7.

To illustrate these sensitivities, we analytically consider a spatially simplified, one-dimensional problem. We consider a spatial domain $[0, L]$ with two ‘vessels’ located at $x = 0$ and $x = L$. We assume that the oxygen concentration $c(x)$ takes its maximal value at each vessel, $c(0) = c(L) = 1$, which we treat as Dirichlet boundary conditions. We then assume the entire domain to be filled with cells that consume oxygen at a constant

(non-Michaelis-Menten) rate $\lambda_i c$, with $i \in \{H, T\}$ representing the differential oxygen consumption between healthy and tumour cells. For simplicity, we take the consumption rate to be constant for a given cell type. Assuming the system to be in equilibrium, we explore the relationships between the vessel spacing, L , the critical oxygen concentration below which death occurs, c_{crit} , and the differential consumption rates λ_i . The steady-state oxygen concentration satisfies

$$D \frac{d^2 c}{dx^2} = \lambda_i c \quad (5.10)$$

subject to $c(0) = c(L) = 1$. The solution is thus given by

$$c(x) = \left(\frac{\sinh(\phi x) + \sinh(\phi(L - x))}{\sinh(\phi L)} \right), \quad (5.11)$$

where we have defined the lumped parameter $\phi = \sqrt{\lambda_i/D}$.

As we vary the values of these critical parameters and plot (Figure 5.8) the resulting solution of equation (5.11), we see the same qualitative behaviours that we saw in Figure 5.7. Specifically, we see parameter combinations which fully support tumour and healthy tissue growth, combinations that support full healthy tissue growth, but in which necrosis emerges for tumour tissue, and combinations in which there is necrosis in both. Further, we can see larger and larger gaps between areas of viable cells, suggesting regimes in which tumour cells starting at one vessel may not be able to invade to the other, a situation resembling the avascular growth limit often seen in early, pre-angiogenic tumours.

Exact solution

To better understand the relationships plotted in Figure 5.8 we solve for the exact relationship between the parameter values (λ, D, L) for which equation (5.11) satisfies $c/c_{\text{max}} \geq c_{\text{ap}}$. Considering the non-dimensional value of $c_{\text{max}} = 1$, we have

$$c_{\text{ap}} = \frac{e^{\phi x} - e^{-\phi x} + e^{\phi(L-x)} - e^{-\phi(L-x)}}{e^{\phi L} - e^{-\phi L}}, \quad (5.12)$$

which can be simplified to a quadratic in $e^{\phi x}$ which yields the solution

$$x = \left(\frac{1}{\phi} \right) \ln \left(\frac{c_{\text{ap}}(e^{\phi L} - e^{-\phi L}) \pm \sqrt{[c_{\text{ap}}(e^{\phi L} - e^{-\phi L})]^2 - 4(1 - e^{-\phi L})(e^{\phi L} - 1)}}{2(1 - e^{-\phi L})} \right), \quad (5.13)$$

representing the positions where the oxygen value crosses c_{ap} . We can then solve for specific cases, depending on the value of the discriminant: the case when there are no real roots (i.e. $[c_{\text{ap}}(e^{\phi L} - e^{-\phi L})]^2 < 4(1 - e^{-\phi L})(e^{\phi L} - 1)$) representing a situation where there is sufficient oxygen throughout the domain to support cells in all positions, as in

the first regime of vascular spacing in Figure 5.7; the case where there is a single root (i.e. $[c_{\text{ap}}(e^{\phi L} - e^{-\phi L})]^2 = 4(1 - e^{-\phi L})(e^{\phi L} - 1)$) representing the onset of necrosis, as in the second regime of vascular spacing in Figure 5.7; and when there are two roots which move away from one another with increasing values of λ and L (i.e. $[c_{\text{ap}}(e^{\phi L} - e^{-\phi L})]^2 > 4(1 - e^{-\phi L})(e^{\phi L} - 1)$), as in the final regime of Figure 5.7. We plot these relationships in Figure 5.9, where we see, as in our qualitative example, that the size of the necrotic region increases monotonically with both basal uptake and domain size (distance between vessels).

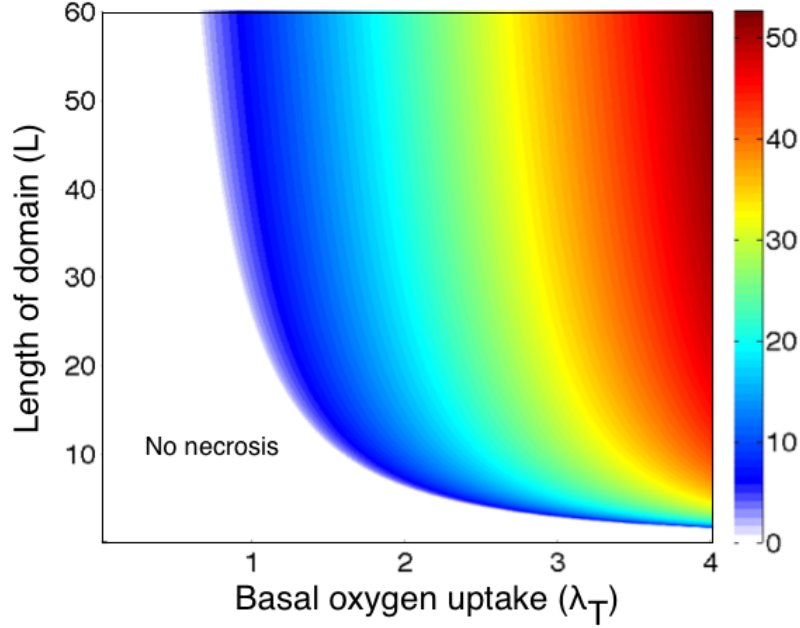


Figure 5.9: Size of necrotic area increases monotonically with oxygen uptake and distance between vessels. We plot the distance between the real roots for equation (5.13), the interface between positive values and zero represents the onset of necrosis and a value of identically zero represents the area in parameter space when there are no real roots and therefore the area in which there is no necrosis. As expected, the size of the necrotic area increases monotonically with both λ_T and L .

Sensitivity analysis in numerical simulation

To assay the numerical model for sensitivity to parameters, we measure the cellularity and cell-oxygen distributions for the regular vascularity example reported in the centre panel of Figure 5.5, with $\Theta = 0.0027$ (or a regular spacing of 14 cell diameters). From the parameter set modelled in Figure 5.5 ($D_c = 0.1, r_c = 1, K_m = 0.01$ and $c_{\text{ap}} = 0.1$) we vary each parameter by approximately three orders of magnitude and report the ranges for each mode of cellular oxygen distribution and cellularity in Table 5.2.

Table 5.2: Sensitivity analysis.

Parameter	Value	Cellularity	Mean cellular-oxygen	Std cellular-oxygen	Skewness cellular-oxygen
D_c					
	0.001	0.0015	0.5198	0.0286	-0.333
	0.01	0.0449	0.4623	0.0916	0.0197
	0.1	0.4336	0.2346	0.1029	1.5947
	0.2	0.8489	0.1931	0.0917	2.2165
r_c					
	0.01	0.9969	0.9796	0.0023	2.2728
	0.1	0.9969	0.7969	0.0227	2.273
	1	0.4336	0.2346	0.1029	1.5947
	10	0.0517	0.3398	0.1296	0.0338
	20	0.0265	0.3716	0.1458	-0.3105
K_m					
	0.001	0.398	0.2705	0.1018	1.4904
	0.01	0.4338	0.2346	0.1029	1.5947
	0.1	0.6862	0.1817	0.0926	2.2194
	0.2	0.9448	0.1702	1.0872	2.494
	0.5	0.9969	0.2816	0.0772	2.3861
c_{ap}					
	0.001	0.8674	0.054	0.0924	2.8366
	0.01	0.6088	0.0849	0.1016	2.3506
	0.1	0.4336	0.2346	0.1029	1.5847
	0.2	0.3495	0.3668	0.0952	1.495
	0.5	0.1883	0.6399	0.0617	1.634

As expected, increasing cellular oxygen consumption (r_c) strongly influences the ability for a given vascular architecture to support cells, with greater consumption correlating with decreased cellularity. The mean oxygen concentration experienced by cells goes down and then up slightly as the number of cells decreases drastically. Variation in the diffusion coefficient, D_c , strongly affects the ability for a domain to support cells, and also the mean and skewness of the resulting oxygen distribution, but affects the standard deviation relatively little. Our choice of threshold for apoptosis, c_{ap} , intuitively has a strong effect on the cellularity, and then inversely on mean cellular oxygen as fewer and fewer cells are competing for the same oxygen. The standard deviation is affected little, and the skewness decreases at first and then increases as the number of cells becomes smaller and smaller.

5.4.2 Tumour growth in heterogeneous vascular domains

Tissue carrying capacity and vessel density

We saw in the previous sections that varying vessel density, on average, changed the carrying capacity and cell-oxygen distribution of the resultant tissue section. Here, we will explore this effect more carefully. We consider a domain of size 73×73 (where 73 is chosen for ease of comparison to regular vessel spacing examples) and vary vessel number from very low until we reach saturation of cells (in this case from 3 to 236 vessels). In each simulation, we initialize the domain full of tumour cells and then allow the tissue to experience oxygen dependent birth and death until a dynamic equilibrium is reached. We define this dynamic equilibrium as when the model simulation has reached a state where there is less than a 1% change in cell number for 50 time steps. We choose this initial condition, instead of seeding the domain with a single cell, because we are not interested in this section with the growth kinetics, but instead on within tumour bulk equilibrium characteristics. From this point, we then record and take the average of 100 time steps of cell and oxygen concentration data. For each vessel number, we run 500 simulations, as detailed above, with random vessel configuration, and plot the result of each set of simulations and the associated standard deviation (inset) in Figure 5.10.

As expected, we see a monotonic approach to cellular saturation as vessel number increases. When we plot the standard deviation, we find that it is significantly higher in the region that is most physiologically relevant, and that these differences can reach as high as 10% (Figure 5.10), making cellularity estimates based on vascular density difficult and unreliable. How the different patterns represented in these families of simulations can affect the heterogeneity of oxygen within the domain, and the resultant cell-oxygen distributions, is yet unknown.

Effect on cell-oxygen distribution in heterogeneous domains

To begin to understand how the patterning of the vasculature affects the cell-oxygen distribution within our simulations, we consider two cases from the sample of our simulations for two separate vessel densities. We plot and compare the vascular patterns which support the maximum and minimum cellular populations from two given vascular densities (Figure 5.11). We plot the minimum cellular population represented in the left column at low (top) and high (bottom) relative vascular densities and the maximum populations in the right column.

In comparing these limiting cases, we observe similarities and differences. In both minimum population cases, there is a large peak of cells near to the hypoxic minimum (c_{ap}) as well as a smaller peak around an oxygen concentration of 0.5. In the maximum

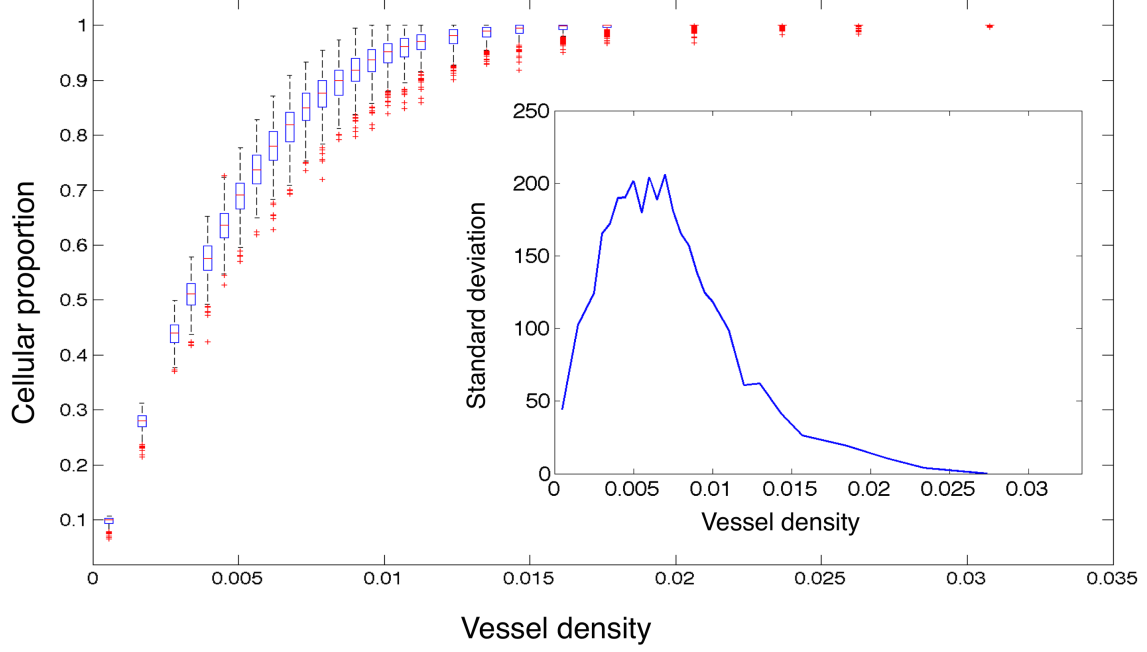


Figure 5.10: Equilibrium cell density versus vessel density in the HCA. We plot the results of several families of simulations seeded with equal numbers of randomly placed vessels on a fixed domain of size 73×73 . Each data point is represented by a box which is centred on the mean of 500 simulations, each of which is the average of 100 time points after dynamic equilibrium is reached. The edges of the boxes represent the 25th and 75th percentile, the whiskers extend to the most extreme data points not considered outliers. Outliers are defined as any simulation outside approximately 2.7 standard deviations, and they are plotted as red crosses. Inset: we plot the standard deviations at each vessel density in raw cell number. A cell density of 1 corresponds to 5329 cells in the domain.

cellularity examples, however, these similarities disappear. In the low vessel density, we see that the hypoxic population is maintained, but we have lost the central, well oxygenated fraction. While in the high vessel density, maximum population, this is reversed, and we lose the hypoxic population and the entire population is centred around normoxia. This opposite effect, losing the hypoxic fraction in one case, and increasing it in the other, would drastically change radiation efficacy in these two cases, suggesting that vascular organization, not just density, could play an important role in clinical radiation therapy, an effect which we will explore in Chapter 6.

5.5 Conclusions

In this chapter we have built, parameterised, and explored an HCA model of vascularised tumour growth. We have used parameters from various experimental sources, mainly focussed on glioblastoma, a deadly primary brain tumour, but endeavour to draw more general, qualitative conclusions from the results. Our initial results highlight the importance not only of vascular density, but also of vascular patterning in the resulting

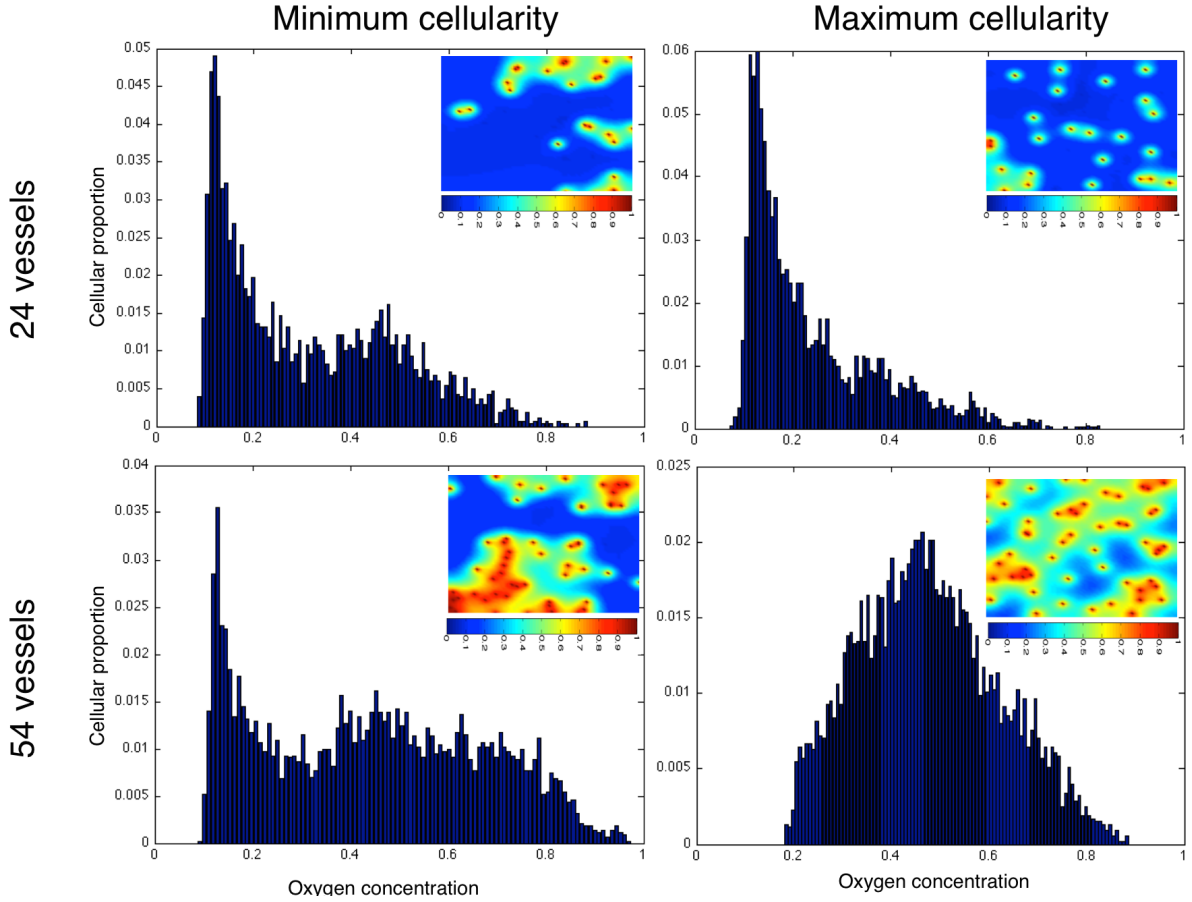


Figure 5.11: Wide variation is present in cell-oxygen distribution between simulations with constant vessel density. We plot the equilibrium cell-oxygen distributions and spatial oxygen distributions (inset) from the minimum (left) and maximum (right) cellularity examples from our family of 500 simulations seen in Figure 5.10. Top is a low vascular density (24 vessels) and bottom is a higher density example (54 vessels).

tissue characteristics, in terms not only of growth kinetics, but importantly, for cellular oxygenation status, a key regulator of radiation therapy efficacy, among other things.

While some of the main qualitative results from this chapter corroborate results reported previously by Alarcón et al. [4] who used a similar hybrid cellular automaton model, our work differs in several key ways. Specifically, the vasculature in their model lies in plane rather than *en face*, a choice which was made consciously in our model to ease future translational validation with patient tissue samples. Further, we have ignored several mechanisms of competition in order to focus on defining summary measures of cell-oxygenation status. These differences aside, Alarcón et al. found, as we did, that heterogeneity on the scale of oxygen concentration affects cell growth, both in overall speed of tumour growth and also in shape of resulting tissue. In both model systems, tumours growing in homogeneous vascular domains grow approximately the same in all directions and those growing in heterogeneous domains grow in strikingly asymmetrical patterns, and, in certain circumstances, are entirely constrained by hypoxia. Further,

they reported significant heterogeneity in steady state oxygen concentration across their domains which was dependent on vascular position and simulated blood flow dynamics. A similar finding was also reported by Al-Shammari et al. in a biophysical model of healthy muscle tissue [3], this time in a system utilizing *en face* oxygen sources, as in our work. We have observed similar changes in cell-oxygen distributions at equilibrium, and indeed, will focus on these going forward.

The current work extends these results from previous authors in some useful ways and also raises a number of novel questions worth investigating. By introducing the non-spatial summary measure of the automaton domain, which we refer to as the ‘cell-oxygen distribution’, we provide an opportunity for efficient computation of radiation efficacy for given vessel patterns (e.g. from histology), which we will explore in the coming chapter. Further, by exploring many different random patterns of vessels, we have begun to quantify the variation possible for different ranges of vascular densities seen on imaging, showing a way in which current estimates of cellularity could be improved through utilization of information from the histologic scale.

In Chapter 6 we quantify the effects of heterogeneous oxygen concentrations on radiation efficacy and determine whether or not this is predictable from vessel patterning which could be gleaned from tissue sections. Our overarching goal is to create a method to link microscopic, histologic, information with imaging scale information to give deeper insight into tissue level metrics for therapeutic benefit.

Chapter 6

Effects of vascular patterning on radiation treatment efficacy

6.1 Introduction

In Chapter 5 we observed both widely varying carrying capacities and also cell-oxygen distributions within, and across, families of simulations of tumour growth with varying vascular densities. These two factors, the raw number of cells and the oxygen concentration they experience, are the main drivers of radiation response in tissue and tumour control probability (TCP) in tumours. In this chapter then, we will work to take these observations and build a measure which could eventually be applied to patient data to help optimise radiation therapy.

6.1.1 Previous work

Radiation biologists have, for many years, understood the importance of cell biological and microenvironmental factors on radiation response. Current radiation therapy dose planning, however, largely neglects this information and is, instead, based on years of clinical experience using intuition and trial and error. As such, there remains limited tailoring of dose planning to an individual patient's tumour. With the advent of modern quantitative histologic [141] and biological imaging methods [79], however, this paradigm is poised to shift.

Research in this area over the last decade [150, 225, 226, 227] has sought to understand the macroscopic spatial distribution of hypoxia within tumours using non-invasive imaging. This information has then been utilised to develop spatially heterogeneous dose-plans to improve tumour control. For example, Malinen et al. [150] inferred average oxygen concentrations from radiocontrast concentrations measured by Dynamic Contrast Enhanced (DCE) MRI in tissue. Other work to understand the effects of radiation in individual patients has utilized human patient MRI scans in combination with

mathematical models of tumour growth incorporating heterogeneity in aggregate tumour phenotype (proliferative or invasive), without consideration of oxygen effects [199], to explain different radiation responses measured by changes in tumour size over time.

Other groups have considered radiation and chemotherapy together, for example Powathil et al. focussed on the problem of treatment optimisation [191, 192] using cell cycle-specific agents and radiation therapy. In these papers, the authors use the software package CompuCell 3d [111] and build a hybrid cellular automaton (CA) model of tumour growth in a heterogeneous vascular domain. They carefully consider the effect of oxygen and radiation on both tumour growth dynamics as well as individual cell cycles, another factor that has been shown to be affected by hypoxia [5]. Their central effort is then focussed on optimisation of combination therapy with cell cycle-specific drug therapy and radiation. Their main conclusion is that the timing of these interdependent treatments is a key driver of treatment success, and that consideration of spatial dependencies is also critical.

Timing of radiotherapy was again considered by Kempf et al. [121] who built a three-dimensional hybrid CA model of *in vitro* tumour spheroid growth. The authors considered cell cycle dynamics and spatially heterogeneous oxygen and nutrients. While there was no vasculature, the nutrients and oxygen were delivered via non-linear diffusion, from the outside of the spheroid, as in an experimental system. The effect of radiation therapy on tumour cells was modelled as a combination of effects from oxygen concentration as well as cell cycle. They then use this system to test re-oxygenation dynamics during multiple, sequential treatments with radiation therapy. Like previously discussed work, the authors described significant heterogeneity in radiation response with differing treatment schedules, but also report predictable re-oxygenation dynamics. They suggest further study into these dynamics with consideration for altering clinical treatment schedules in light of their findings. More recently, a computational model of a tumour comprised of populations with proliferative heterogeneity [10] was used to compare tumour response to a range of spatially heterogeneous radiation dose plans.

Of direct relevance to this work as well, but published after this work appeared in *PLoS Computational Biology* [212], is work done by Grogan et al. [98], in which a model of cancer growth in a heterogeneously vascularised domain in three spatial dimensions is built. The model is based on the same framework as the one we present here: the hybrid, discrete-continuous cellular automaton framework. Vessels are considered, like in our work, as constant valued sources from which oxygen can diffuse, and considered Dirichlet boundary conditions for a reaction-diffusion PDE. While vessels in this work are off lattice, they are interpolated to lattice positions for calculations. Cells are somewhat more complex in their formulation than ours, as they have a cell cycle built in as well,

as in Owen et al. [179]. Like our model, cells can be quiescent in hypoxia and respond to radiation in the same way we will present later in this chapter, obeying the oxygen enhancement effect. Besides being in three spatial dimensions, Grogan et al. also simulate the effects of regrowth on the overall response to radiation therapy. Their findings mostly corroborate ours in the most comparable scenarios (two-dimensional approximation with point sources), but do show that microvasculature architecture is much less important when the full three-dimensional simulation is considered. Further, their work shows that different geometrical representations of tissue slices (vessels running through or in plane with the slice) can over-, or under-estimate cell kill, an effect which is exacerbated in certain oxygen consumption regimes.

6.1.2 Aim of work

The aforementioned models cover a wide range of mathematical techniques and many different biological aspects of tumour growth. The impact of tumour growth and the resultant tissue and vascular organisation on radiation treatment, however, remains much less well studied. For this reason, we will keep our model as simple as possible in order to be able to draw the most general conclusions. As we aim to understand how spatially heterogeneous concentrations of oxygen affect radiation therapy efficacy, and to eventually translate this knowledge to the clinic, we will build on the previous work using the hybrid CA paradigm.

In this chapter, we aim to build on the observations in Chapter 5 concerning cell-oxygen distributions to study how these can impact on the efficacy of radiation therapy. By identifying broad relationships between *in silico* tissue architecture and radiation response, we aim to progress toward a useful measure to apply to radiation plan optimization from information extracted from biopsies and imaging.

6.1.3 Structure of chapter

The remainder of this chapter is structured as follows. We begin, in Section 6.2, with a brief overview of the basic biology underlying radiation response, including the equations for surviving fraction under homogeneous and heterogeneous oxygen concentrations. In Section 6.3, we illustrate how different vascular patterns lead to widely varying radiation efficacy and present a simplified model to understand the factors driving this. In Section 6.4, we present an overview of several methods of measuring vessel patterning and then compare these measures in their ability to predict how different vascular patterning scenarios respond to radiation in Section 6.5. Finally, we conclude by discussing how

these findings apply under the current understanding of drugs that normalise vasculature, and present potential avenues for future research and translational applications of these findings.

6.2 Established radiobiology

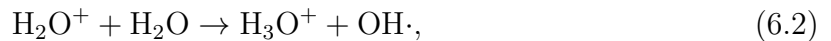
Here, we introduce the basics of radiobiological modelling. We will review concepts that include direct and indirect mechanisms of DNA damage, TCP, the linear-quadratic model of cell survival and the concept of oxygen enhancement ratio (OER).

6.2.1 Proposed mechanisms for DNA damage

X-ray induced DNA damage has been posited to occur mainly via two mechanisms, termed direct and indirect damage, schematised in Figure 6.1. Both mechanisms begin with an incident photon, γ , striking a valence electron in a molecule in the cell. At the energies relevant for radiotherapy modelling, incident photons interact with matter through Compton scattering [44]. Through this process, the photons interact with valence electrons (often termed free electrons) and impart some of their energy, causing a deflection of the photon, and a ‘scattering’ of the electron, freeing it from its orbit. This free electron can then either directly interact with the DNA, causing a break or breaks in the strand, or interact with another nearby molecule. Interaction of the scattered photon with a water molecule also causes a chain of events which begins with the ionisation of the water molecule, giving it an overall positive charge:



This ion then interacts with another water molecule to form the highly reactive hydroxyl radical ($\text{OH}\cdot$),



which can then diffuse locally, causing ‘indirect’ DNA damage. This indirect pathway is thought to be responsible for the majority of DNA damage from photon irradiation [100].

6.2.2 The linear-quadratic model of cell kill

The standard description of the surviving fraction (SF), proposed by Lea [132] and since extended by many [25, 76], of cells after a given dose (d) of photon irradiation (measured in Gray (Gy), which is the amount of energy in J absorbed per kg of tissue) in n fractions is given by

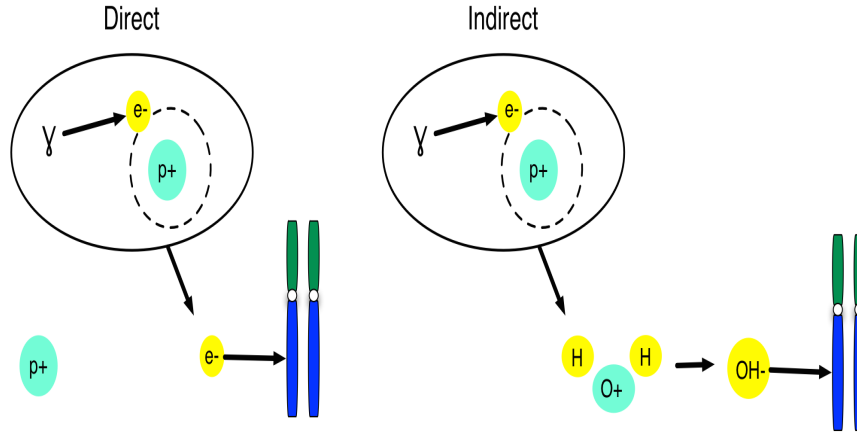


Figure 6.1: A schematic describing the difference between direct (left) and indirect (right) radiation damage. In direct damage, an electron (yellow e^-) is freed from its orbit around a positively charged nucleus (p^+) through interaction with a γ particle and secondarily strikes the DNA which is represented here as a chromosome, causing damage. In indirect damage, the electron freed interacts with a water molecule (H_2O) and subsequently forms an hydroxyl radical, which can then diffuse, causing DNA damage.

$$SF = e^{-n(\alpha d + \beta d^2)}. \quad (6.3)$$

Here, α and β are radiobiologic parameters associated phenomenologically with cell kill resulting from ‘single hit’ events (α) and ‘double hit’ events (β), respectively, which we have schematised in Figure 6.2. This measure, while originally derived to fit *in vitro* data [132], has since been widely used to also describe clinical efficacy of radiation [48, 49].

6.2.3 Dependence of cell kill on oxygen concentration

As described in Section 6.2.1, the basic interaction driving DNA damage from radiation is mediated either by free electrons or by free radicals formed by the interaction of photons with water. The DNA damage instantiated by these mechanisms can be repaired via reduction by locally available biomolecules containing sulfhydryl groups [100]. This damage can be made more permanent, however, by the presence of molecular oxygen, which binds with the DNA radical, forming a ‘non-restorable’ organic peroxide. This damage ‘fixation’ by oxygen creates a situation in which the damage now requires enzymatic repair, which occurs on a much longer timescale [100]. This explanation of the radiation sensitising effect of oxygen is termed the ‘oxygen fixation hypothesis’, and remains the standard explanation, but has been called into question [70]. Regardless of the mechanism of radiation sensitisation by oxygen, the effect is well documented, and has been modelled by replacing the parameters α and β in equation (6.3) by

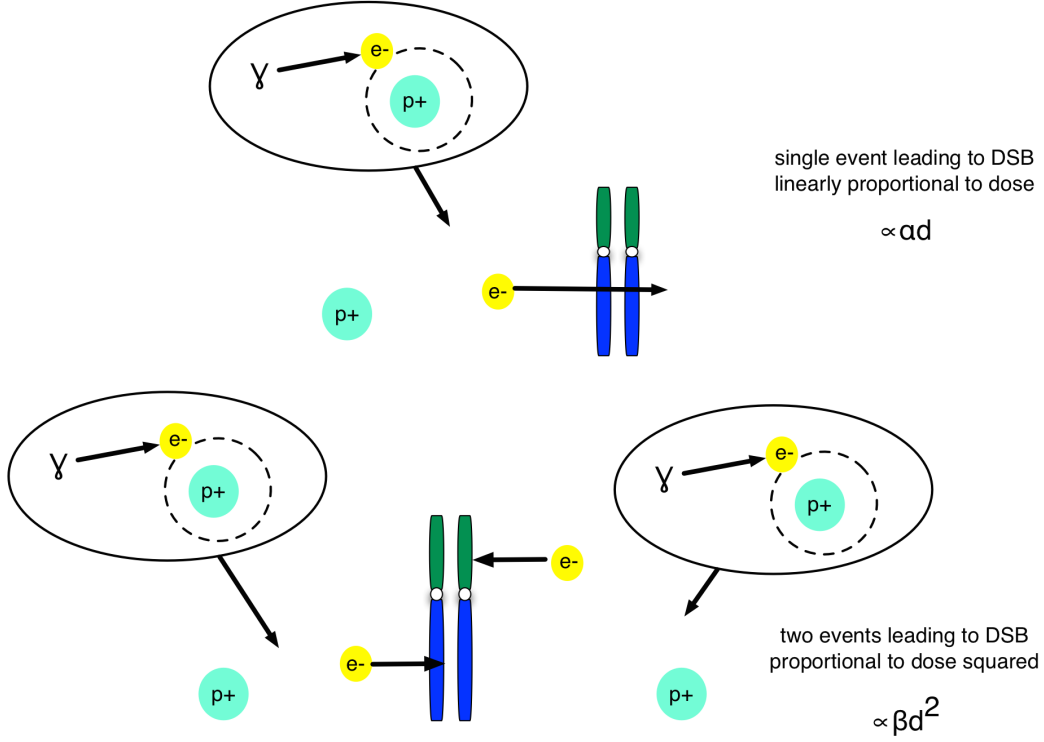


Figure 6.2: A schematic describing the linear-quadratic model of radiation effect. In each case, a single photon (γ) interacts with the electron cloud of some molecule, and frees a valence electron by Compton scattering. In the case of a ‘single hit’ (top), when a single event causes a double strand break (DSB) in the DNA, the probability of cell death is linearly proportional to dose (d), whereas the probability of cell death in the case of a ‘double hit’ is proportional to dose squared.

$$\alpha_i = \frac{\alpha_{max}}{\text{OER}_\alpha(c_i)}, \quad \beta_i = \frac{\beta_{max}}{\text{OER}_\beta(c_i)^2}. \quad (6.4)$$

The terms α_{max} and β_{max} are, respectively, the values of α and β under fully physiologically oxygenated conditions. α_i , β_i and c_i are, respectively, the values of α , β and oxygen c , at oxygen concentration i . We define the OER as a function of the oxygen concentration by using the relation established by experimental correlation by Chapman et al. [40], and Palcic and Skarsgard [182]. In this formulation

$$\text{OER}(c_i) = \frac{(\text{OER}_{\max} - \text{OER}_{\min})K_m}{c_i + K_m} + \text{OER}_{\min}, \quad (6.5)$$

where $K_m = 3.28\text{mmHg}$, $\text{OER}_{\alpha,\min} = \text{OER}_{\beta,\min} = 1$, $\text{OER}_{\alpha,\max} = 1.75$ and $\text{OER}_{\beta,\max} = 3.25$ [40, 182, 249] and we solve for the full range of c_i from 0 to 80mmHg.

In Figure 6.3 we utilize the OER (equations (6.4) and (6.5)) to plot, as an example, in blue, the oxygen concentration specific surviving fraction (equation (6.3)) for a fixed dose of radiation, $d = 2\text{Gy}$. It is evident that the assumption that all cells in a given domain experience the average oxygen concentration (green trace) will, depending on

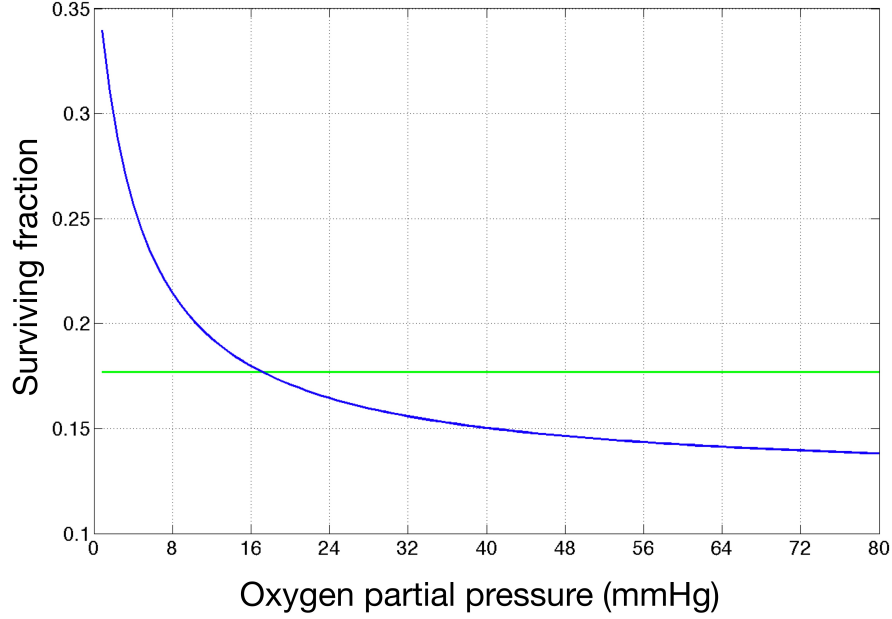


Figure 6.3: Surviving fraction is a non-linear function of oxygen concentration. We plot, in blue, the surviving fraction from equation (6.3) calculated for each oxygen concentration as modified by the OER in equations (6.4) and (6.5). This is in contrast to the green line which represents surviving fraction assuming all cells experience the mean oxygen of a given domain.

the actual distribution of oxygen, cause errors in estimates for cell kill under radiation therapy. This error will be increasingly exacerbated by populations of cells whose cell-oxygen distributions are skewed left (toward low oxygen concentrations), highlighting the need to consider not just the mean oxygen in the domain, but instead the true distribution experienced by cells.

6.2.4 Tumour control probability

TCP is a measure that approximates the probability of eradicating all cells within a tumour and thereby controlling (in this case ‘curing’) it. While a number of methods, both deterministic and stochastic [53], have been proposed to define such a measure (recently reviewed by Zaider and Hanin [253]), the most commonly used formulation assumes that the number of surviving clonogenic cells after radiation follows Poisson statistics and that the TCP is calculated by

$$\text{TCP} = e^{-N_0 \text{SF}}, \quad (6.6)$$

where N_0 is the initial number of clonogens (cells capable of creating viable colonies of cells, which are all cells in our simulations), and SF is the surviving fraction calculated by equation (6.3).

As we are simulating microscopic oxygen and cellular dynamics, extending our model to realistic tumour cell numbers ($\mathcal{O}10^{11}$ th cells) is infeasible. To compare the effect of radiation on different populations of cells in our model, we calculate the total number of cells surviving a given therapeutic intervention. To this end, we consider the individual survival probability of cells in a given spatial domain experiencing a given oxygen concentration. Let N_i^k be the number of cells in our non-dimensionalised cellular automaton model that experience a oxygen concentration in the interval $[i\Delta c, (i+1)\Delta c)$ at time step $t_k = k\Delta t$, where Δt is as defined in Chapter 5 and $i \in \{0, \dots, M\}$. Here we define $\Delta c = 0.01$, and $M\Delta c = 1$, reflecting the fact that the oxygen concentration is bounded by its non-dimensional value in the arterial blood, which is 1. We make the assumption that radiation effect is instantaneous, and represent the time step after radiation as t_{k+1} . The number of cells experiencing an oxygen concentration in the interval $[i\Delta c, (i+1)\Delta c)$ immediately following a single radiation dose d is thus given by

$$N_i^{k+1} = N_i^k e^{-(\alpha_i d + \beta_i d^2)}, \quad (6.7)$$

where we use the average oxygen in each interval to calculate α_i and β_i . Thus the total number of surviving cells at time $k+1$ is

$$N^{k+1} = \sum_{i=0}^M N_i^{k+1}. \quad (6.8)$$

While the assumption that the kill from radiation is instantaneous is somewhat strong, for the purposes of our study it is appropriate as we will only compare the results of single doses of radiation without consideration of repopulation dynamics. That is, while the resultant oxygen dynamics would be changed in the ensuing time steps differently dependent on how many cells survived, since we will compare only the effects of the single dose, these effects will not carry forward. This assumption will have to be loosened if we desire, in future, to study multiple doses of radiation, or the resultant tumour growth after radiation.

6.3 The effect of heterogeneous vascular patterns on surviving cell number

As we have seen in Chapter 5, the organisation of vessels can have a significant effect on both the cell carrying capacity of a domain and also the resulting cell-oxygen distribution of the cells inhabiting the domain. Our computation of total surviving cells, a measure critical for comparison across heterogeneous samples, and to understanding radiation efficacy at larger scales, is directly a function of cell number and surviving fraction. These

two phenomena are, in turn, directly driven by cell-oxygen distribution and carrying capacity, both of which are affected by the distribution of oxygen from blood vessels. How these phenomena are governed by blood vessel organisation, however, we have not yet elucidated. In this section we will investigate the effects of differing vascular patterns on radiation efficacy.

6.3.1 Surviving cell number after radiation varies widely with vessel pattern

To begin to understand how the changes in cell-oxygen distribution observed in Chapter 5 modulate the effects of radiation, we calculate the number of cells surviving after a single simulated dose of 2Gy of radiation in each of the simulations reported in Figure 5.10 and plot the results as a function of vessel density (Figure 6.4). In Chapter 5, we found that as the vessel density increases, the mean number of cells supported by the domain increases. All other things being equal, this should translate into an increase in the number of surviving cells after radiation. We see that this relation holds for the low vessel density simulation families, but as the vessel density increases above 0.01, the mean number of surviving cells begins to decrease. Further, near this transition point, we observe that the standard deviation of surviving cells within each family of simulations peaks (Figure 6.4, inset). In order to better predict the number of surviving cells in a domain, we will need a more comprehensive understanding of the cellular oxygen distribution. While this measurement is simple to perform *in silico*, it is impossible *in vivo* or clinically. To address this, we will endeavour to glean this information from a proxy in the form of a measure of vascular organisation.

6.3.2 A simplified case: two vessels at varying distance

To better understand how vessel organisation affects the number of surviving cells after radiation, we first investigate a simplified model in which there are only two vessels. We will begin by analysing the situation where the vessels are far enough apart such that the cellular populations associated with their oxygen fields do not overlap, and then bring them together to understand how varying levels of overlap affects radiation therapy cell kill.

Changes in oxygen distribution

We have seen in Chapter 5 that for varying vessel densities and configurations the equilibrium cell-oxygen distribution can vary drastically (Figure 5.10). This variation is due to

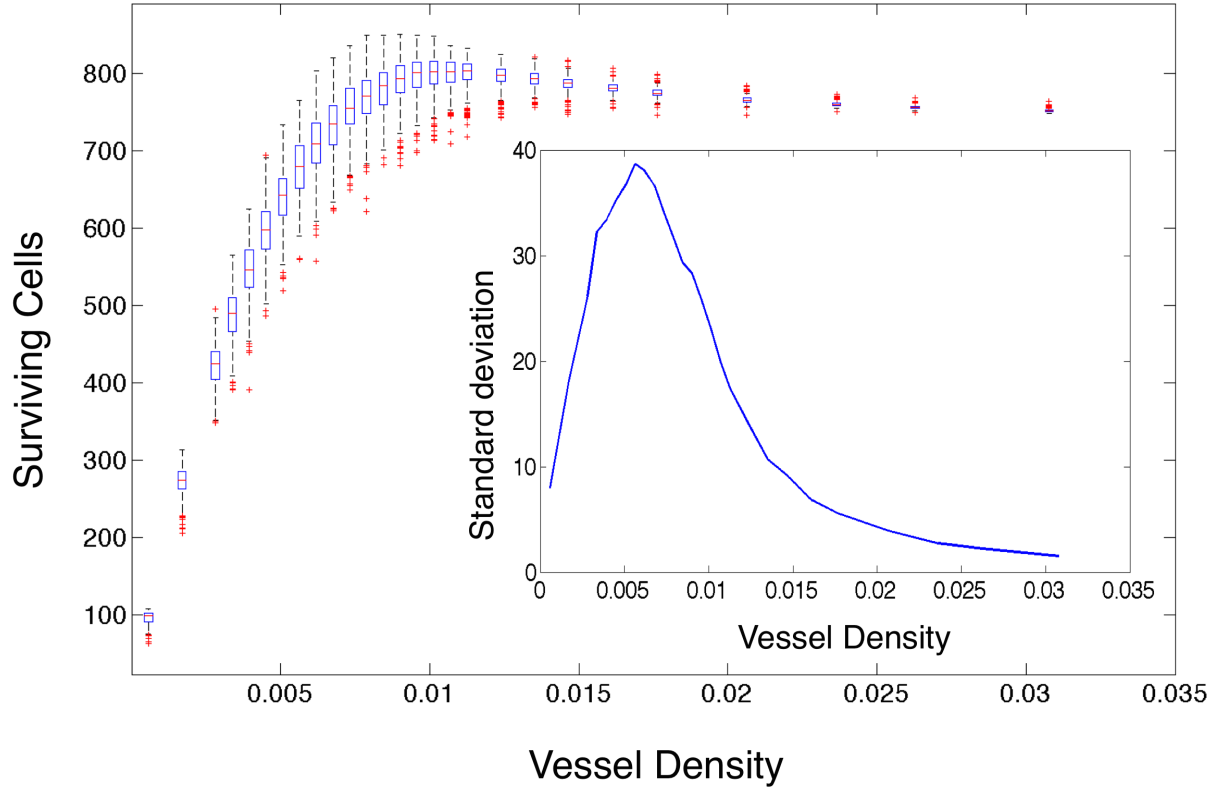


Figure 6.4: Surviving cell number versus vessel density in the HCA. We plot the number of surviving cells after 2Gy of simulated radiation in each simulation as calculated using equation (6.3) modified by the OER from equations (6.4) and (6.5) versus the number of vessels in each case for each of 500 stochastic simulations with constant vessel number, but random placement, on domain size 73×73 at dynamic equilibrium. The edges of the boxes represent the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers. Outliers are defined as any simulation outside approximately 2.7 standard deviations, and they are plotted as red crosses. Inset we plot the standard deviation for each family of simulations versus the vessel number.

the aggregated ‘overlap’ of the associated oxygen field from neighbouring vessels. Depending on the proximity of these vessels, there can be quite varied oxygen distributions. For example, if two vessels lie far enough apart then they may support two non-overlapping cell populations in equilibrium. This configuration results in the most left-skewed cellular oxygen distribution possible, and also the largest total population possible (cf. Figure 6.5). As the distance between two neighbouring vessels is reduced, the populations begin to overlap and the resultant cell-oxygen distribution changes by skewing to the right as the cells in the overlap area receive oxygen supplied by two different vessels (Figure 6.6, right). As the vessels approach and eventually abut one another, we recover a cell-oxygen distribution nearly equivalent to the non-overlapping vessels (Figure 6.5), but with approximately half the total cells.

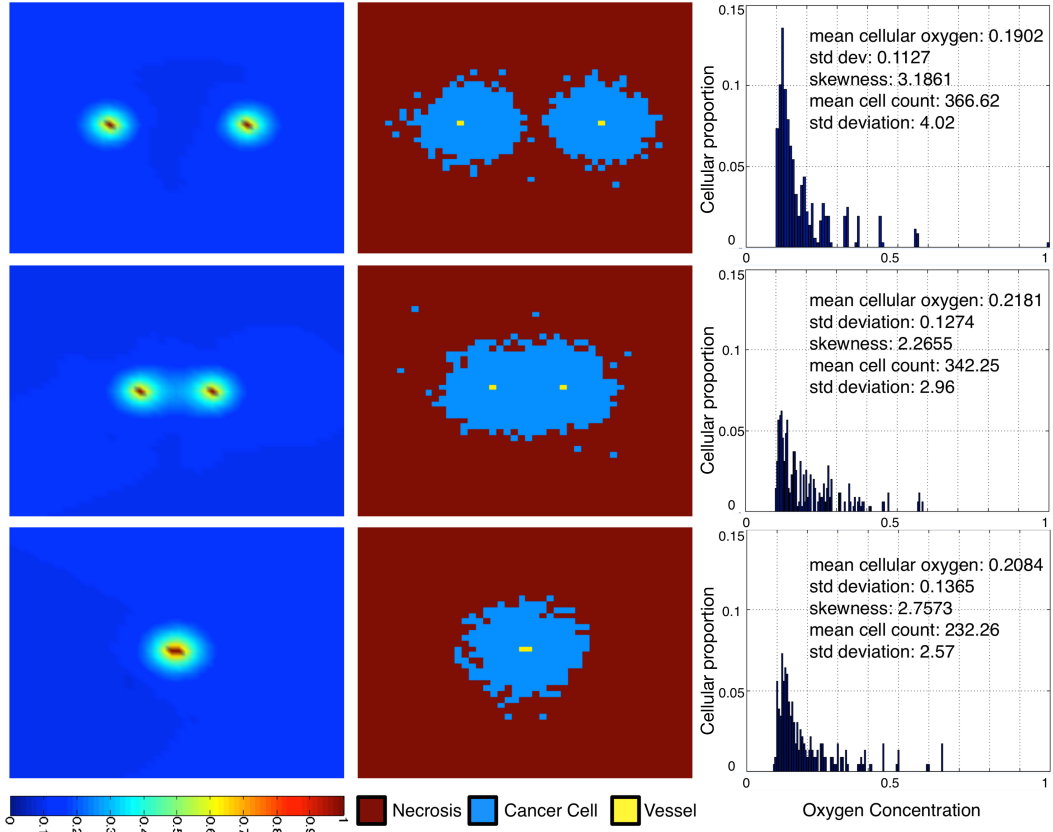


Figure 6.5: Simulation results for a simplified two-vessel case. Vessels that are far enough apart so as to not interact (top), support non-overlapping populations which have both the most cells, and also the most left-skewed cellular oxygen concentration. As the vessels approach one another, and the cellular populations begin to overlap (middle), the cell-oxygen distribution becomes less skewed. When the vessels abut, they support a population more like in the original, non-overlapping case, but with much smaller cell numbers. All parameter values are as listed in Table 5.1, domain size is 50×50 . For each case, a single stochastic simulation is performed. Measures reported, shown in the inset, are averages of the 500 time steps after dynamic equilibrium is reached. As before, dynamic equilibrium is defined as the time from which cell number varies no more than 1% from maximum. The visualised state of the model is the final time step and therefore may not represent the distribution reported exactly.

Carrying capacity change

In addition to the change in skewness, there is the concomitant change in the number of cells that the two vessels can support. To understand these interactions, we plot the relationship between the number of supported cells over 50 simulations at dynamic equilibrium as a function of distance between the two vessels (Figure 6.6, left). In this situation, we see a monotonic increase in the number of cells supported until a plateau at a distance of 14 cell diameters, after which we see no appreciable change, and the distributions essentially become those of separated islands of cells.

We next plot the number of cells surviving after a single radiation dose of 2Gy. In Figure 6.6 on the right we plot two survival curves: the number of surviving cells, assum-

ing that each cell experiences the mean oxygen in the domain (red), and the surviving number of cells using the actual oxygen concentration each individual cell experiences (blue). Here, we calculate a mean oxygen concentration for the entire domain, which is simply the mean oxygen value of all occupied lattice sites. We observe a non-monotonic change as cell kill first increases as the vessels are moved apart as more cells experience higher oxygen concentrations (Figure 6.5, right), and therefore more efficient killing due to radiation damage. As the vessels move farther apart, however, and the equilibrium cell population approaches the separate ‘islands’ of cells state, the surviving cell number drops slightly below the maximum value. In the inset, we plot the difference in these two plots, highlighting the effect that skewness has on our estimates.

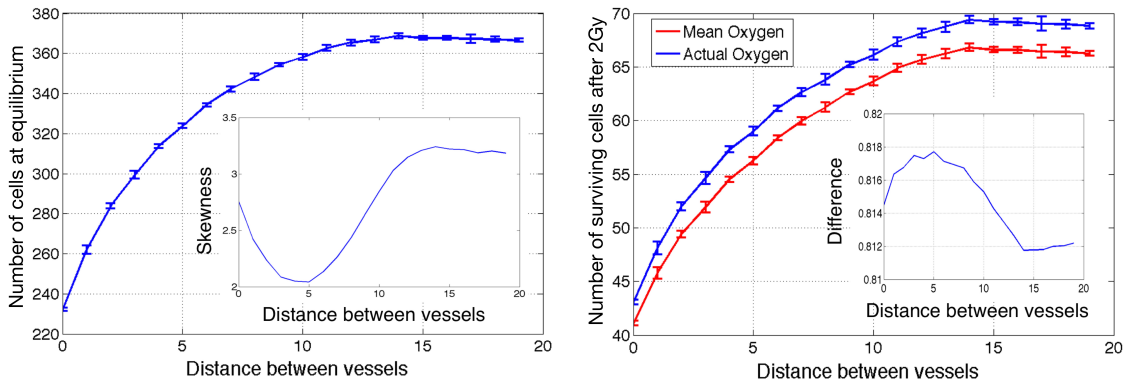


Figure 6.6: Carrying capacity and cellular oxygen distribution are competing effects on cell survival after radiation. (Left) Carrying capacity versus distance between vessels in the two-vessel case. Beyond a certain distance, there is no significant change in carrying capacity, and as the vessels approach one another, it monotonically decreases. Inset: we plot the skewness of the underlying cell-oxygen distributions, which exhibits a bi-phasic dependence on vessel separation. (Right) Surviving fraction assuming all cells experience mean oxygen concentration vs actual cell-oxygen concentration. The blue line represents the effects of the OER in equations (6.4) and (6.5) while the red line represents cell kill assuming all cells experience the mean oxygen concentration of the distribution at dynamic equilibrium, highlighting the effect of population skew on survival as suggested by Figure 6.3.

6.3.3 Vessel density is insufficient to predict radiation effect

We have previously shown that vascular density is a strong predictor for the carrying capacity of the tissue in our model (Figure 5.10). This is intuitive and, for regularly vascularized domains, as we expect in healthy tissue, is sufficient to explain not only the carrying capacity, but also other measures of the cellular distribution (e.g. the shape of the cell-oxygen distribution). We have now shown that this measure is not sufficient when we begin to consider heterogeneously vascularized domains, as we know exist in cancer [63], wherein a single domain could have many possible spatial patterns that

exhibit the same density. The mapping from cell-oxygen distribution to surviving cells is straightforward in the simulations we have presented, as the cell-oxygen distribution can be measured directly, and the subsequent surviving cell number can be calculated using equation (6.8). In clinical practice, however, it is not possible to measure oxygen concentrations at the cellular resolution. It would therefore be of translational value to define a measure by which cell-oxygen distribution, or the subsequent radiation efficacy, could be inferred from a readily measurable surrogate, such as the distribution (and density) of microvessels within these samples.

6.4 Methods for quantifying vessel organisation

To understand how the spatial organization of vessels affects the cell-oxygen distribution, and therefore radiation effect on a tissue, we consider several measures of spatial organisation, and test them each for predictive capacity.

6.4.1 Spatial autocorrelation

Several measures of spatial autocorrelation have been suggested. The first to be suggested and used widely was Moran's I , first defined in 1950 by P.A.P. Moran [165]. The form of this measure is

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}, \quad (6.9)$$

where N is the number of discretised spatial units in a lattice indexed by i and j , X is the variable of interest, in this case oxygen concentration, and $\bar{X} = \sum_i X_i / N$ is the mean of X . The matrix w_{ij} is a matrix that contains spatial weights, which take the value of 1 if the elements i and j are adjacent, and 0 otherwise. Another metric of spatial autocorrelation, Geary's C , is inversely related to Moran's I [85] and shares this adjacency based spatial weight matrix.

These measures have been useful in a number of ecological contexts where the questions are related to spatial relationships between several different species (for example) which are comprised of discrete elements. For our purposes, however, as we are interested in the relative spatial relationship between discrete points (vessels) and the associated continuous oxygen field, these metrics become less useful. As suggested by Legendre [133], the residual spatial autocorrelation between two variables (in our case cells and oxygen), and the short relative length scale on which oxygen varies as compared to vessel presence/absence, makes all landscapes appear to be well correlated, that is, have a Moran's I measure that approaches zero and cannot be used to discern between distributions.

6.4.2 A simple metric of vascular organisation

To begin to understand the vascular organisation, we introduce a new measure Φ , which we term the ‘aggregation coefficient’. We aim to define a measure that gives greater weight to pairs of nearby vessels than to those that are more distant from one another, therefore measuring ‘clumpiness’ of vessels. The simplest way to do this is to take the reciprocal of the Euclidean distance between each pair. We can then compute this number for the entire domain by taking the sum of the reciprocal of the distance between each vessel and normalizing by the total number of permutations of two-vessel pairs in the domain. Formally, we define the aggregation coefficient by

$$\Phi = \frac{1}{n(n-1)} \sum_{(a,b)} \frac{1}{\|\mathbf{x}_a - \mathbf{x}_b\|}, \quad (6.10)$$

where $\mathbf{x}_a$ and $\mathbf{x}_b$ are the positions of vessel a and b in the domain, respectively, and n is the number of vessels. So, for example, if the domain had 3 vessels adjacent to one another in a line, the aggregation coefficient would be $\Theta = (1 + 1 + 1 + 1/2 + 1/2)/6 = 5/6$, whereas the same 3 vessels in a line but with 2 spaces between each subsequent vessel would be $\Theta = (1/3 + 1/3 + 1/3 + 1/3 + 1/6 + 1/6)/6 = 5/18$, so a higher value of Θ means ‘tighter’ clustering.

To explore the relationship between this measure and carrying capacity, we first compare two cases, with high or low vascular density and differing vascular aggregation: vessels placed at random throughout the entire domain, vessels placed at random but constrained to a single quadrant, and vessels placed at random but constrained to one sixteenth of the domain. After placing the vessels, we then fill the domain with healthy cells and allow the system to come to equilibrium, defined as at least 100 time steps during which the population remains unchanged to within 1%. We find that there is a striking difference in the carrying capacity exhibited when comparing high to low aggregation coefficients. While increasing vascular density continues to monotonically increase carrying capacity, it is at a much diminished rate when the aggregation coefficient is higher (Figure 6.7). This initial analysis suggests that this simple measure would be useful in predicting carrying capacity, but as to how it predicts radiation efficacy, we can not yet be sure. We will explore this specific question in the coming section.

6.4.3 Voronoi tessellation

In the field of physiological modelling, many different methods have been used to approximate oxygen concentrations from the distribution of capillaries. Beginning with regular tilings of the plane using overlapping circular regions, or hexagonal tilings, the field has

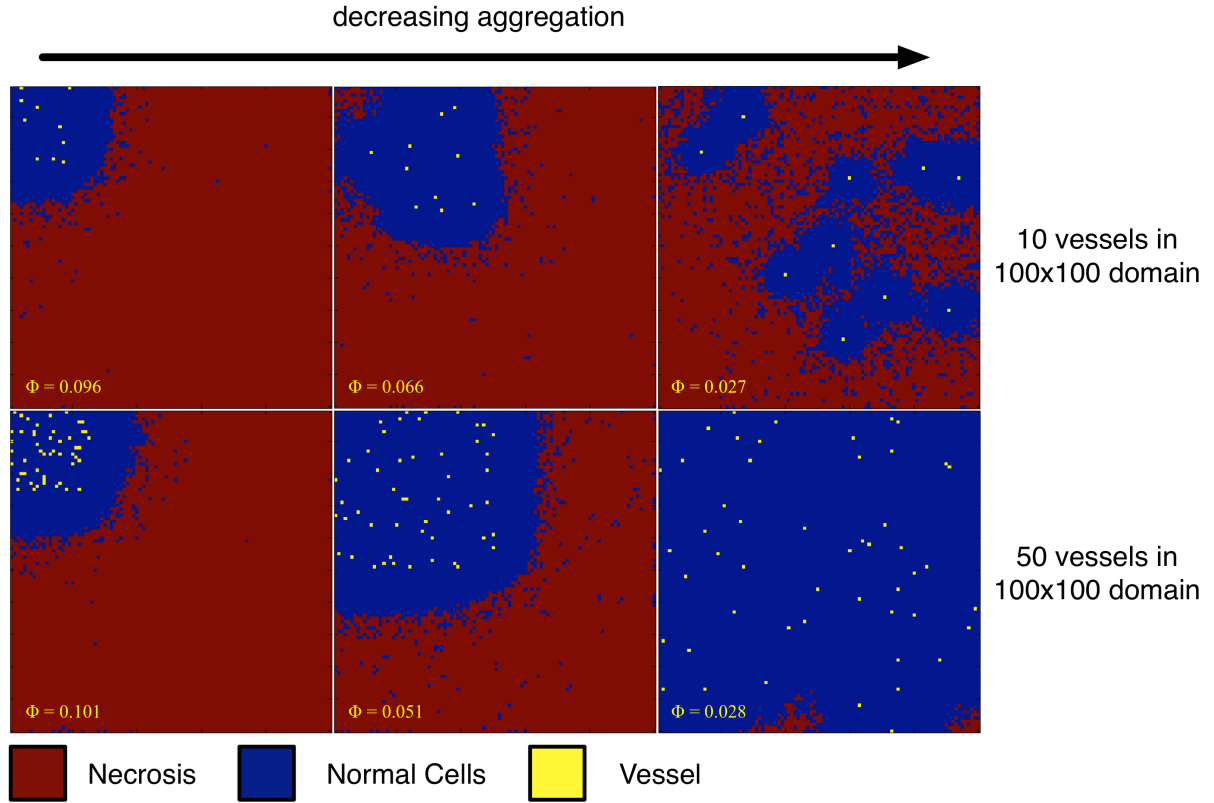


Figure 6.7: Varying the vascular aggregation affects the carrying capacity of normal tissue independently of vascular density. We plot a series of healthy tissue growth patterns with decreasing aggregation coefficient (looser clustering) from left to right and compare the effect in high (bottom row) and low (top row) vascular density. Each simulation begins with the 100×100 domain full of normal cells and is allowed to reach a dynamic equilibrium where the births and deaths are balanced (within 1% of an equilibrium value for 100 timesteps). We show the aggregation coefficient, Φ , as computed by equation (6.10) and the cellularity (cell number at equilibrium/domain area). Top left: $\Phi = 0.096$, cellularity = 0.115. Top centre: $\Phi = 0.066$, cellularity = 0.242. Top right: $\Phi = 0.027$, cellularity = 0.386. Bottom left: $\Phi = 0.101$, cellularity = 0.152. Bottom centre: $\Phi = 0.051$, cellularity = 0.410. Bottom right: $\Phi = 0.028$, cellularity = 0.987. Parameters values are listed in Table 5.1.

evolved to now using Voronoi tessellations to approximate the heterogeneity of oxygen concentration in a given tissue, or even the size of hypoxic areas (see review by Egginton and Gaffney [64]). For example, Al-Shammari and colleagues [2, 3] utilized the shape of the distribution of computed Voronoi polygon areas (VPAs) to understand which vascular patterns could support more or less muscle tissue. In particular, they found that left-skewed distributions, with ‘fat-tails’ representing large polygon areas, predicted for areas of extreme hypoxia. Because of the success of Al-Shammari and others in predicting the non-trivial consequences of heterogeneous oxygen concentrations in complex tissues using this measure, we will utilise it here. Specifically, we will use the *en face* vessels as points to generate Voronoi polygons to tessellate the plane, and then determine if the distribution of the area of the polygons formed is useful in the prediction of radiation

efficacy.

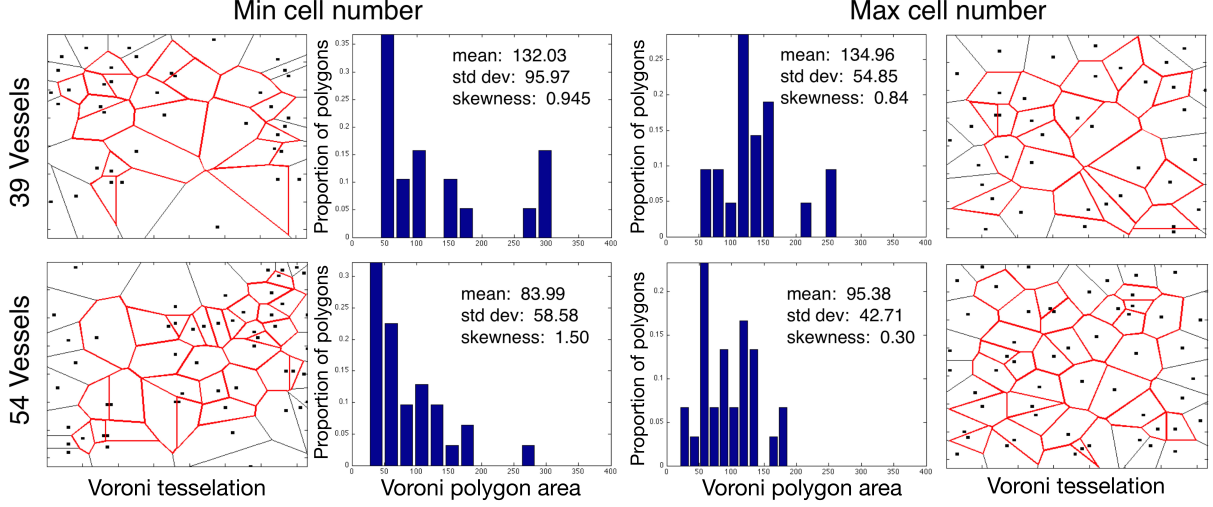


Figure 6.8: Examples of Voronoi polygon plots with VPA distributions. We plot representative samples of Voronoi polygon and VPA distributions from the maximum (right) and minimum (left) cell numbers at equilibrium for each of the 500 simulations in the 39 (top) and 54 (bottom) vessel cases. We report the statistical properties of each distribution. In each domain, we consider only the areas of polygons that are closed within the domain (outlined in red).

We define the Voronoi polygon containing the i th vessel, $\mathbf{x}_i$, as the closure of the set of points in domain D which are closer to $\mathbf{x}_i$ than to any other:

$$V_i = \{\mathbf{x} : \mathbf{x} \in D; |\mathbf{x} - \mathbf{x}_i| \leq |\mathbf{x} - \mathbf{x}_k|; i \neq k\}. \quad (6.11)$$

By this definition, we have that $\mathcal{V} = \{V_1, \dots, V_n\}$ tessellates our domain. A complication of this method in finite domains is that the polygons closest to the edge of the domain must either be chosen to be bounded by the domain edge, which does not reflect the underlying point pattern, or considered to have infinite area as they are not bounded by edges of other Voronoi polygons. To correct for this, we exclude those polygons when we consider the distribution of VPAs as seen in Figure 6.8, with included polygons highlighted by red edges, an assumption also utilised by Al-Shammari et al. [2].

In their work on modelling oxygen supply to muscle tissue, Al-Shammari et al. found that ‘fat tails’ and skewed distributions of Voronoi polygon areas correlated with areas of hypoxia [2, 3]. Because of this finding, it is reasonable to hypothesize that the statistical properties of this distribution could be predictive in our model system as well. To this end we first test two different examples of vessel number that support the most and least number of cells in a given family of simulations. Specifically, we plot the Voronoi tessellation and associated VPA distribution of the maximum and minimum cellular carrying capacity in each of the 39 and 54 vessel cases (Figure 6.8). These two examples do not

provide clear evidence to support or refute our hypothesis, but do show that the statistical properties of the VPA distributions vary from one vessel pattern to another. We will more rigorously explore the utility of this measure in the full family of simulations in the coming section.

6.4.4 Ripley's K

Another spatial measure, invented specifically to quantify the relationship between discrete points, is Ripley's K [197], and a variance stabilized version, Ripley's L . These measures, which are functions of distance, describe the average number of points within a given distance of any other point. For Ripley's K , we have

$$\hat{K}(r) = \lambda^{-1} \sum_{i \neq j} \frac{I(d_{ij} < r)}{n}, \quad (6.12)$$

where λ is the average density of n points in the domain, and I is the indicator function which is defined by

$$I(d_{ij} < r) = \begin{cases} 1 & \text{if the Euclidian distance, } d_{ij}, \text{ between vessels } i \text{ and } j \text{ is less than } r, \\ 0 & \text{otherwise.} \end{cases} \quad (6.13)$$

For perfectly homogeneously spread points, $\hat{K}(r) = \pi r^2$. We will utilize the variance stabilized version of this measure, $\hat{L}(r)$, which is given by

$$\hat{L}(r) = \frac{\hat{K}(r)^{1/2}}{\pi}, \quad (6.14)$$

which has an expected value of $\hat{L}(r) = r$ for homogeneous data with constant variance. To correct for edge effects, we implement the correction suggested by Ripley [197], which changes the value of the indicator function, for points assayed within a distance r of the edge, to the reciprocal of the proportion of the circle (of radius r) which falls outside the study area. Specifically, we consider two cases: one in which the circle falls within r of a single boundary (left, Figure 6.9), and one in which the circle falls within r of two boundaries (Figure 6.9, lower right). In the former case, we find the correction factor, w , to be

$$w = \left(1 - \frac{2\theta}{2\pi}\right)^{-1}, \quad (6.15)$$

where $\theta \in (0, \pi/2) = \arccos(d/r)$ and d is the distance from the centre of the circle to the boundary. This takes its maximum value when the centre point lies infinitely close to

the boundary such that $\theta \rightarrow \pi/2$, and $w \rightarrow 2$. In the latter case, we find the correction factor to be

$$w = \left(1 - \frac{\theta + \phi + \pi/2}{2\pi}\right)^{-1}, \quad (6.16)$$

where $\theta \in (0, \pi/2) = \arccos(d/r)$ is the angle subtended by the shortest line segment (of length d) connecting one boundary and the circle's centre, $\phi \in (0, \pi/2) = \arccos(p/r)$ is another angle subtended by the shortest line segment (of length p) connecting the other boundary and the circle's centre. This takes its maximum value when the centre point lies infinitely close to a corner of the domain and $\theta \rightarrow \pi/2$, $\phi \rightarrow \pi/2$ and $w \rightarrow 4$. Utilising the boundary correction, we modify the values measured for these three circles of the same radius encircling identical clouds of points from a range of 4 – 7 to 6.75 – 7.5, significantly improving the measure.

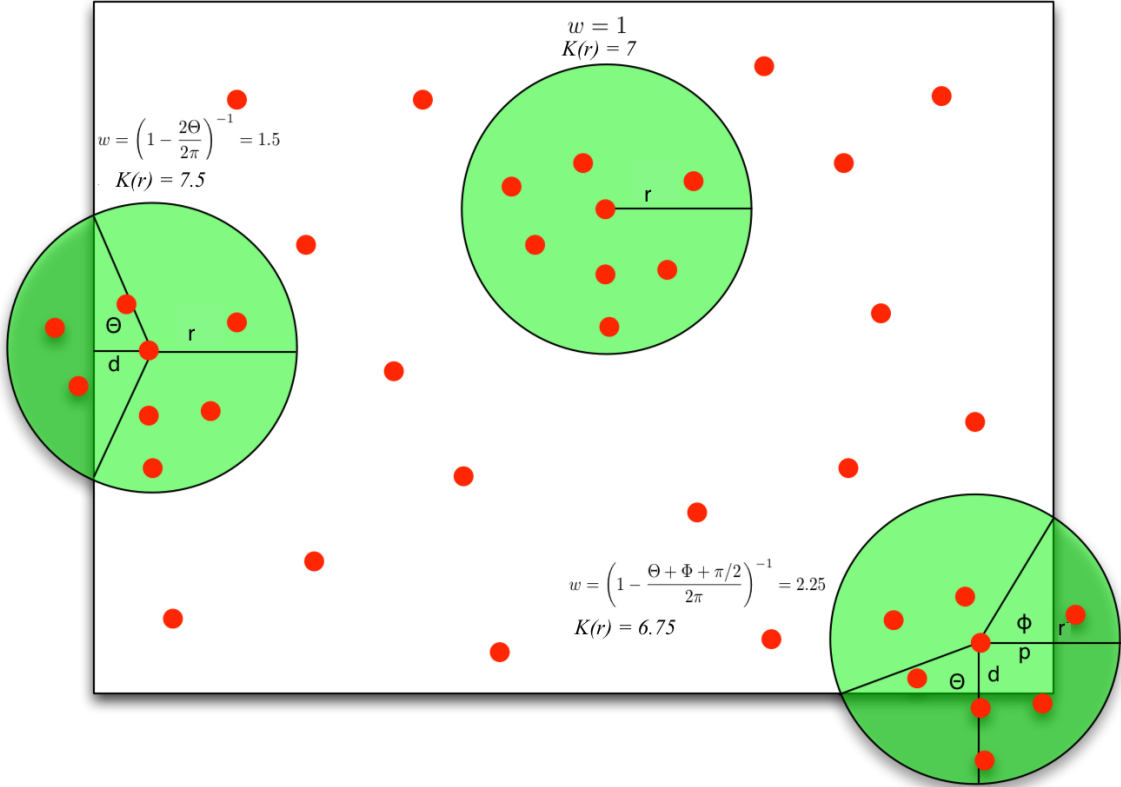


Figure 6.9: Ripley's K edge correction. To correct Ripley's K when points lie near the domain edge, we weight each point within a distance r of the point of interest by the inverse of the circumference of the portion of the circle of radius r that lies outside the domain. To demonstrate, we plot three identical circles of radius r enclosing the same number and pattern of points, but at different distances from the edge, causing some points to fall outside the domain. We calculate the correction factor, w , and value of Ripley's K for each and find without edge correction the value of the measure varies from 4 to 7, while with correction it varies only from 6.75 to 7.5.

To ascertain the value of Ripley's L in discriminating between families of simulations, as we have done for the previous measures we compare, in Figure 6.10, two examples of the measure on different vascular arrangements. When we compare the maximum and minimum supported populations for a low (24 vessel) and moderate (39 vessel) vessel simulation, we see a strong separation between the curves. We will more systematically evaluate the predictive ability of this measure for radiation response in the next section, and compare it with that of the other spatial metrics reviewed previously.

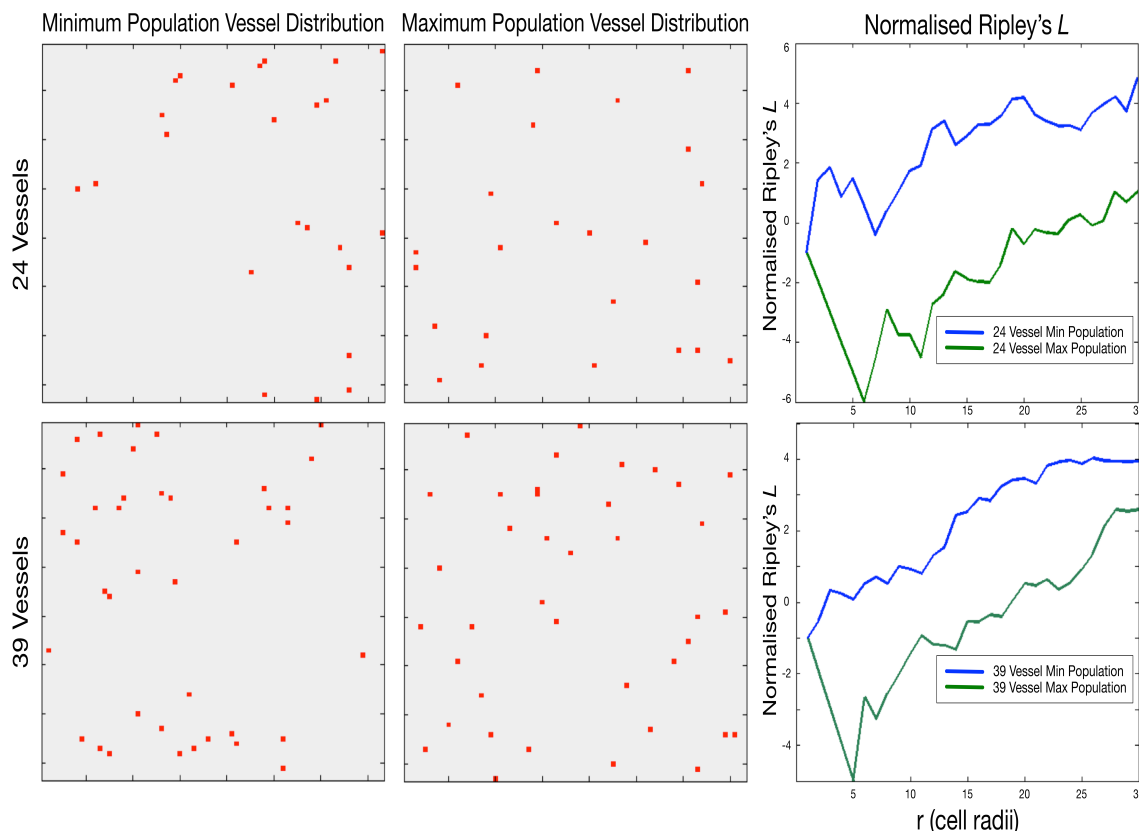


Figure 6.10: Examples of Ripley's L for specific vessel distributions. Vessel distributions and value of Ripley's L for the maximum and minimum populations simulated in two different vessel number simulation families from Chapter 5. We see a strong separation, suggesting that the metric could be useful as a predictor for carrying capacity.

6.5 Using spatial metrics to predict radiation effect

In Section 6.4, we presented three spatial measures: the aggregation coefficient, the VPA distribution and Ripley's L . We found that each had some potential for predicting at least carrying capacity (see Figures D.1, D.3 and D.4), but as this is only a part of the information needed, and indeed can be estimated from modern MRI technology, we must more specifically assay for correlation with survival after simulated radiation therapy. To this end we will, for each of the 500 simulations in the differing cases of vessel number,

compute each of the measures presented in the previous section and analyse their effect on surviving cells after radiation. We will present these raw data in scatter plots across the values of vessel number and then present their correlation coefficients in aggregate, and compare them head to head.

To calculate the number of surviving cells in a given domain after simulated radiation, we first calculate the individual values of α_i and β_i for each interval of oxygen concentration using equations (6.4) - (6.5). Once we have these oxygen specific values, we can then compute the surviving number of cells in the domain experiencing oxygen within a given range using equation (6.8). We take as input the number of cells and associated cellular oxygen distribution at dynamic equilibrium from each of the simulations represented in Figure 5.10 and, assuming that there is a single dose ($n = 1$) of radiation given in a fraction size of $d = 2\text{Gy}$, compute the surviving number of cells using equation (6.8).

6.5.1 Aggregation coefficient

Utilising the method for calculating the aggregation coefficient presented in Section 6.4.2, we plot the surviving number of cells in each domain against the aggregation coefficient for each of the 500 simulations. In Figure 6.11 we present scatter plots of the relationship between the aggregation coefficient and surviving cells after radiation in 12 representative families of simulations with increasing vessel numbers. Here, the axes are scaled to the data for individual subplots for ease of visualisation. We see that for vessel numbers below 78 there is a negative correlation between the aggregation coefficient of the domain and the number of surviving cells. This metric loses its ability to predict survival around 78 vessels and then the correlation becomes positive with a larger aggregation coefficient predicting more surviving cells in the larger vessel numbers. In all situations, except for the 78 vessel case, the correlation is significant ($p < 0.05$), which we show in detail in Figure 6.14.

6.5.2 Voronoi polygon area distribution

Utilising the method for calculating the Voronoi polygon area distribution presented in Section 6.4.3, we briefly examine which moment of the distribution is most correlated with carrying capacity. We find that while all of the moments correlate, the skewness of the VPA distribution is the strongest (see Figure D.2). Accordingly, we plot it against the surviving number of cells after the simulated radiation in each domain for each of the 500 simulations. Here, the axes are scaled to the data for individual subplots for ease of visualisation. In Figure 6.12 we present scatter plots of the relationship between the VPA skewness and surviving cells after radiation in 12 representative families of simulations

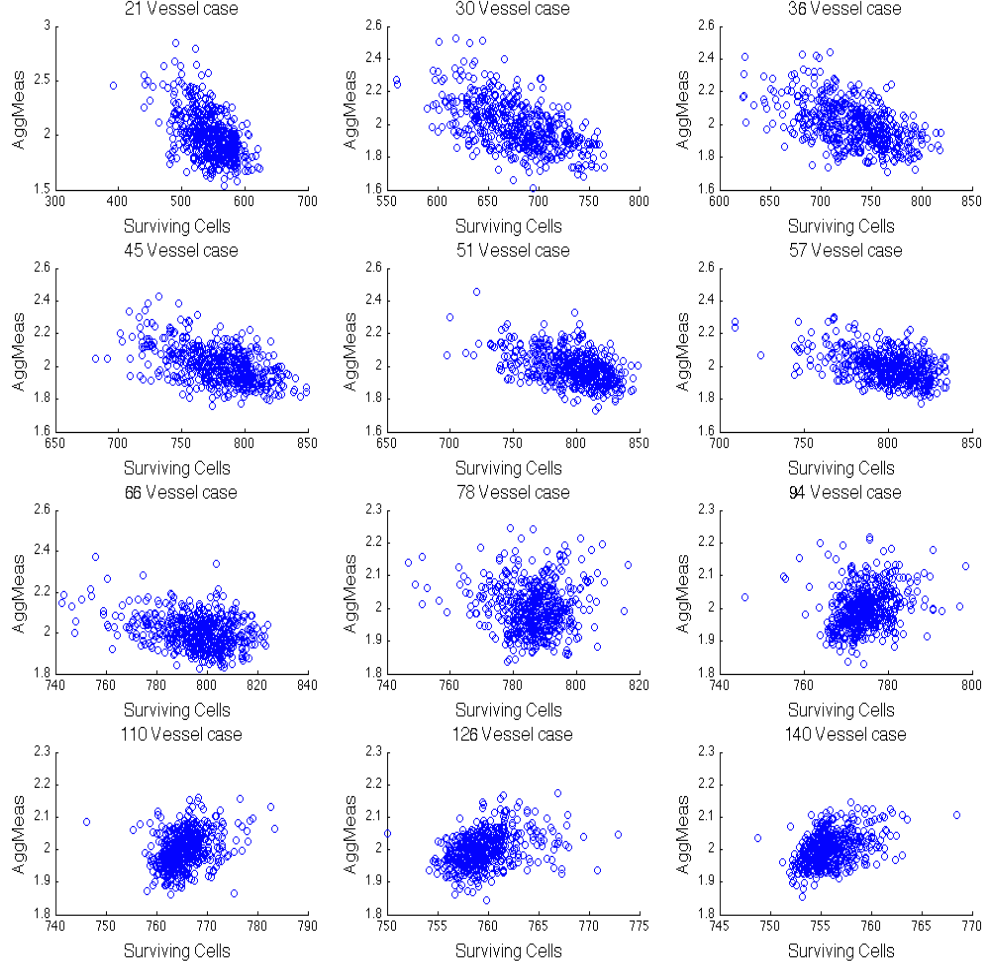


Figure 6.11: Aggregation coefficient versus surviving cells after radiation. We show twelve scatter plots illustrating the relationship in each of the 500 simulations represented in Figure 5.10 for a given initial vessel density between cell number surviving after 2Gy of radiation (x -axis) and the Aggregation Measure (y -axis). We find that there is a negative correlation in the low vessel densities, and a positive correlation in the high vessel densities. Axes are scaled to the individual subplot data for ease of visualisation.

with increasing vessel numbers. Similar to the aggregation measure, we find a negative correlation in the low vessel densities which changes to a positive one at higher densities, however the correlation is weaker here (correlations fully described in Figure 6.14).

6.5.3 Ripley's L

To utilise this same method for correlating Ripley's L with surviving cells requires a small simplification. Each of the other measures was given by a single number, but the Ripley's L is a function of distance. To make these comparable, we take the mean value of Ripley's L from $r = 0$ to $r = 20$, which is approximately the distance after which the two vessels in question no longer affect each others oxygen distributions, calculated as presented in Section 6.4.4. The results of this are plotted against the surviving number of

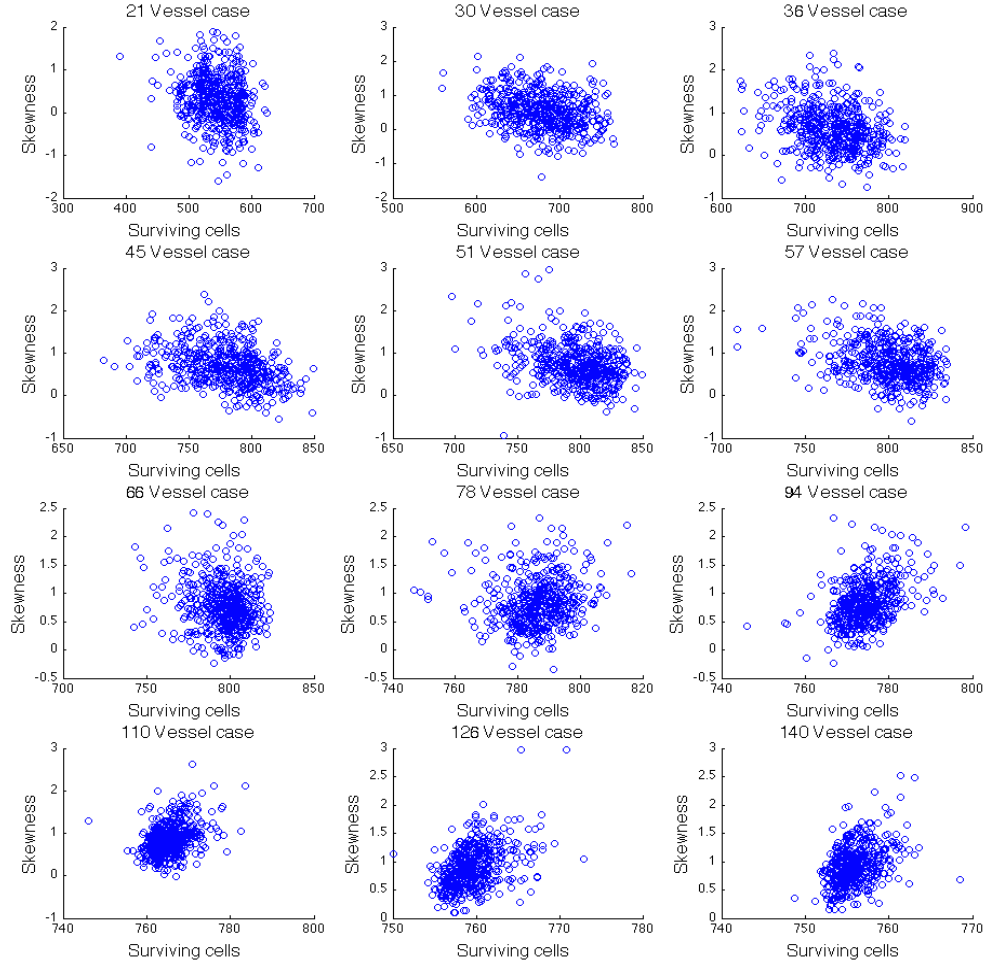


Figure 6.12: Voronoi polygon area distribution skewness versus surviving cells after radiation. We show twelve scatter plots illustrating the relationship in each of the 500 simulations represented in Figure 5.10 for each listed initial vessel density between the skewness of the Voronoi Polygon area and the cells surviving after 2Gy of simulated radiation. We find that there is a positive correlation in the low vessel densities, and a negative correlation in the high vessel densities. Axes are scaled to the individual subplot data for ease of visualisation.

cells in each domain for each of the 500 simulations. Here, the axes are scaled to the data for individual subplots for ease of visualisation. In Figure 6.13 we present scatter plots of the relationship between the aggregation measure and surviving cells after radiation in 12 representative families of simulations with increasing vessel numbers. Again, we see a negative correlation in the low vessel densities which changes to a positive correlation in the high vessel densities.

6.5.4 Comparison across metrics

Each of the spatial measures that we tested were able to predict, to some degree, the number of surviving cells in a given domain after radiation. Each of the metrics also had the same changing relationship: a negative correlation at low vessel number and a positive

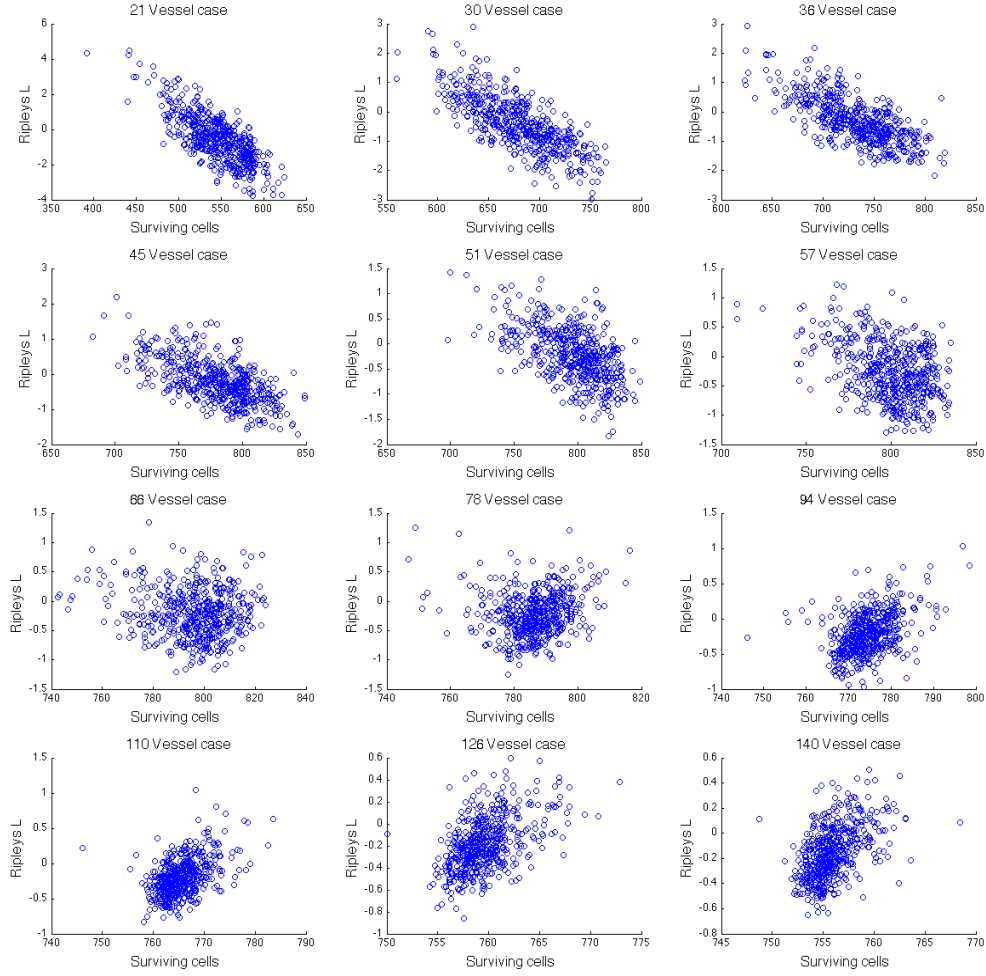


Figure 6.13: Ripley’s L versus surviving cells after radiation. We show twelve scatter plots illustrating the relationship in each of the 500 simulations represented in Figure 5.10 for a given initial vessel density between cell number surviving after 2Gy of radiation (x-axis) and the mean of Ripley’s L (y-axis). We find that there is a negative correlation in the low vessel densities, and a positive correlation in the high vessel densities. Axes are scaled to the individual subplot data for ease of visualisation.

correlation at higher vessel number. Additionally, each passed through ‘no correlation’ at approximately the same vessel density. To compare the relative value of these metrics, we plot their correlation coefficients and the associated p-values. While each of the metrics has the same general shape to the correlation versus vessel number plot (Figure 6.14, left), and are significant ($p < 0.05$) for all but one vessel number (Figure 6.14, right), the Ripley’s L measure outperforms the others.

6.6 Conclusions

In this chapter we have seen that the assumption that mean vessel density, and subsequently mean oxygenation, can be used as a surrogate for the number of surviving cells

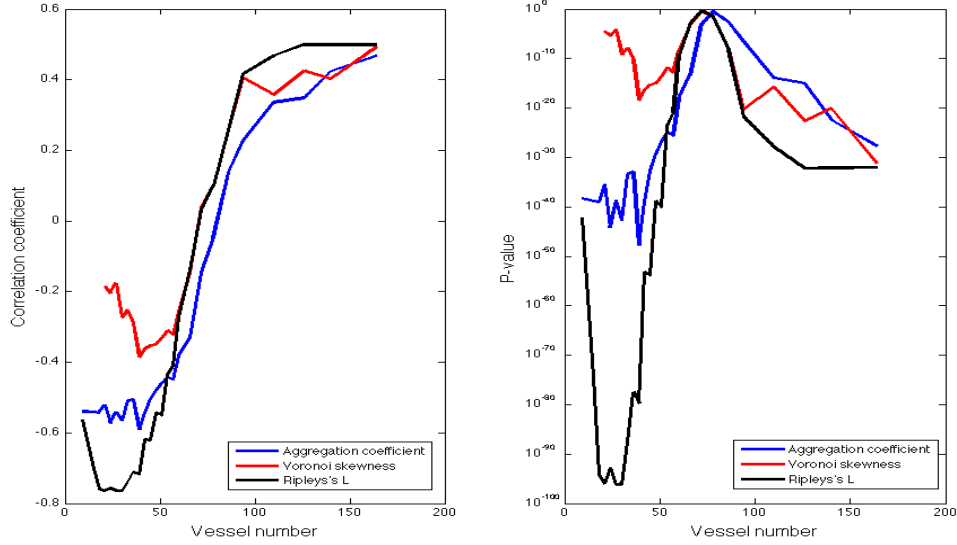


Figure 6.14: Comparison of the correlation of the three measures of vascular patterning with radiation effect. We show the correlation coefficient versus vessel number for each of the measures of vessel organisation against surviving cell number (left) with the associated p-values (right) for each of Ripley’s L (black), the VPA distribution skewness (red) and the aggregation measure (blue). While all three measures share the same qualitative shape and regions of statistical significance, the Ripley’s L measure performs the best.

after radiation is insufficient for understanding radiation efficacy. Further, we found that the relationship between each of our vessel pattern measures and the surviving cells after radiation exhibits a sign change in the mid-vessel density range. This change of sign in correlation means that the patterns, within a given vessel density, that take the extreme values of each measure of vascular organisation can represent opposite extremes of radiation response from one another. This suggests that if one were to perturb the vascular organisation, for example toward a more homogeneously distributed with a vascular normalising therapy, one could induce opposite effects on radiation response, depending on the vessel density.

6.6.1 Effect of vessel normalisation on radiation response

With Folkman’s discovery, in 1971, of a master tumour angiogenesis regulating factor [75], the world thought that the suggested method of blocking this factor, by which it was promised that we could ‘starve’ tumours of their oxygen supply, would dominate cancer research. It was thought that a cure for cancer would come about in a short time period. However, early trials of single agent anti-angiogenic drugs failed to show efficacy [116]. Later trials, with combinations of chemotherapy and anti-angiogenic drugs, however, showed promise, but even this was discordant with the leading hypothesis describing the

mechanism of anti-angiogenic therapy, which was thought to entirely starve tumours of blood supply.

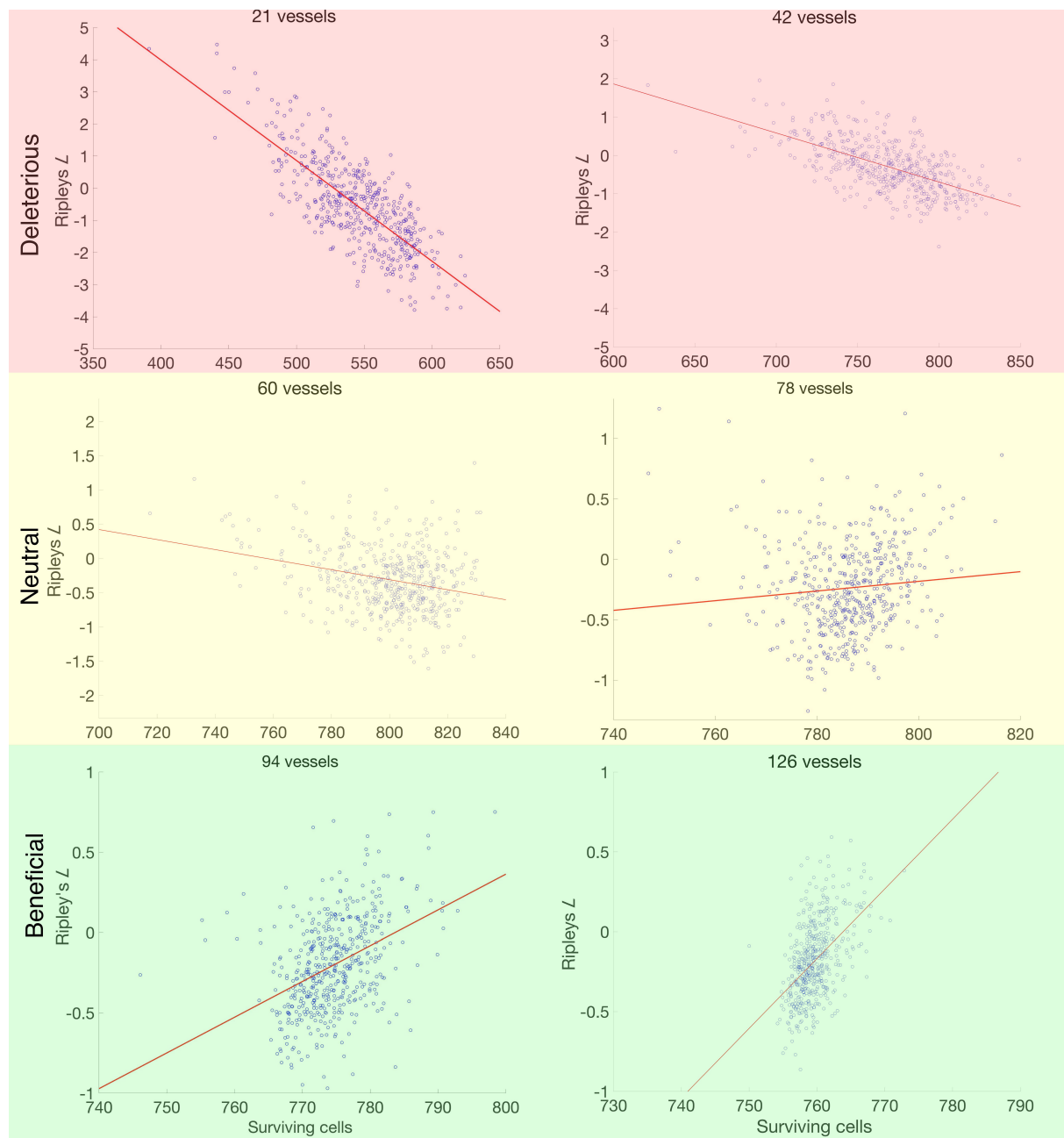


Figure 6.15: Normalising vascular patterning has opposite effects on radiation efficacy in low versus high vascular density. We plot six scatter plots showing the relationship in each of the 500 simulations represented in Fig. 5.10 for a given initial vessel density between cell number surviving after 2 Gy of radiation (x -axis) and the mean of Ripley's L function (y -axis). We find that there is a positive correlation for the low vessel densities, and a negative correlation for the high vessel densities. All correlations are significant ($p < 0.05$, see Figure 6.14). Axes ranges are normalised across all data presented.

It was not until 2001, when Jain suggested the ‘vascular normalisation hypothesis’ [113], that these results could be understood under a single rubric. Jain suggested that

anti-angiogenic drugs, instead of entirely blocking new vessel formation, worked to normalise vasculature, pruning inefficient vessels and creating a more regular lattice, thereby improving drug and oxygen delivery, the effects of which on chemotherapeutic response have been modelled in a theoretical model system [7]. More recent iterations of this hypothesis also include improvement of the efficiency of existing vessels (reviewed by Jain [114, 115]).

While this advance in our thinking has provided a way to explain the counter-intuitive results of many trials, it still does not explain why these combinations of anti-angiogenic drugs with chemotherapeutics (or radiation therapy) does not help all patients. Using our simple model system, we have observed a changing correlation with spatial measures and radiation response. This suggests that in certain cases, ‘improving’ the homogeneity of vascularisation would diminish the radiation effect, whilst in other cases it would help: a heterogeneous response to normalisation of vascular patterning. We show that for certain cases, with vessel density held constant, normal vascular patterning can respond either better or worse to radiation therapy (Figure 6.15). This raises the possibility of using this knowledge to predict in which cases (patients) vascular normalisation therapy (VNT, e.g. bevacizumab) would be of use before radiation therapy, and for which patients it would not. Alternatively, this suggests the possible utility of partial radiation treatment (to the areas where normalisation would hinder radiation) given before VNT, followed by radiation to the remainder of the tumour after vascular normalisation.

6.6.2 Future directions: translation to the clinic

To translate these conclusions to the clinic would first require biological validation of these hypotheses, as well as overcoming a number of model limitations. Specifically, the assumption that all vessels are *en face* and all vessels are the same in terms of size and efficiency. Validation could be achieved either through *in vivo* window chamber experiments, or indirectly through examination of post-radiation surgical specimens. Model development, to include the addition of in-plane vessels, angiogenesis, more realistic geometry and vessel heterogeneity, could be used to extend these predictions to more realistic vasculatures.

If these limitations were addressed, and validation were achieved, the technology exists currently to take advantage of this new idea. Specifically, macroscopic oxygen concentrations could be inferred from DCE MRI (or other advanced imaging) to create optimised dose plans, as suggested by Malinen et al. [150]. This information on putative vessel density could then be coupled with histologic measurements using automated localisation of vessels [140] and an algorithm to calculate Ripley’s K . This combination of information would then allow for determination of whether radiation would be best

given before or after VNT. These two pieces of information could then be incorporated by creating a temporally and a spatially distinct radiation plan whereby the appropriate radiation dose would be delivered before VNT to the area that would suffer after normalisation, and then the final radiation could be delivered after VNT. This concept, which would represent both a biologically and physically optimised treatment plan, is schematised in Figure 6.16.

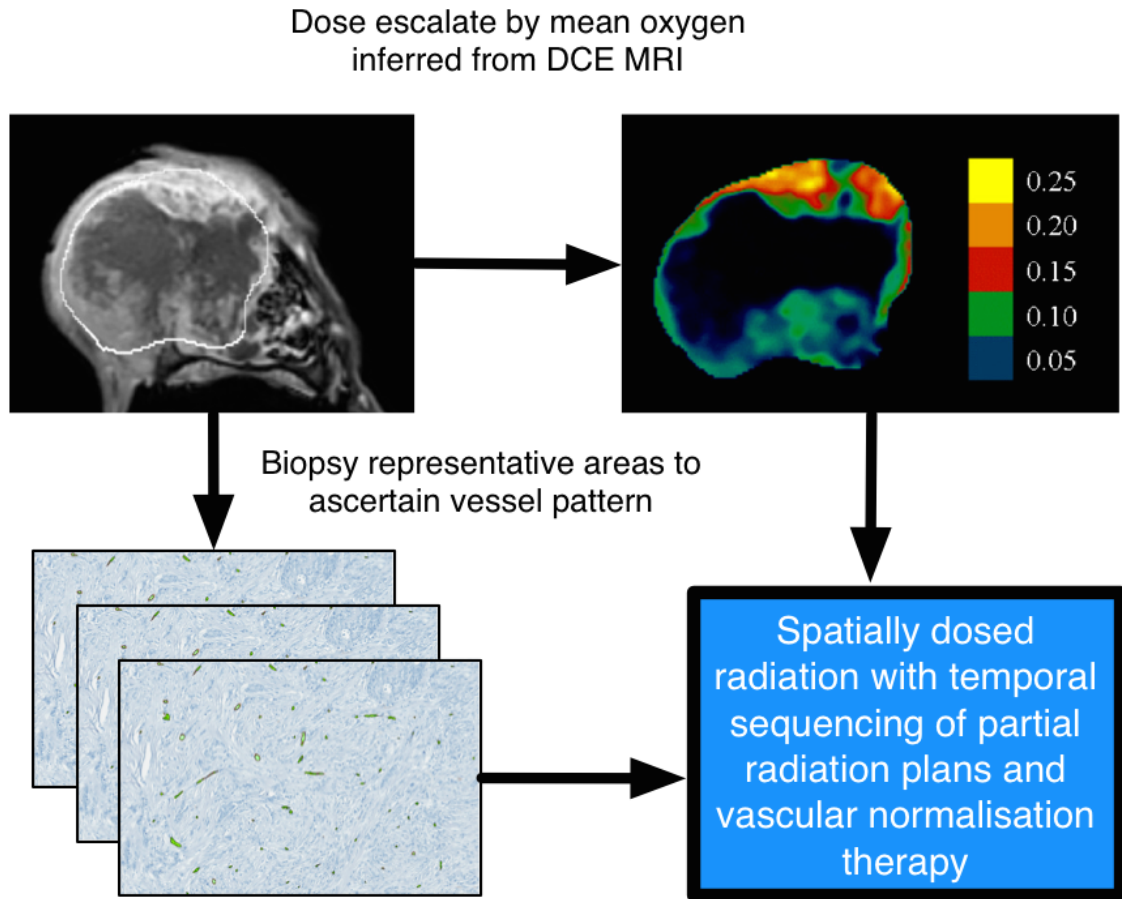


Figure 6.16: Schematic of a method by which spatially and temporally optimise radiation therapy treatment and VNT scheduling. Tumours can be segmented by advanced imaging such as DCE MRI to understand macroscopic oxygen tension (upper left) and segmented to create optimal dose plans (upper right) as suggested by Malinen et al. [150]. Secondly, biopsies can be taken from representative areas of the tumour to determine vascular patterning, and a determination could be made as to whether a given area should be treated before or after VNT.

Chapter 7

Conclusions and future work

In this thesis, we have gained insight into several clinically relevant questions about tumour growth by building and analysing theoretical models of simplified representations of cancer growth and treatment. We attacked these questions in two parts. In Part 1, which is comprised of Chapters 3 and 4, we worked to determine if we could use any clinically available surrogate metrics to understand the symmetric division probability of a classic, cancer stem cell-driven tumour. And then, in Part 2, which is comprised of Chapters 5 and 6, we worked to understand the relationship between spatial oxygen distributions, vascular patterning, tumour growth and radiation response.

7.1 Part 1: Conclusions and future directions

7.1.1 Diversity, colony size and shape, and imperfect information: comparisons of models across symmetric division probability

To approach the problem of inferring symmetric division probability, we built a pair of models of a growing tumour under a cancer stem cell hierarchy with neutral evolution. We then compared the two models, one spatially constrained by contact inhibition on a two-dimensional lattice, and one unconstrained, using a number of different measures. In our comparisons, we found that the two models behaved in opposite manners in terms of changes in diversity with increasing symmetric division probability. Specifically, we found that the measured Shannon index of the non-spatial model decreased monotonically with increasing symmetric division probability. Further, the results of Shannon index measurements were tightly clustered within a given family of simulations (for a specific symmetric division probability) and the families were all statistically different from one another – allowing us to reliably discriminate between them. This phenomenon was not found to be conserved in the spatial model, where instead, if anything, the opposite

occurred. While the differences were only significant for the highest symmetric division probabilities, the trend was an increasing one: higher symmetric division rates had higher measures of diversity. Disappointingly, we were unable to reliably discriminate between families of differing symmetric division probability, with the exception of being able to say whether or not a given simulation was stem-driven or not, i.e. had symmetric division probability, $\alpha = 1$, or $\alpha < 1$.

Beyond measures of diversity, we also observed differences in the colony size distributions between the two models, but they were more subtle. All the observed relationships between symmetric division probability trends were the same, but less noticeable in the spatial model, where differences were rarely significant. These results are suggestive of a simpler method of inference which could be tested against more experimentally difficult ones using the same data, like that of Driessens et al. [61] studying *in vivo* lineage tracing, but one that would require stains for a number of aberrations in a single sample. Further, as we mentioned in Chapter 3, data from CRISPR/Cas9 gene editing studies are beginning to make colony size distributions available, though this is clearly only for well controlled laboratory systems.

Next, we asked how many of these observations could be replicated using ‘imperfect’ information. To simulate this imperfect information, we generated an algorithm to take ‘biopsies’ from our simulation outputs, which we modelled as circular, non-overlapping samples of a given size, taken uniformly at random from the domain without replacement. We then tested how accurately one could approximate the total diversity of simulations using symmetric division probabilities of $\alpha = 0.2$ and $\alpha = 1$. While the accuracy was quite poor, and we were unable to discern the difference between the simulations, we found that the relative accuracy of measurement changed in favour of non-TIC-driven tumours. A more thorough exploration of the parameter space also suggested that more biopsies are more valuable than larger ones, when compared holding the total number of cells gathered constant.

7.1.2 Inferring symmetric division probability from phylogenetic measures

In Chapter 4, we worked to remedy our inability to predict symmetric division probability in our spatial model of a stem driven tumour with neutral evolution. To do this, we developed an algorithm to reconstruct the phylogenetic tree for a given simulation. We next compared a number of statistical measures of the resulting tree topologies across changing symmetric division probability and found several measures which were able to discriminate between the families of simulations. Further, we found that these measures were stable in their ability to discriminate between simulations with differing symmetric

division probability through perturbations in mutation rate and tumour size. We secondarily attempted to discriminate between families of simulations based on reconstructed trees from ‘biopsies’ of the simulations without success.

7.1.3 Future work

Secondary to the observations in Chapter 3, there are two opportunities to increase our knowledge using clinical samples. First is an opportunity to discriminate, in solid tumours, between stem driven (symmetric division rate, $\alpha = 1$) and non-stem driven (i.e. $\alpha < 1$) tumours; and second, to understand relative symmetric division probabilities for liquid tumours (not spatially constrained) like leukemias and to correlate these with clinical outcomes. These findings could be tested clinically using methods similar to Martinez et al. [151] to measure diversity in a variety of tumour types, both solid and liquid. Determining the ‘ground truth’ for symmetric division rate would be more difficult, but surrogate measures, like surface markers or sphere forming assays could be used.

While it was encouraging to see in Chapter 4 that measures of reconstructed phylogenetic trees enabled us to discriminate between underlying symmetric division probability in our spatial model, the approach is not yet ready to be translated to clinical practice as we are not, *ab initio*, able to infer symmetric division probability unless all other parameters are known. We are, however, able to compare two simulations and determine relative symmetric division probabilities. To understand the reason for this, and the next step, let us consider Bayes’ theorem:

$$P(\mathbf{S}|\mathbf{I}) = \frac{P(\mathbf{S})P(\mathbf{I}|\mathbf{S})}{P(\mathbf{I})}. \quad (7.1)$$

Here, we represent the conditional probability of having a true symmetric division probability ($\mathbf{S}$) given a measure of a phylogenetic tree index ($\mathbf{I}$). In Chapter 4, we generated the prior conditional probability distribution of tree indices given symmetric division rate, $P(\mathbf{I}|\mathbf{S})$. To be able to directly infer the posterior probability we desire, we need to know more about the distribution of tree indices, $P(\mathbf{I})$, and symmetric division rates, $P(\mathbf{S})$, in real cancers. This information has yet to be described, but we can still observe that $P(\mathbf{S}|\mathbf{I}) \propto P(\mathbf{I}|\mathbf{S})$. Therefore, having the prior distributions we have generated does allow us to compare two tumours, or two samples of the same tumour, through time. While not a perfect answer to our question, this at least opens several avenues for research. In particular, this information would allow us to study the change in symmetric division rate of a tumour before and after a given therapy, and also methods of stratifying patients based on relative symmetric division rate for clinical trials of novel agents.

Another meaningful extension to this work would be to consider the effects of non-neutral evolution on these phylogenetic measures. While selection has been modelled before, in particular in HIV and other viruses [95] (Figure 7.1, upper panel), what happens to topologic measures in the setting of positive selection in a spatial tumour model, has not. Further, several different methods of modelling non-neutral evolution have been posited, and which is most correct for a given cancer, is *terra incognita*. McFarland et al. [155, 156] considered a ‘passenger-driver’ framework in which most mutations were nearly neutral, but actually slightly deleterious (passengers), while a very small proportion of mutations were strongly beneficial (drivers). They specifically modelled the fitness of a given cell using multiplicative epistasis where

$$B(d, p) = \frac{(1 + s_d)^d}{(1 + s_p)^p}, \quad (7.2)$$

where $B(d, p)$ is the birth rate and s_d and s_p are parameters describing the fitness effects of the d driver, and p passenger, mutations. Alternative methods of assigning fitness include driver only paradigms (where every mutation provides a benefit to fitness) that can be modelled as multiplicative, or simply additive, epistasis. Whatever the method of calculating fitness, these methods can allow, as Williams et al. have done, to observe the effects of selection on a surrogate measure of population level genetics, in their case variant allele frequency [247], and then using these ‘signatures’ (see Figure 7.1, lower panel), infer the selective advantage of clones in real data. Beyond simply finding an alternative signature of positive selection, this extension of our model could be used to tease apart differing models of positive selection, which in itself could be a useful step forward, and determine how spatial constraints effect these results.

7.2 Part 2: Conclusions and future directions

In the second part of the thesis, we worked toward a translatable method of predicting radiation response using only information gleaned from modern radiologic imaging and histologic analysis of biopsy specimens.

7.2.1 Understanding tumour progression and cell-oxygen distributions using a hybrid cellular automaton model with vasculature

To approach this problem, we first built and analysed, in Chapter 5, a hybrid cellular automaton model of tumour growth in a vascularised domain. After carefully parameterising the model we quantified relationships between vascular density and dynamic

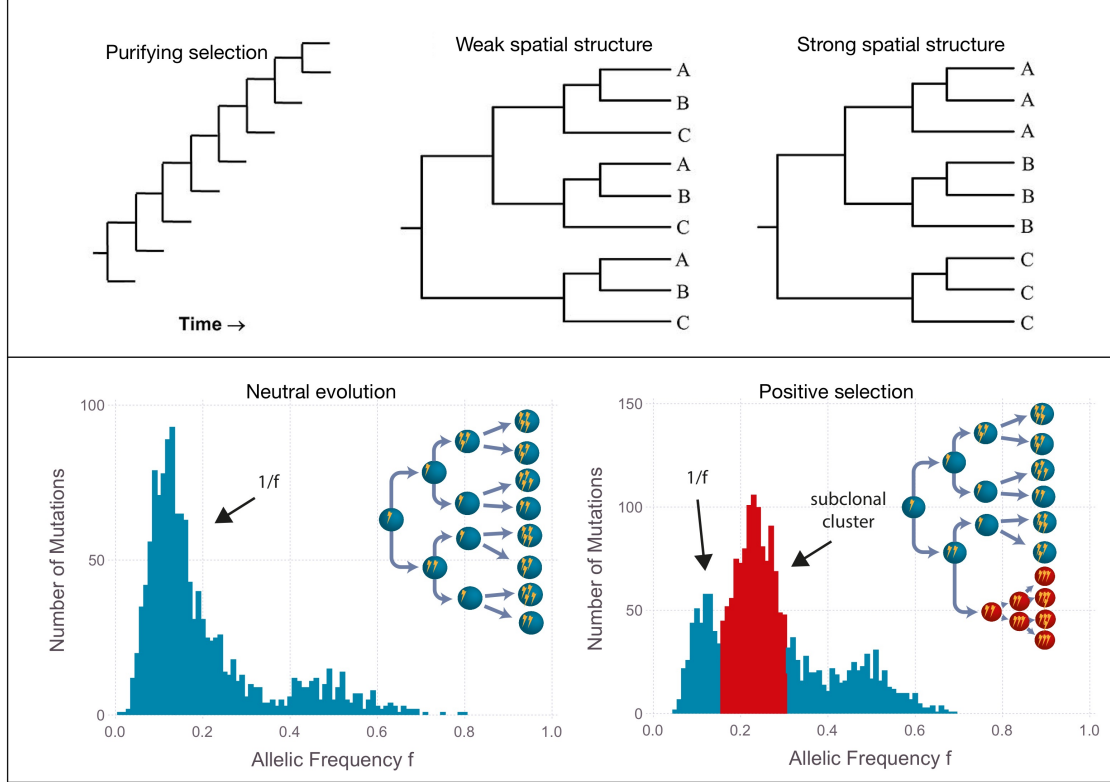


Figure 7.1: Signatures of positive selection can be seen in bulk sequencing data and phylogenies. Idealised tree shapes under immune pressure for viruses within host have been postulated, with and without spatial and population effects (upper panel). Using a combination of mathematical modelling and genomic data, Williams et al. [247] found signatures of positive selection, and were able to calculate the fitness advantage and time of inception of fit subclones (lower panel). Figure adapted from Williams et al. [247] and Grenfell et al. [95].

regimes of tumour growth in regularly organised vascular architectures. We next defined and described the cell-oxygen distribution, which is a summary representation of a population of cells that describes how many cells experience a given oxygen concentration. We found that with regular vascular organisation the mean oxygen concentration experienced by the cells would change with vascular density as expected, but the higher moments of these distributions were conserved. With irregular vascular organisation however, while the mean continued to change predictably, we found that the higher moments of the cell-oxygen distributions could vary widely – strongly influencing both the ability of the simulated tissue to support cells, and other oxygen dependent phenomena, such as radiation therapy.

7.2.2 Predicting radiation response from spatial measures of vascular architecture

To better understand the effect of this variation on radiation response in Chapter 6, we computed the surviving number of cells after a simulated dose of radiation using the oxygen enhancement ratio calculated for each individual cell. We next calculated a number of measures of vascular organisation including Ripley’s K , the Voronoi polygon area distribution and one we invented ourselves, which we termed the aggregation coefficient. We then correlated these measures with radiation response for a large family of stochastic simulations and found the same pattern for each measure: at low vascular density, vessel patterns that were highly ‘clumped’ responded better to radiation, but the opposite was true of high vessel density simulations. While all the measures showed this qualitative pattern, we found that Ripley’s K performed the best in terms of predicting radiation response.

These correlations suggested a novel explanation for the failure of vascular normalisation therapies, like bevacizumab, together with radiation therapy. Specifically, depending on vascular density, normalising the vasculature (making it more regularly organised) which is the putative effect of these therapies, *could improve or worsen radiation efficacy* – making clinical trial results difficult to interpret. Worse is that within a single tumour, because of spatial heterogeneity, it may be that we are improving radiation response in one part of the tumour while decreasing it in another. A possible solution to this is rationally sequencing and targeting radiation. With improving technique, we are now able to treat some areas of a tumour and leave others untreated. This opens the door to the possibility of radiating one part of the tumour (the part that would be made less radiation responsive by VNT), before giving that VNT, and treating the other part, which would be made more radiation sensitive by VNT, afterwards.

7.2.3 Future work

Since publication of the second part of the thesis in *PLoS Computational Biology* [212], work done by Grogan et al. [98] has corroborated our findings in two spatial dimensions, but found that when three spatial dimensions are considered, radiation response is less sensitive to vascular architecture. What is still unknown though, is how other biological phenomena react to changes in vascular architecture. While it is possible than any results in two-dimensions could be different to those in three, it is worth considering several extensions in the two-dimensional model. Specifically, how would a hierarchically organised tumour behave with heterogeneous oxygen fields? Would vascular organisation affect the distribution of TICs vs TACs? Could the well established ‘peri-vascular’ and

‘peri-necrotic’ stem cell niches be recapitulated in a simplified model? If so, what could we learn about the different dynamics present in each?

A preliminary simulation suggests that the dynamics between a peri-vascular and peri-necrotic niche could be very different (Figure 7.2). Here, we show an illustrative result using an earlier version of the model system presented in Chapter 5 with the addition of the proliferative hierarchy presented in Chapter 3 without neutral mutation. While this represents only an example simulation, and not a rigorous analysis, we can see that the population dynamics of the stem cells differ significantly between the two ‘niches’: in the peri-necrotic niche there is significant turnover, suggesting the possibility of increased rate of evolution. This raises the question: could the peri-necrotic niche be a source of novel phenotypes?

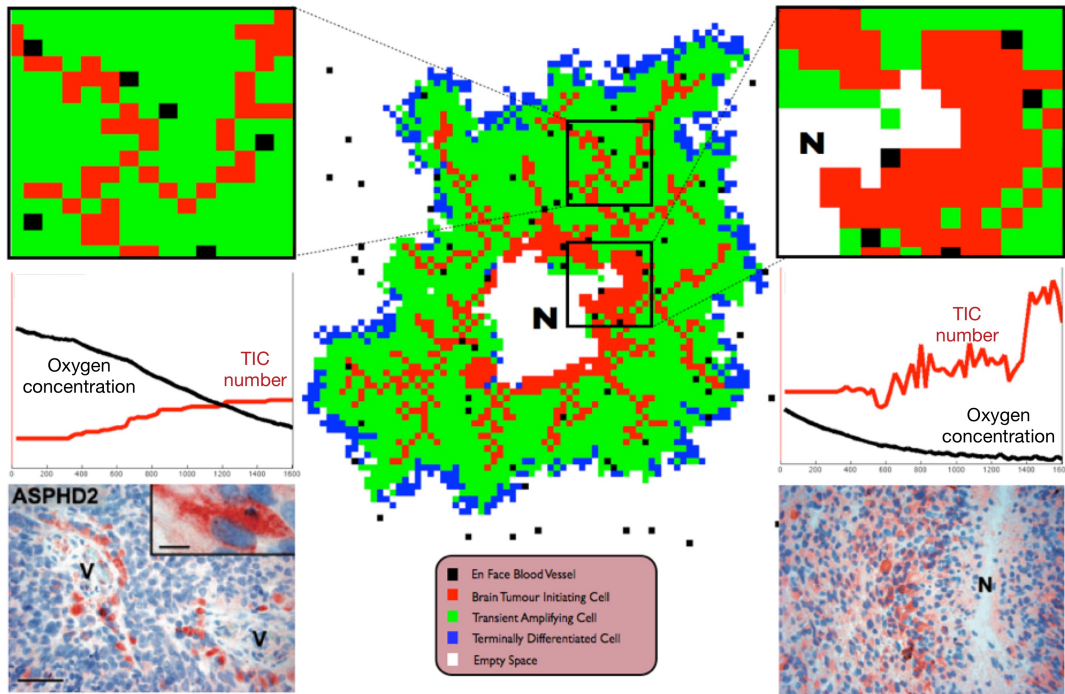


Figure 7.2: A simple model of a hierarchical tumour with irregular vasculature shows differing TIC dynamics in different TIC niches within the same tumour. While it is known that TICs inhabit several specific niches within tumours, how the dynamics differ, and indeed whether or not they differ, is an open question. Our initial simulation results suggest that the peri-necrotic niche (right) could harbour both more birth and more death than the peri-vascular niche (left), setting up a situation in which the rate of evolution could be accelerated.

7.3 Concluding remarks

The work presented in this thesis further represents a series of open doors, rather than final answers. For these open doors to be useful to patient care however, we need to work

to connect the theory here to data in real tumours. We have already discussed a fair number of the hurdles set up between final translation and the existing theory, but it is worth discussing some of them again here. In the first part of the thesis, we developed theory to enable prediction of the underlying dynamics of the TIC-driven proliferative hierarchy. To connect this theory to reality, and to make it clinically useful, would require carefully conducted measures of stem proportion, like those conducted with cell-labelling experiments in mice [61], over time and throughout the course of treatment. Alternatively, and possibly more relevant, would be using temporally distinct measures, such as serial biopsies in leukemia, or multi-region sequencing done on lesions that appeared at different times during the course of therapy: a primary and late metastasis, for example.

To test the predictions in the second part of the thesis, concerning the organization of vessels and radiation effect, would be impossible in human samples. Instead, I propose using a window chamber model, either in the flank or calvarium of a mouse. With this technique, together with sequenced radiation and vascular normalisation therapy, vessel organisation could be measured and correlated with radiation effect, before and after vessel normalisation. If these experiments validated the theory, then there would still be a step required to correlate imaging findings with vessel density and organisation. Vessel density has already been correlated with apparent diffusion coefficient maps using MRI, but organisation has not. These studies could be performed on autopsy specimens where MRI studies could be registered with histology, or alternatively using an animal model system like the above. So, while there remain several steps before the theory here can be applied clinically, we have provided steps forward which can help researchers to connect in the ways I have proposed here, or other ways they see fit.

Regardless of the utility of the work now, as I am nearing the end of this process, I feel as though I am better able to approach the career of cancer research ahead of me, and better able to operate in the world of mathematical biology and oncology at large. Through my course of study, learning new techniques and learning to question the dogmas inherent in the teachings of clinical oncology, I feel as though I have seen something I can no longer unsee. Like the story of the emperor's new clothes, I feel as though oncology research today is misleading itself, but there are a few of us who can see it for what it is – and we can cry out the truth that we see. And, thanks to those before us, we are no longer shouting into the wind, but instead are part of a movement, based on years of hard work, that is gaining ground – and this can only be a good thing for those living with, and those who will one day get, cancer.

Appendix A

Chapter 2 supplemental material

A.1 Size and number of biopsies: diminishing returns

To more exhaustively explore the relationship between biopsy size and the diversity estimate, we take 200 biopsies at random, with replacement, from the computational domain of increasing size. We then plot the distributions on a single figure (Figure A.1, left) to illustrate the increase in average diversity with increasing biopsy size. We also plot the mean and standard deviation of each of these distributions (Figure A.1, right) and notice that there seems to be diminishing returns in size increase suggesting that bigger is better, but that investing in larger biopsies does not necessarily return similar gains in diversity estimation.

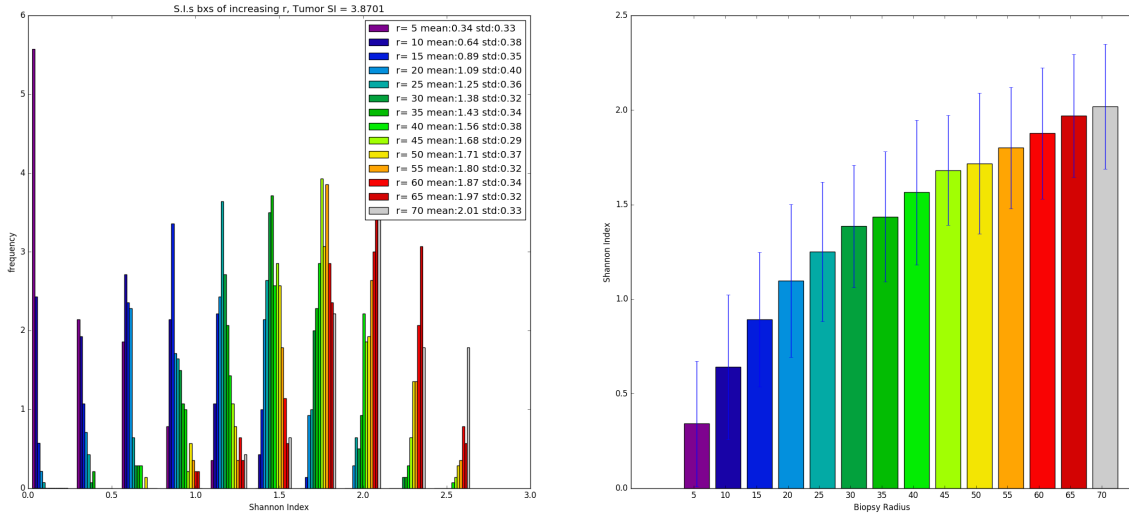


Figure A.1: Exhaustive exploration of biopsy size changes. We plot the distributions (left) of Shannon index from 200 biopsies taken at random, with replacement, of 14 different radii. We find as the biopsies get larger, the mean shifts to the right and approaches the true diversity and the standard deviation tightens. Right, we plot the mean and standard deviation of these distributions. As expected, larger biopsies on average better approximate the diversity of the tumour.

A.2 Shape normalisation

In Section 3.3.3 we found that without normalising the centroid distance by area, we lost all possibility of signal because, as we see in Figure A.2, the size distribution is heavily skewed.

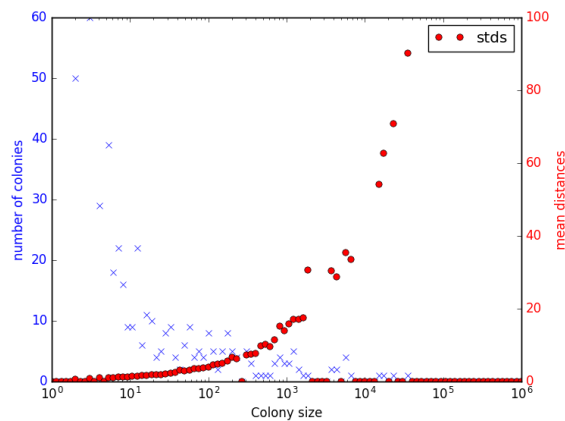


Figure A.2: Plot of colony size versus frequency (blue) and mean distance to centroid versus colony size (red) for the same colonies. The many, very small colonies ended up washing out what signal there may have been.

Appendix B

Chapter 3 supplemental information

B.1 Temporal dynamics

While the number of mutations differs, the rate of acquisition is not captured by these static plots. While we are mostly interested in the spatial patterning, the growth kinetics of these tumours are also of interest. While much of this has been described in the past, by Enderling and others [65, 187], we plot the resulting temporal dynamics of each simulation. In Figure B.1 we plot on the left the number of mutations vs. time for all tested symmetric division rates (0.2, 0.4, 0.6, 0.8 and 1.0) and find, as expected, that the tumours with the higher symmetric division rate acquire mutations more quickly, and more of them over the course of the simulation.

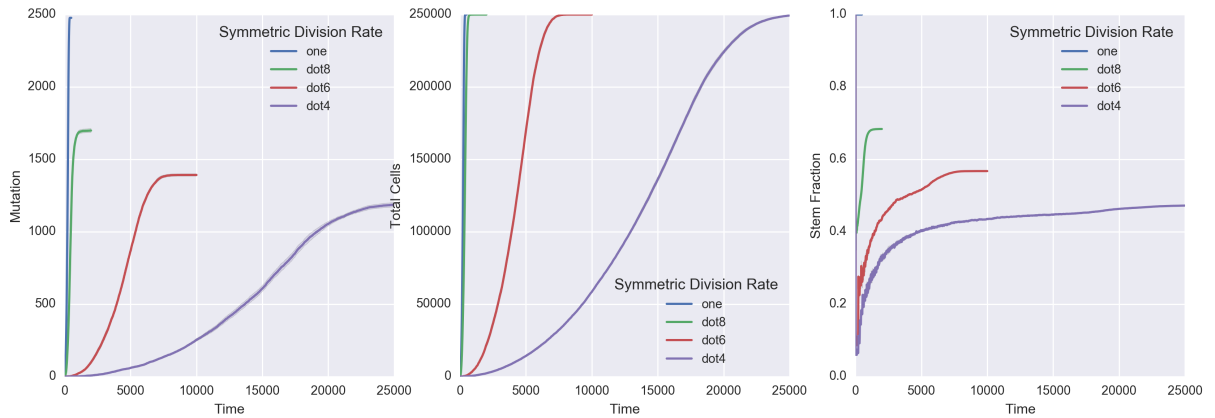


Figure B.1: Tumour growth dynamics for a range of symmetric division rates. For each different symmetric division rate, we plot the mutations vs. time (left), tumour size vs. time (middle), and stem fraction vs. time (right). Error bars represent the range of the data for 75 simulations (very tight).

This ‘protection’ against mutation in the lower symmetric division probability simulations is something Gatenby et al. pointed out in their perspective on colonic crypt dynamics and carcinogenesis [84], suggesting that stemness is protective against (cancer causing) mutations in the colonic crypt. This is somewhat trivial, as we are counting

only mutations in the TIC compartment. The middle panel shows a plot for the same simulations, but of tumour size vs. time for each simulation. Also as expected, and as reported before, the higher the symmetric division rate, the faster the overall tumour growth. In the right panel, we plot TIC fraction vs. time.

B.2 Non-spatial Model comparative statistics

We present here an exhaustive list of the pairwise comparisons of distributions presented in Section 4.3 for the non-spatial model.

Table B.1: First and second moments of distributions of tree measures across symmetric division probability for the non-spatial model.

Means					
	0.2	0.4	0.6	0.8	1.0
Sackin	2128.6	2198.5	2799.6	3763.6	5036.9
Sackin _{PDA}	0.085	0.074	0.060	0.051	0.043
Sackin _{Yule}	-10.17	-10.61	-11.31	-11.97	-12.61
$\bar{N}$	2.49	2.28	2.17	2.14	2.10
B1	105.71	73.00	69.78	78.80	89.68
Standard deviations					
	0.2	0.4	0.6	0.8	1.0
Sackin	239.2	256.9	134.2	296.9	237.6
Sackin _{PDA}	0.009	0.008	0.003	0.004	0.002
Sackin _{Yule}	0.271	0.268	0.131	0.160	0.089
$\bar{N}$	0.27	0.26	0.11	0.16	0.08
B1	10.51	8.70	8.00	7.64	8.92

Table B.2: Pairwise Wilcoxon rank sum comparisons for the non-spatial model.

Sackin					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	0.03				
0.6	5.80e-16	4.17e-14			
0.8	6.86e-18	6.86e-18	6.86e-18		
1.0	6.86e-18	6.86e-18	6.86e-18	1.10e-16	
Sackin _{PDA}					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	1.12e-11				
0.6	6.86e-18	1.75e-16			
0.8	6.86e-18	6.53e-17	1.31e-16		
1.0	6.86e-18	6.86e-18	6.86e-18	6.16e-17	
Sackin _{Yule}					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	2.00e-13				
0.6	6.86e-18	8.72e-18			
0.8	6.86e-18	6.86e-18	1.24e-16		
1.0	6.86e-18	6.86e-18	6.86e-18	6.86e-18	
$\bar{N}$					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	5.56e-08				
0.6	1.47e-13	5.84e-05			
0.8	5.49e-16	1.31e-09	0.001		
1.0	8.62e-16	1.90e-13	5.35e-08	0.01	
B1					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	2.70e-17				
0.6	1.14e-17	0.094			
0.8	9.53e-17	0.0005	2.13e-07		
1.0	3.02e-11	5.43e-12	1.25e-14	1.81e-08	

B.3 Comparative pairwise statistics of spatial model

We present here an exhaustive list of the pairwise comparisons of distributions presented in Section 4.3 for the spatial model.

Table B.3: First and second moments of distributions of tree measures across symmetric division probability for the spatial model.

Means					
	0.2	0.4	0.6	0.8	1.0
Sackin	2965.1	3565.6	4645.0	5452.3	7200.4
Sackin _{PDA}	0.116	0.106	0.101	0.089	0.069
Sackin _{Yule}	-9.283	-9.637	-9.836	-10.34	-11.31
$\bar{N}$	0.332	0.366	0.385	0.396	0.215
B1	90.06	97.46	107.84	127.6	217.6
Standard deviations					
	0.2	0.4	0.6	0.8	1.0
Sackin	331.4	408.3	527.9	640.3	523.5
Sackin _{PDA}	0.011	0.0113	0.011	0.011	0.005
Sackin _{Yule}	0.319	0.364	0.381	0.398	0.212
$\bar{N}$	0.332	0.365	0.385	0.396	0.215
B1	7.53	8.05	8.64	9.57	11.70

Table B.4: Pairwise Wilcoxon rank sum comparisons for the spatial model.

Sackin					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	6.35e-11				
0.6	1.04e-17	4.96e-15			
0.8	6.86e-18	9.26e-18	2.87e-09		
1.0	6.86e-18	6.86e-18	6.86e-18	6.31e-16	
Sackin _{PDA}					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	3.75e-05				
0.6	1.52e-08	0.085			
0.8	5.38e-15	2.37e-10	8.49e-08		
1.0	6.86e-18	6.86e-18	8.21e-18	1.10e-16	
Sackin _{Yule}					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	2.86e-06				
0.6	1.81e-10	0.016			
0.8	4.90e-16	2.04e-12	1.52e-08		
1.0	6.86e-18	6.86e-18	6.86e-18	2.13e-17	
$\bar{N}$					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	0.863				
0.6	0.002	0.010			
0.8	0.236	0.308	0.078		
1.0	0.002	0.013	7.56e-08	0.0001	
B1					
	0.2	0.4	0.6	0.8	1.0
0.2					
0.4	1.24e-05				
0.6	2.45e-14	2.05e-07			
0.8	6.86e-18	8.72e-18	2.58e-14		
1.0	6.86e-18	6.86e-18	6.86e-18	6.86e-18	

B.4 Raw vs. pruned trees

When we first investigated trees, and plotted them, in Figure 4.2, we pruned each tree of all terminal nodes (leaves), for ease of visualisation. While we used the full trees for all calculations going forward in the chapter, it is worth noting that the measures we utilised were not qualitatively affected in terms of their rank. Here, in Figure B.2, we plot the measured statistics for each pair of trees, both pruned and unpruned, and find that the rank of each measure is conserved.

B.5 Algorithm for developing individual cell ‘genomes’ from mutational flag and life history.

Here we describe the algorithm we created to develop the individual cells ‘genomes’ from the mutational flag and the life history. Using this reconstruction algorithm allowed for significant increase in speed of our CA model and reduced memory requirements by several orders of magnitude, and also allowed for easy implementation of the PAUP* algorithm for phylogenetic tree inference.

Data: Dictionary of unique Parent:Child pairs and spatial array of unique mutation flags at time point of interest.

Result: Array of bitstrings representing ‘genomes’ of cells in array.

for All cells in array **do**

if mutation ID = 0 **then**

 | break

end

 set bitstring to '1' + maxval(mutation ID) '0';

 final-parent = 2;

if mutation ID = 1 **then**

 | finalize bitstring

end

while final-parent > 1 **do**

 | final-parent = lookup parent(cell of interest) in dictionary;

 | flip bitstring at position(cell of interest) to '1';

end

 finalize bitstring;

end

Algorithm 1: Pseudo-code describing algorithm to reconstruct genomes from unique mutation flags and family history.

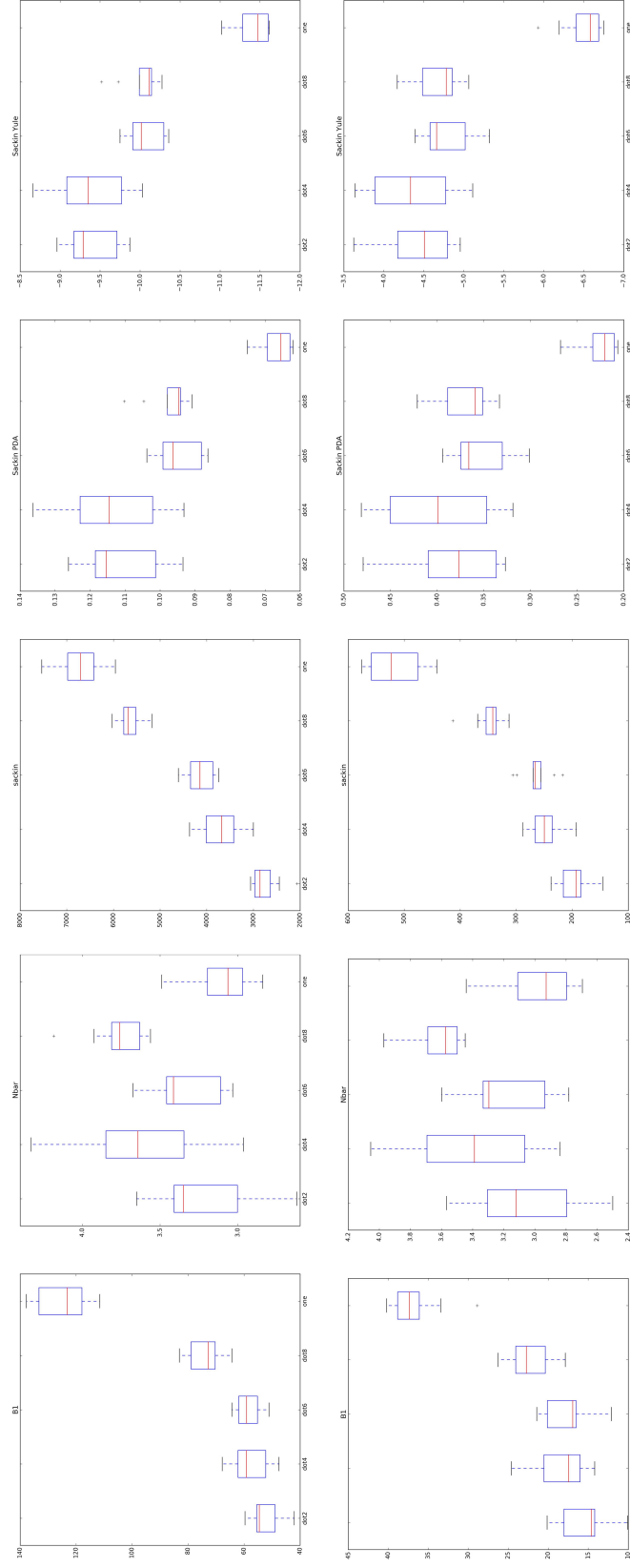


Figure B.2: Raw and pruned trees give rise to qualitatively similar summary measures with rank preserved. We plot the tree measures in each pair of plots with the raw measure on top and pruned on the bottom. In each pair, we plot the results from 10 simulations for each of the tested symmetric division probabilities. From left to right, we plot the B1 statistic, $\bar{N}$, the Sackin index, the PDA normalised Sackin index and finally the Yule normalised Sackin index.

Appendix C

Chapter 4 supplemental information

C.1 Analytic solution to diffusion equation

As a preliminary check of our numerical method, we consider a simplified scenario of the non-dimensional problem in which the entire left- and right-hand boundaries of the domain represent vessels running along the plane, and no cells are present. Any horizontal slice through this domain should then approximate the solution to the one-dimensional diffusion equation

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (\text{C.1})$$

for $t \geq 0$ and $x \in [0, L]$, with boundary conditions $c(0, t) = c(L, t) = 1$ and initial condition $c(x, 0) = 0$.

To solve, we observe that $\tilde{c} = c - 1$ satisfies equation (C.1) with Dirichlet boundary conditions subject to $\tilde{c}(0, t) = \tilde{c}(L, t) = 0$ and $\tilde{c}(x, 0) = -1$, and we find the general solution

$$\tilde{c}(x, t) = \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right) \exp\left[-\frac{n^2 \pi^2 D_C t}{L^2}\right]. \quad (\text{C.2})$$

To find the value of the coefficients, C_n , we impose the initial condition $\tilde{c}(x, 0) = -1$. We have

$$\tilde{c}(x, 0) = -1 = \sum_{n=1}^{\infty} C_n \sin\left(\frac{n\pi x}{L}\right), \quad (\text{C.3})$$

hence

$$C_n = -\frac{2}{L} \int_0^L \sin\left(\frac{n\pi x}{L}\right) dx = \frac{2}{n\pi} (\cos n\pi - 1). \quad (\text{C.4})$$

Thus we obtain the analytic solution

$$c(x, t) = \sum_{n=1}^{\infty} \left[\frac{2}{n\pi} (\cos n\pi - 1) \right] \sin \left(\frac{n\pi x}{L} \right) \exp \left[- \frac{n^2 \pi^2 D_C t}{L^2} \right] + 1. \quad (\text{C.5})$$

C.2 Calibration of model using tumour growth in an avascular domain: a tumour spheroid simulation

Cancers are often modelled as growing in an avascular domain with oxygen diffusing from the outside [32, 78, 96, 88]. This is a situation which has also been well studied *in vitro* by means of the well-established ‘tumour spheroid’ assay pioneered in the 1950s [168] (see [130] for a review) which has allowed researchers to study early tumour dynamics in a well controlled, three-dimensional experimental set up. A well conserved observation from many studies of this sort is a maximum size of spheroid supportable by only external oxygen before the appearance of central necrosis (see middle panel, Figure C.1).

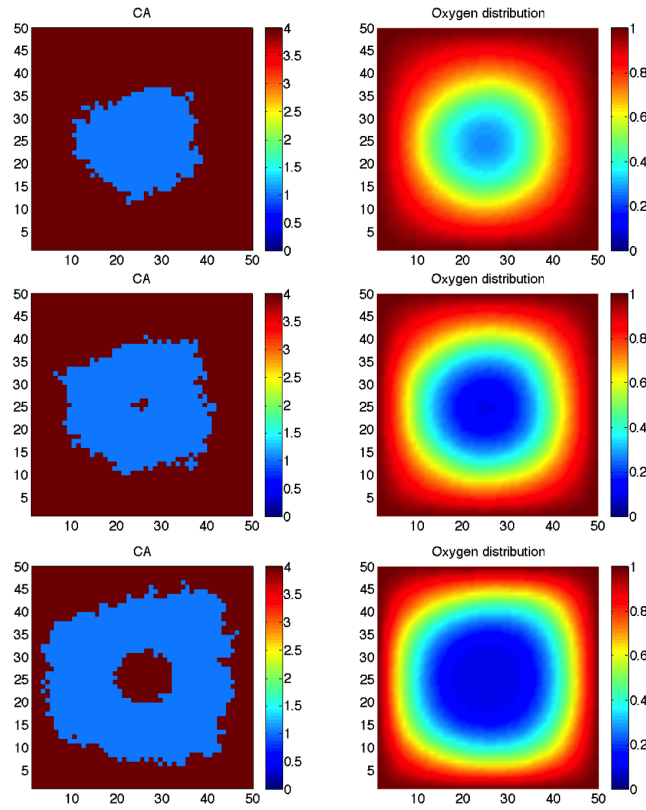


Figure C.1: Tumour growth in an avascular domain with oxygen diffusion from the outside displays characteristics of tumour ‘spheroid’ growth. Here, an otherwise empty domain is initiated with a single cancer cell. The oxygen at the edge of the domain is set to $c = 1$. Cells (left) and oxygen concentration (right) are plotted at three time points: before the onset of central necrosis (top), initiation of central necrosis (middle) and later in progression (bottom) when a nearly constant sized band of viable tumour cells is observed. From this calibration, we find that the maximal oxygen uptake rate most appropriate for our model is $r_c = 7.5 \times 10^{-4} \text{ s}^{-1}$.

As discussed in Section 5.2.3 the parameters in our model have been drawn from various sources, to include *in vitro* studies, human and non-human measurements and a variety of different cancer subtypes. This often leads to a range of potential parameter

choices and combinations. Therefore, to ensure that it recapitulates some well-established phenomena, we examine the impact of changing key parameters on model behaviour. One of these phenomena is the oxygen diffusion length, often measured to be on the order of 10 cancer cell diameters [78, 96] in tumour spheroid models. In Figure C.1, we show the appearance of central necrosis in a simplified model of a spheroid using our model system. We have adjusted the cellular oxygen uptake until the diffusion length of oxygen is approximately 10 cells. From this calibration, we find that the maximal oxygen uptake rate most appropriate for our model is $r_c = 7.5 \times 10^{-4} \text{ s}^{-1}$.

Appendix D

Chapter 5 supplemental information

D.1 Predicting carrying capacity

D.1.1 Aggregation coefficient

To understand the aggregation coefficient metric in terms of our simulation dataset, we relate it to each of the 500 simulations from Chapter 5. Figure D.1 shows scatter plots of representative vessel numbers with the total cellularity on the x -axis (averaged over 100 timesteps after dynamic equilibrium is reached) plotted against the aggregation measure on the y -axis. Across the simulations, we observe a consistent negative correlation between the two variables, as was suggested in our simple example in Figure 6.7, but this loses statistical significance as the domain reaches confluence as all of the populations become (approximately) equal to the size of the domain (bottom row, Figure D.1).

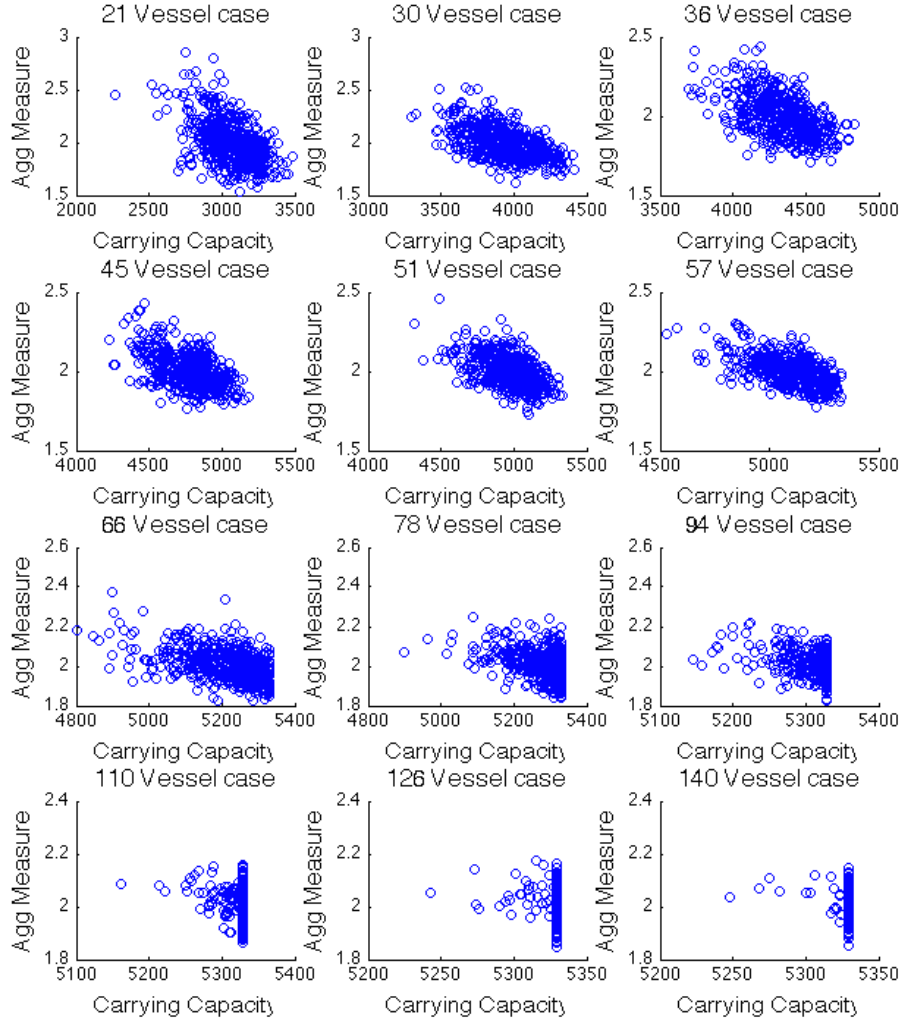


Figure D.1: Aggregation measure versus carrying capacity. We show twelve scatter plots illustrating the relationship in each of the 500 simulations represented in Figure 5.10 for a given initial vessel density between cell number at equilibrium (x -axis) and the aggregation coefficient, Φ (y -axis). We find that there is a negative correlation in the low vessel densities (vessel cases 21 to 66) which loses predictive capability as the domain becomes entirely filled, as most of the data points aggregate at the same cell number (beginning to be seen by vessel number 78).

D.1.2 Voronoi Polygon Area

In order to more clearly understand the relationship between the VPA distribution and carrying capacity, we plot the correlation coefficient (Figure D.2, left) and associated p-value (Figure D.2, right) of each moment of the VPA distribution against the carrying capacity. From this we see that the skewness of the VPA distribution is a significant ($p < 0.05$) predictor of carrying capacity for the majority of the vessel cases, including all the vessel cases below 94, specifically the cases where the aggregation measure was also of use. To illustrate this relationship we plot the VPA distribution skewness versus the associated cellularity at equilibrium in each of the 500 simulations for each vessel number in Figure D.3.

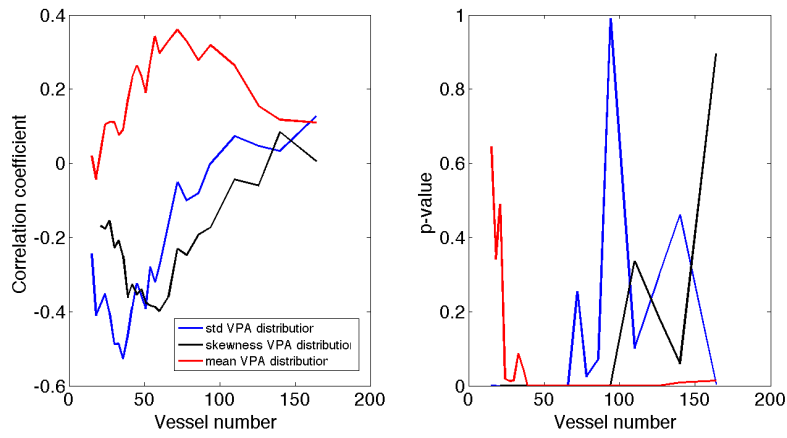


Figure D.2: Comparison of correlation of VPA distribution moments and carrying capacity. We show the correlation coefficient (left) and associated p-value (right) of each moment (mean (red), standard deviation (blue), and skewness (black)) of the VPA distribution for each of the 500 simulations in each vessel number family versus the carrying capacity for the associated simulation.

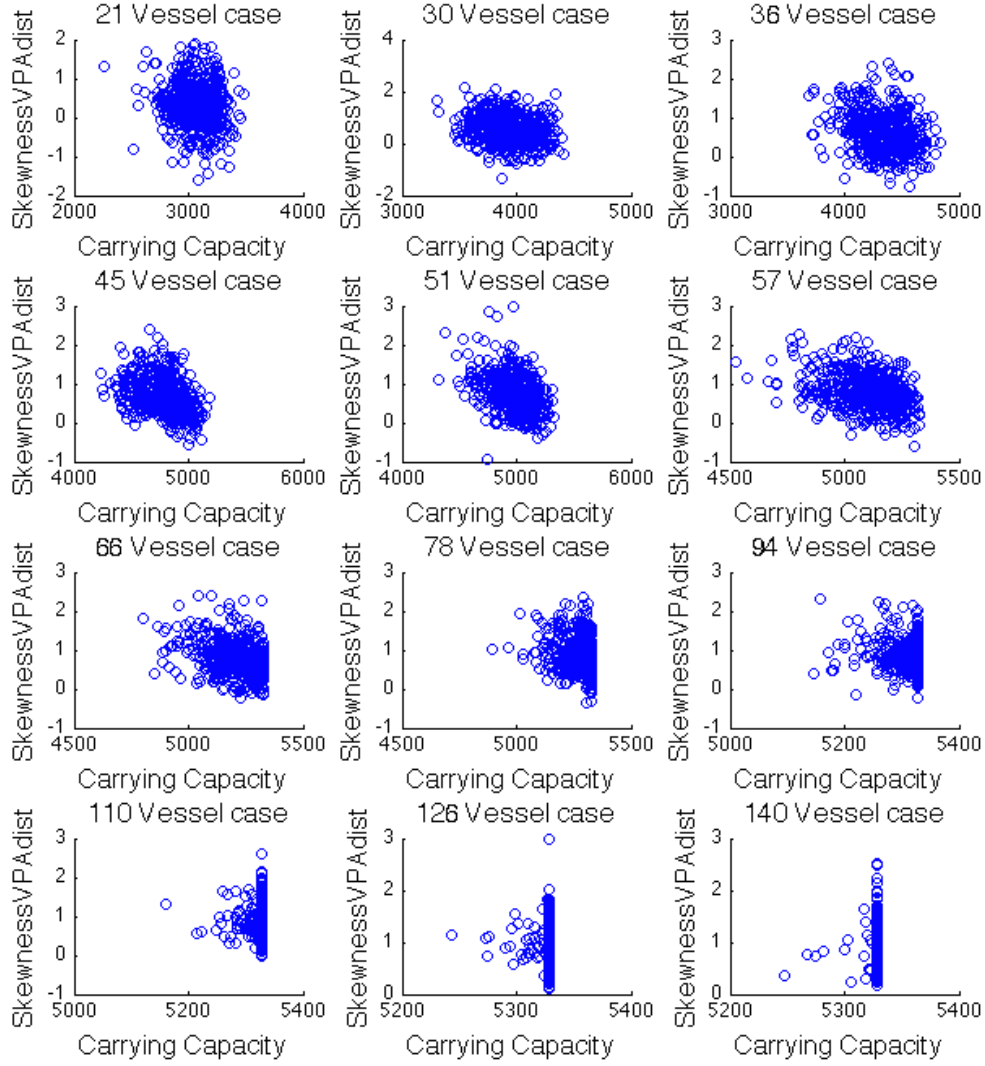


Figure D.3: VPA distribution skewness versus carrying capacity. We show twelve scatter plots illustrating the relationship in each of the 500 simulations represented in Figure 5.10 for a given initial vessel density, between cell number at equilibrium (x -axis) and the skewness of the Voronoi polygon area distribution (y -axis). We find that there is a poor negative correlation in the low vessel densities which loses predictive capability as the domain becomes entirely filled, as most of the data points aggregate at the same cell number.

D.1.3 Ripley's K

We calculate the mean of the variance stabilised Ripley's K (Ripley's L) for $r = 0$ to $r = 20$ for all of the 500 simulations in each vessel number case and plot it against the associated carrying capacity (Figure D.4). We find a significant negative correlation for the lower vessel densities which loses statistical significance (p-values not shown) as the domains begin to reach confluence for all but the minority of vessel arrangements.

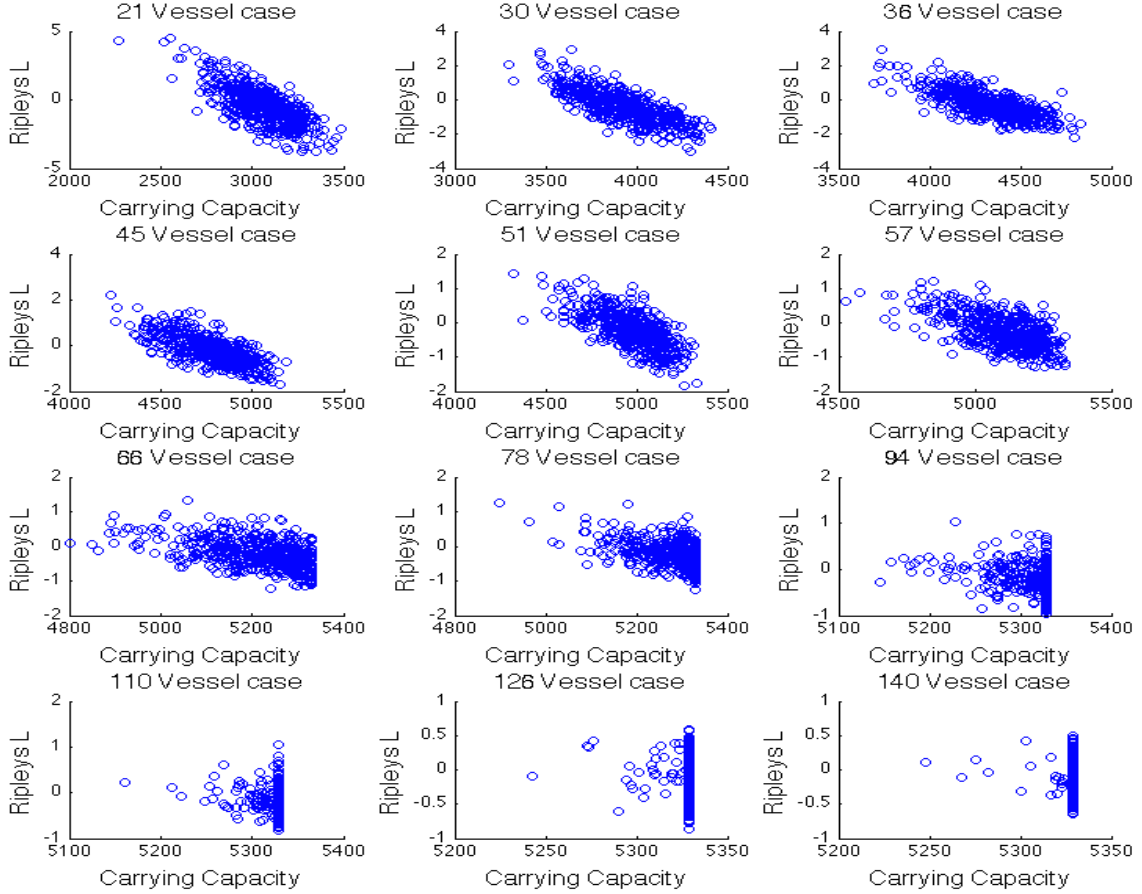


Figure D.4: Ripley's L versus carrying capacity. We show twelve scatter plots illustrating the relationship in each of the 500 simulations represented in Figure 5.10 for a given initial vessel density between cell number at equilibrium (x -axis) and Ripley's L (y -axis). We find that there is a significant negative correlation in the low vessel densities which loses predictive capability as the domain becomes entirely filled and all of the data points align at the full carrying capacity.

References

- [1] M Al-Hajj, MS Wicha, A Benito-Hernandez, SJ Morrison, and MF Clarke. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA, 100(7):3983, 2003.
- [2] AA Al-Shammari, EA Gaffney, and S Egginton. Re-evaluating the use of voronoi tessellations in the assessment of oxygen supply from capillaries in muscle. Bull Math Biol, 74(9):2204–2231, 2012.
- [3] AA Al-Shammari, EA Gaffney, and S Egginton. Modelling capillary oxygen supply capacity in mixed muscles: Capillary domains revisited. J Theor Biol, 356:47–61, 2014.
- [4] T Alarcón, HM Byrne, and PK Maini. A cellular automaton model for tumour growth in inhomogeneous environment. J Theor Biol, 225(2):257–274, 2003.
- [5] T Alarcón, HM Byrne, and PK Maini. A mathematical model of the effects of hypoxia on the cell-cycle of normal and cancer cells. J Theor Biol, 229(3):395–411, 2004.
- [6] T Alarcón, HM Byrne, and PK Maini. Towards whole-organ modelling of tumour growth. Prog Biophys Mol Biol, 85(2):451–472, 2004.
- [7] T Alarcón, MR Owen, HM Byrne, and PK Maini. Multiscale modelling of tumour growth and therapy: the influence of vessel normalisation on chemotherapy. Comp Math Methods Med, 7(2-3):85–119, 2006.
- [8] DJ Aldous. Probability distributions on cladograms. In Random Discrete Structures, pages 1–18. 1996.
- [9] DJ Aldous. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Statistical Science, pages 23–34, 2001.
- [10] JCL Alfonso, N Jagiella, L Núñez, MA Herrero, and D Drasdo. Estimating dose painting effects in radiotherapy: A mathematical model. PLOS ONE, 9(2):e89380, 2014.
- [11] PM Altrock, LL Liu, and F Michor. The mathematics of cancer: integrating quantitative models. Nat Rev Cancer, 15(12):730–745, 2015.
- [12] ARA Anderson. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. Math Med Biol, 22(2):163, 2005.

- [13] ARA Anderson and MAJ Chaplain. Continuous and discrete mathematical models of tumor-induced angiogenesis. Bull Math Biol, 60(5):857–899, 1998.
- [14] ARA Anderson, M Hassanein, KM Branch, J Lu, NA Lobdell, J Maier, D Basanta, B Weidow, A Narasanna, CL Arteaga, AB Reynolds, V Quaranta, L Estrada, and AM Weaver. Microenvironmental independence associated with tumor progression. Cancer Res, 69(22):8797–8806, 2009.
- [15] ARA Anderson and V Quaranta. Integrative mathematical oncology. Nat Rev Cancer, 8(3):227–234, 2008.
- [16] ARA Anderson, KA Rejniak, P Gerlee, and V Quaranta. Microenvironment driven invasion: a multiscale multimodel investigation. J Math Biol, 58(4):579–624, 2009.
- [17] ARA Anderson, IM Young, BD Sleeman, BS Griffiths, and WM Robertson. Nematode movement along a chemical gradient in a structurally heterogeneous environment. Fund Appl Nematol, 20(2):157–163, 1997.
- [18] K Anderson, C Lutz, FW Van Delft, CM Bateman, Y Guo, SM Colman, H Kempski, AV Moorman, I Titley, J Swansbury, L Kearney, T Enver, and M Greaves. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature, 469(7330):356–361, 2011.
- [19] L Arakelyan, Y Merbl, and Z Agur. Vessel maturation effects on tumour growth: validation of a computer model in implanted human ovarian carcinoma spheroids. Eur J Cancer, 41(1):159–167, 2005.
- [20] L Arakelyan, V Vainstein, and Z Agur. A computer algorithm describing the process of vessel formation and maturation, and its use for predicting the effects of anti-angiogenic and anti-maturation therapy on vascular tumor growth. Angiogenesis, 5(3):203–214, 2002.
- [21] RP Araujo and DLS McElwain. New insights into vascular collapse and growth dynamics in solid tumors. J Theor Biol, 228(3):335–346, 2004.
- [22] S Astanin and L Preziosi. Mathematical modelling of the warburg effect in tumour cords. J Theor Biol, 258(4):578–590, 2009.
- [23] S Bao, Q Wu, RE McLendon, Y Hao, Q Shi, AB Hjelmeland, MW Dewhirst, DD Bigner, and JN Rich. Glioma stem cells promote radioresistance by preferential activation of the dna damage response. Nature, 444(7120):756–760, 2006.
- [24] A-L Barabási and R Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.
- [25] GW Barendsen. Dose fractionation, dose rate and iso-effect relationships for normal tissue responses. Int J Rad Biol Phys, 8(11):1981–1997, 1982.

- [26] David Basanta, Jacob G Scott, Mayer N Fishman, Gustavo Ayala, Simon W Hayward, and Alexander RA Anderson. Investigating prostate cancer tumour–stroma interactions: clinical and biological insights from an evolutionary game. British journal of cancer, 106(1):174, 2012.
- [27] M Bjerknes. Expansion of mutant stem cell populations in the human colon. J Theor Biol, 178(4):381–385, 1996.
- [28] K Bloch, H Smith, V van Hamel Parsons, D Gavaghan, C Kelly, AG Fletcher, PK Maini, and R Callaghan. Metabolic alterations during the growth of tumour spheroids. Cell Biochem Biophys, 68(3):615–628, 2014.
- [29] MGB Blum and O François. On statistical tests of phylogenetic tree imbalance: the Sackin and other indices revisited. Math Biosci, 195(2):141–153, 2005.
- [30] D Bonnet and JE Dick. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med, 3(7):730–7, 1997.
- [31] CJW Breward, HM Byrne, and CE Lewis. A multiphase model describing vascular tumour growth. Bull Math Biol, 65(4):609–640, 2003.
- [32] HM Byrne. Modelling avascular tumour growth. Cancer Modelling and Simulation, pages 75–120, 2003.
- [33] HM Byrne. Dissecting cancer through mathematics: from the cell to the animal model. Nat Rev Cancer, 10(3):221–230, 2010.
- [34] HM Byrne and MAJ Chaplain. Necrosis and apoptosis: distinct cell loss mechanisms in a mathematical model of avascular tumour growth. Comput Math Methods Med, 1(3):223–235, 1998.
- [35] DP Cahill, KW Kinzler, B Vogelstein, and C Lengauer. Genetic instability and Darwinian selection in tumours. Trends Cell Biol, 9(12):M57–M60, 1999.
- [36] P Calabresi, PS Schein, and SA Rosenberg. Medical Oncology: Basic Principles and Clinical Management of Cancer. MacMillan Pub Co, New York, NY, 1985.
- [37] A Carreau, BE Hafny-Rahbi, A Matejuk, C Grillon, and C Kieda. Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. J Cell Mol Med, 15(6):1239–1253, 2011.
- [38] Lewis Carroll, Sir John Tenniel, et al. Alice’s Adventures in Wonderland: And, Through the Looking-glass. Hartsdale House, 1885.
- [39] JJ Casciari, SV Sotirchos, and RM Sutherland. Variations in tumor cell growth rates and metabolism with oxygen concentration, glucose concentration, and extracellular pH. J Cell Physiol, 151(2):386–394, 1992.

- [40] JD Chapman, CJ Gillespie, AP Reuvers, and DL Dugle. The inactivation of chinese hamster cells by X-rays: The effects of chemical modifiers on single-and double-events. Rad Res, 64(2):365–375, 1975.
- [41] CY Chen, HM Byrne, and JR King. The influence of growth-induced stress from the surrounding medium on the development of multicell spheroids. J Math Biol, 43(3):191–220, 2001.
- [42] J Chen, Y Li, T Yu, RM McKay, DK Burns, SG Kernie, and LF Parada. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature, 488(7412):522–6, 2012.
- [43] S Chiou, IP Winters, J Wang, S Naranjo, C Dudgeon, FB Tamburini, JJ Brady, D Yang, BM Grüner, C Chuang, DR Caswell, H Zeng, P Chu, GE Kim, Carpizo DR, SK Kim, and MM Winslow. Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. Genes Dev, 29(14):1576–1585, 2015.
- [44] AH Compton. A quantum theory of the scattering of X-rays by light elements. Phys Rev, 21(5):483, 1923.
- [45] SJ Conley, E Gheordunescu, P Kakarala, B Newman, H Korkaya, AN Heath, SG Clouthier, and MS Wicha. Antiangiogenic agents increase breast cancer stem cells via the generation of tumor hypoxia. Proc Natl Acad Sci USA, 109(8):2784–9, 2012.
- [46] D Corwin, C Holdsworth, RC Rockne, AD Trister, Maciej M Mrugala, JK Rockhill, RD Stewart, M Phillips, and KR Swanson. Toward patient-specific, biologically optimized radiation therapy plans for the treatment of glioblastoma. PLOS ONE, 8:e79115, 2013.
- [47] K Curtius, C Wong, WD Hazelton, AM Kaz, A Chak, JE Willis, WM Grady, and EG Luebeck. A molecular clock infers heterogeneous tissue age among patients with Barrett’s esophagus. PLoS Comput Biol, 12(5):e1004919, 2016.
- [48] RG Dale. The application of the linear-quadratic dose-effect equation to fractionated and protracted radiotherapy. Br J Radiol, 58(690):515–528, 1985.
- [49] RG Dale. Time-dependent tumour repopulation factors in linear-quadratic equations—implications for treatment strategies. Radiother Oncol, 15(4):371–381, 1989.
- [50] C Darwin. The origin of species. 1872.
- [51] M Dean, T Fojo, and S Bates. Tumour stem cells and drug resistance. Nat Rev Cancer, 5(4):275–84, 2005.
- [52] A Dhawan, TA Graham, and AG Fletcher. A computational modeling approach for deriving biomarkers to predict cancer risk in premalignant disease. Cancer Prev Res, 9(4):283–295, 2016.

- [53] A Dhawan, M Kohandel, R Hill, and S Sivaloganathan. Tumour control probability in cancer stem cells hypothesis. PLOS ONE, 9(5):e96093, 2014.
- [54] A Dhawan, SAM Tonekaboni, JH Taube, Stephen Hu, N Sphyris, SA Mani, and M Kohandel. Mathematical modelling of phenotypic plasticity and conversion to a stem-cell state under hypoxia. Sci Rep, 6, 2016.
- [55] M Diehn, RW Cho, and MF Clarke. Therapeutic implications of the cancer stem cell hypothesis. Semin Rad Oncol, 19(2):78–86, 2009.
- [56] M Diehn, RW Cho, NA Lobo, T Kalisky, M Dorie, AN Kulp, D Qian, JS Lam, LE Ailles, M Wong, J Benzion, MJ Kaplan, I Wapnir, FM Dirbas, G Somlo, C Garbergolio, B Paz, J Shen, SK Lau, Quake SR, JM Brown, IL Weissman, and MF Clarke. Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature, 458(7239):780–783, 2009.
- [57] M Diehn and MF Clarke. Cancer stem cells and radiotherapy: new insights into tumor radioresistance. J Natl Cancer Inst, 98(24):1755–1757, 2006.
- [58] D Dingli and F Michor. Successful therapy must eradicate cancer stem cells. Stem Cells, 24(12):2603–2610, 2006.
- [59] F Doetsch, L Petreanu, I Caille, JM Garcia-Verdugo, and A Alvarez-Buylla. EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells. Neuron, 36(6):1021–1034, 2002.
- [60] A Dress, V Moulton, and W Terhalle. T-theory: An overview. European J Combin, 17(2):161–175, 1996.
- [61] G Driessens, B Beck, A Caauwe, BD Simons, and C Blanpain. Defining the mode of tumour growth by clonal analysis. Nature, 488(7412):527–530, 2012.
- [62] BJ Druker, M Talpaz, DJ Resta, B Peng, E Buchdunger, JM Ford, NB Lydon, H Kantarjian, R Capdeville, S Ohno-Jones, and CL Sawyers. Efficacy and safety of a specific inhibitor of the bcr-abl tyrosine kinase in chronic myeloid leukemia. N Engl J Med, 2001(344):1031–1037, 2001.
- [63] A Eberhard, S Kahlert, V Goede, B Hemmerlein, KH Plate, and HG Augustin. Heterogeneity of angiogenesis and blood vessel maturation in human tumors: implications for antiangiogenic tumor therapies. Cancer Res, 60(5):1388–1393, 2000.
- [64] S Egginton and E Gaffney. Tissue capillary supply—it’s quality not quantity that counts! Exp Physiol, 95(10):971–979, 2010.
- [65] H Enderling, ARA Anderson, MAJ Chaplain, A Beheshti, L Hlatky, and P Hahnfeldt. Paradoxical dependencies of tumor dormancy and progression on basic cell kinetics. Cancer Res, 69(22):8814–8821, 2009.
- [66] H Enderling, L Hlatky, and P Hahnfeldt. Migration rules: tumours are conglomerates of self-metastases. Br J Cancer, 100(12):1917–1925, 2009.

- [67] A Eramo, F Lotti, G Sette, E Piloizzi, M Biffoni, Aet Di Virgilio, C Conticello, L Ruco, C Peschle, and R De Maria. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death Differ, 15(3):504–514, 2008.
- [68] P Erdős and A Rényi. On the evolution of random graphs. Publ Math Inst Hung Acad Sci, 5(17-61):43, 1960.
- [69] Steven A Eschrich, Jimmy Pramana, Hongling Zhang, Haiyan Zhao, David Boulware, Ji-Hyun Lee, Gregory Bloom, Caio Rocha-Lima, Scott Kelley, Douglas P Calvin, et al. A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. International Journal of Radiation Oncology* Biology* Physics, 75(2):489–496, 2009.
- [70] D Ewing. The oxygen fixation hypothesis: a reevaluation. Am J Clin Oncol, 21(4):355–361, 1998.
- [71] BM Faltas, D Prandi, ST Tagawa, AM Molina, DM Nanus, C Sternberg, J Rosenberg, JM Mosquera, B Robinson, O Elemento, A Sboner, H Beltran, F Demichelis, and MA Rubin. Clonal evolution of chemotherapy-resistant urothelial carcinoma. Nat Genet, 2016.
- [72] H Farhidzadeh, B Chaudhury, JG Scott, DB Goldgof, LO Hall, RA Gatenby, RJ Gillies, and M Raghavan. Signal intensity analysis of ecological defined habitat in soft tissue sarcomas to predict metastasis development. In SPIE Medical Imaging, pages 97851H–97851H. International Society for Optics and Photonics, 2016.
- [73] Isaiah J Fidler. The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. Nature Reviews Cancer, 3(6):453–459, 2003.
- [74] A Fitzgibbon, M Pilu, and RB Fisher. Direct least square fitting of ellipses. IEEE Trans Pattern Anal Mach Intell, 21(5):476–480, 1999.
- [75] J Folkman, E Merler, C Abernathy, and G Williams. Isolation of a tumor factor responsible for angiogenesis. J Exp Med, 133(2):275–288, 1971.
- [76] JF Fowler. The linear-quadratic formula and progress in fractionated radiotherapy. Br J Radiol, 62(740):679–694, 1989.
- [77] JP Freyer, E Tustanoff, AJ Franko, and RM Sutherland. In situ oxygen consumption rates of cells in V-79 multicellular spheroids during growth. J Cell Physiol, 118(1):53–61, 1984.
- [78] HB Frieboes, ME Edgerton, JP Fruehauf, FRAJ Rose, LK Worrall, RA Gatenby, M Ferrari, and V Cristini. Prediction of drug response in breast cancer using integrative experimental/computational modeling. Cancer Res, 69(10):4484–4492, 2009.

- [79] S Gaedicke, F Braun, S Prasad, M Machein, E Firat, M Hettich, R Gudihal, X Zhu, K Klingner, J Schüler, CC Herold-Mende, A Grosu, M Behe, W Weber, H Mäcke, and G Niedermann. Noninvasive positron emission tomography and fluorescence imaging of CD133+ tumor stem cells. Proc Natl Acad Sci USA, page 201314189, 2014.
- [80] J Gallaher, A Babu, S Plevritis, and ARA Anderson. Bridging population and tissue scale tumor dynamics: A new paradigm for understanding differences in tumor growth and metastatic disease. Cancer Res, 74(2):426–435, 2014.
- [81] RA Gatenby and ET Gawlinski. A reaction-diffusion model of cancer invasion. Cancer Res, 56(24):5745–5753, 1996.
- [82] RA Gatenby, ET Gawlinski, AF Gmitro, B Kaylor, and RJ Gillies. Acid-mediated tumor invasion: a multidisciplinary study. Cancer Res, 66(10):5216–5223, 2006.
- [83] RA Gatenby and RJ Gillies. Why do cancers have high aerobic glycolysis? Nat Rev Cancer, 4(11):891–899, 2004.
- [84] RA Gatenby, RJ Gillies, and JS Brown. The evolutionary dynamics of cancer prevention. Nat Rev Cancer, 10(8):526–527, 2010.
- [85] RC Geary. The contiguity ratio and statistical mapping. Inc Stat, pages 115–146, 1954.
- [86] P Gerlee. The model muddle: in search of tumor growth laws. Cancer Res, 73(8):2407–2411, 2013.
- [87] P Gerlee and ARA Anderson. Stability analysis of a hybrid cellular automaton model of cell colony growth. Phys Rev E, 75(5):051911, 2007.
- [88] P Gerlee and ARA Anderson. A hybrid cellular automaton model of clonal evolution in cancer: The emergence of the glycolytic phenotype. J Theor Biol, 250(4):705–722, 2008.
- [89] M Gerlinger, S Horswell, J Larkin, AJ Rowan, MP Salm, I Varela, R Fisher, N McGranahan, N Matthews, CR Santos, P Martinex, B Phillimore, S Begum, A Rabinowitz, B Spencer-Dene, S Gulati, PA Bates, G Stamp, L Pickering, M Gore, DL Nicol, S Hazell, PA Futreal, A Stewart, and C Swanton. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet, 46(3):225–233, 2014.
- [90] M Gerlinger, AJ Rowan, S Horswell, J Larkin, D Endesfelder, E Gronroos, P Martinez, N Matthews, A Stewart, P Tarpey, I Varela, B Phillimore, S Begum, NQ McDonald, A Butler, D Jones, K Raine, C Latimer, CR Santos, M Nohadani, AC Eklund, B Spencer-Dene, G Clark, L Pickering, G Stamp, M Gore, Z Szallasi, J Downward, PA Futreal, and C Swanton. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med, 366(10):883–92, 2012.

- [91] RJ Gilbertson and TA Graham. Cancer: Resolving the stem-cell debate. Nature, 488(7412):462–463, 2012.
- [92] John H Gillespie. A simple stochastic gene substitution model. Theoretical population biology, 23(2):202–215, 1983.
- [93] RJ Gillies, D Verduzco, and RA Gatenby. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. Nat Rev Cancer, 12(7):487–93, 2012.
- [94] M Greaves and CC Maley. Clonal evolution in cancer. Nature, 481(7381):306–13, 2012.
- [95] BT Grenfell, OG Pybus, JR Gog, JLN Wood, JM Daly, JA Mumford, and EC Holmes. Unifying the epidemiological and evolutionary dynamics of pathogens. Science, 303(5656):327–332, 2004.
- [96] DR Grimes, C Kelly, K Bloch, and M Partridge. A method for estimating the oxygen consumption rate in multicellular tumour spheroids. J R Soc Interface, 11(92):20131124, 2014.
- [97] JA Grogan, PK Maini, J Pitt-Francis, and HM Byrne. Simulating tumour vasculature at multiple scales. In In Silico Oncology and Cancer Investigation (IARWISOCI), 2014 6th International Advanced Research Workshop on, pages 1–4. IEEE, 2014.
- [98] JA Grogan, B Markelc, AJ Connor, RJ Muschel, JM Pitt-Francis, PK Maini, and HM Byrne. Predicting the influence of microvascular structure on tumour response to radiotherapy. IEEE Trans Biomed Eng, 9(8), 2016.
- [99] James A Grogan, Anthony J Connor, Bostjan Markelc, Ruth J Muschel, Philip K Maini, Helen M Byrne, and Joe M Pitt-Francis. Microvessel chaste: An open library for spatial modeling of vascularized tissues. Biophysical Journal, 112(9):1767–1772, 2017.
- [100] EJ Hall and AJ Giaccia. Radiobiology for the Radiologist. Wolters Kluwer Health/Lippincott Williams and Wilkins, Philadelphia, 7th ed edition, 2012.
- [101] D Hanahan and RA Weinberg. The hallmarks of cancer. Cell, 100(1):57–70, 2000.
- [102] D Hanahan and RA Weinberg. Hallmarks of cancer: the next generation. Cell, 144(5):646–674, 2011.
- [103] MO Hill. Diversity and evenness: a unifying notation and its consequences. Ecology, 54(2):427–432, 1973.
- [104] AB Hjelmeland, Q Wu, JM Heddleston, GS Choudhary, J MacSwords, JD Lathia, R McLendon, D Lindner, A Sloan, and JN Rich. Acidic stress promotes a glioma stem cell phenotype. Cell Death Differ, 18(5):829–840, 2011.
- [105] EC Holmes and BT Grenfell. Discovering the phylodynamics of rna viruses. PLoS Comput Biol, 5(10):e1000505, 2009.

- [106] X Hu, L Ghisolfi, AC Keates, J Zhang, S Xiang, D Lee, and CJ Li. Induction of cancer cell stemness by chemotherapy. Cell Cycle, 11(14):2691–8, 2012.
- [107] S Huang. Non-genetic heterogeneity of cells in development: more than just noise. Development, 136(23):3853–3862, 2009.
- [108] A Humphries, B Cereser, LJ Gay, DSJ Miller, B Das, A Gutteridge, G Elia, E Nye, R Jeffery, R Poulsom, MR Novelli, M Rodriguez-Justo, SAC McDonald, NA Wright, and TA Graham. Lineage tracing reveals multipotent stem cells maintain human adenomas and the pattern of clonal expansion in tumor evolution. Proc Natl Acad Sci USA, 110(27):E2490–E2499, 2013.
- [109] CB Hyo-eun, DA Ruddy, VK Radhakrishna, JX Caushi, R Zhao, MM Hims, AP Singh, I Kao, D Rakiec, P Shaw, M Balak, A Raza, E Ackley, N Keen, MR Schlabach, M Palmer, RJ Leary, DY Chiang, WR Sellers, F Michor, VG Cooke, JM Korn, and F Stegmeier. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. Nat Med, 21(5):440–448, 2015.
- [110] A Ibrahim-Hashim, JW Wojtkowiak, MC Ribeiro, V Estrella, KM Bailey, HH Cornell, RA Gatenby, and RJ Gillies. Free base lysine increases survival and reduces metastasis in prostate cancer model. J Cancer Sci Ther, (4), 2011.
- [111] Jesús A Izaguirre, Rajiv Chaturvedi, Chengbang Huang, Trevor Cickovski, J Cofland, G Thomas, Gabor Forgacs, M Alber, G Hentschel, Stuart A Newman, et al. Compucell, a multi-model framework for simulation of morphogenesis. Bioinformatics, 20(7):1129–1137, 2004.
- [112] TL Jackson and HM Byrne. A mathematical model to study the effects of drug resistance and vasculature on the response of solid tumors to chemotherapy. Math Biosci, 164(1):17–38, 2000.
- [113] RK Jain. Normalizing tumor vasculature with anti-angiogenic therapy: a new paradigm for combination therapy. Nat Med, 7(9):987–989, 2001.
- [114] RK Jain. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science, 307(5706):58–62, 2005.
- [115] RK Jain. Normalizing tumor microenvironment to treat cancer: bench to bedside to biomarkers. J Clin Oncol, 31(17):2205–2218, 2013.
- [116] RK Jain, DG Duda, JW Clark, and JS Loeffler. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. Nat Clin Prac Oncol, 3(1):24–40, 2006.
- [117] P Jha and R Peto. Global effects of smoking, of quitting, and of taxing tobacco. N Engl J Med, 370(1):60–68, 2014.
- [118] MR Junttila and FJ de Sauvage. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature, 501(7467):346–354, 2013.

- [119] Artem Kaznatcheev, Jacob G Scott, and David Basanta. Edge effects in game-theoretic dynamics of spatially structured tumours. Journal of The Royal Society Interface, 12(108):20150154, 2015.
- [120] Artem Kaznatcheev, Robert Vander Velde, Jacob G Scott, and David Basanta. Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. British Journal of Cancer, 116(6):785–792, 2017.
- [121] H Kempf, M Bleicher, and M Meyer-Hermann. Spatio-temporal dynamics of hypoxia during radiotherapy. PloS ONE, 10(8):e0133357, 2015.
- [122] Mi-Young Kim, Thordur Oskarsson, Swarnali Acharyya, Don X Nguyen, Xiang H-F Zhang, Larry Norton, and Joan Massagué. Tumor self-seeding by circulating cancer cells. Cell, 139(7):1315–1326, 2009.
- [123] M Kimura. Evolutionary rate at the molecular level. Nature, 217(5129):624–626, 1968.
- [124] M Kimura. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics, 61(4):893, 1969.
- [125] M Kimura. The Neutral Theory of Molecular Evolution. Cambridge University Press, 1984.
- [126] M Kirkpatrick and M Slatkin. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution, pages 1171–1181, 1993.
- [127] Allon M Klein and Benjamin D Simons. Universal patterns of stem cell fate in cycling adult tissues. Development, 138(15):3103–3111, 2011.
- [128] R Kostadinov, CC Maley, and MK Kuhner. Bulk genotyping of biopsies can create spurious evidence for heterogeneity in mutation content. PLoS Comput Biol, 12(4):e1004413, 2016.
- [129] A Krogh. The number and distribution of capillaries in muscles with calculations of the oxygen pressure head necessary for supplying the tissue. J Physiol, 52(6):409–415, 1919.
- [130] LA Kunz-Schughart. Multicellular tumor spheroids: intermediates between monolayer culture and in vivo tumor. Cell Biol Int, 23(3):157–161, 1999.
- [131] JD Lathia, M Hitomi, J Gallagher, SP Gadani, J Adkins, A Vasanji, L Liu, CE Eyler, JM Heddleston, Q Wu, S Minhas, A Soeda, DJ Hoepfner, R Ravin, RDG McKay, RE McLendon, D Corbeil, A Chenn, AB Hjelmeland, DM Park, and JN Rich. Distribution of CD133 reveals glioma stem cells self-renew through symmetric and asymmetric cell divisions. Cell Death Dis, 2:e200, 2011.
- [132] DEA Lea. Actions of radiations on living cells. Cambridge: University Press, 1946.
- [133] P Legendre. Spatial autocorrelation: trouble or new paradigm? Ecology, 74(6):1659–1673, 1993.

- [134] B Lemasson, S Valable, R Farion, A Krainik, C Rémy, and EL Barbier. In vivo imaging of vessel diameter, size, and density: a comparative study between MRI and histology. Magn Reson Med, 69(1):18–26, 2013.
- [135] M Leszczyniecka, T Roberts, P Dent, S Grant, and PB Fisher. Differentiation therapy of human cancer: basic science and clinical applications. Pharmacol Ther, 90(2):105–156, 2001.
- [136] GE Leventhal, R Kouyos, T Stadler, V Von Wyl, S Yerly, J Böni, C Celleraï, T Klimkait, HF Günthard, and S Bonhoeffer. Inferring epidemic contact structure from phylogenetic trees. PLoS Comput Biol, 8(3):e1002413, 2012.
- [137] SF Levy, JR Blundell, S Venkataram, DA Petrov, DS Fisher, and G Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature, 2015.
- [138] LA Liotta, GM Saidel, and J Kleinerman. Diffusion model of tumor vascularization and growth. Bull Math Biol, 39(1):117–128, 1977.
- [139] S Liu, C Ginestier, Sing J Ou, SG Clouthier, SH Patel, F Monville, H Korkaya, A Heath, J Dutcher, CG Kleer, Y Jung, G Dontu, R Taichman, and MS Wicha. Breast cancer stem cells are regulated by mesenchymal stem cells through cytokine networks. Cancer Res, 71(2):614–24, 2011.
- [140] MC Lloyd, KO Alfarouk, D Verduzco, MM Bui, RJ Gillies, ME Ibrahim, JS Brown, and RA Gatenby. Vascular measurements correlate with estrogen receptor status. BMC Cancer, 14(1):279, 2014.
- [141] MC Lloyd, P Allam-Nandyala, CN Purohit, N Burke, D Coppola, and MM Bui. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: How reliable is it? J Pathol Inform, 1:29, 2010.
- [142] MC Lloyd, JJ Cunningham, MM Bui, RJ Gillies, JS Brown, and RA Gatenby. Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces. Cancer research, 76(11):3136–3144, 2016.
- [143] Carlos Lopez-Garcia, Allon M Klein, Benjamin D Simons, and Douglas J Winton. Intestinal stem cell replacement follows a pattern of neutral drift. Science, 330(6005):822–825, 2010.
- [144] Y Luo, K Dallaglio, Y Chen, WA Robinson, SE Robinson, MD McCarter, J Wang, R Gonzalez, DC Thompson, DA Norris, DR Roop, V Vasiliou, and M Fujita. ALDH1A isozymes are markers of human melanoma stem cells and potential therapeutic targets. Stem Cells, 30(10):2100–2113, 2012.
- [145] P Macklin, S McDougall, ARA Anderson, Mark AJ Chaplain, V Cristini, and J Lowengrub. Multiscale modelling and nonlinear simulation of vascular tumour growth. J Math Biol, 58(4-5):765–798, 2009.

- [146] DR Maddison, DL Swofford, and WP Maddison. NEXUS: an extensible file format for systematic information. Syst Biol, 46(4):590–621, 1997.
- [147] JA Magee, E Piskounova, and SJ Morrison. Cancer stem cells: impact, heterogeneity, and uncertainty. Cancer Cell, 21(3):283–296, 2012.
- [148] AE Magurran. Measuring biological diversity. Blackwell Publishing, Oxford, 2013.
- [149] CC Maley, PC Galipeau, JC Finley, VJ Wongsurawat, X Li, CA Sanchez, TG Paulson, PL Blount, R Risques, PS Rabinovitch, and BJ Reid. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet, 38(4):468–473, 2006.
- [150] E Malinen, Å Søvik, D Hristov, ØS Bruland, and DR Olsen. Adapting radiotherapy to hypoxic tumours. Phys Med Biol, 51(19):4903, 2006.
- [151] P Martinez, MR Timmer, CT Lau, S Calpe, M del Carmen Sancho-Serra, D Straub, A Baker, SL Meijer, FJW Ten Kate, RC Mallant-Hent, AHJ Naber, AHAM van Oijen, LC Baak, P Scholten, CJM Böhmer, P Fockens, JJGHM Bergman, CC Maley, TA Graham, and KK Krishnadath. Dynamic clonal equilibrium and predetermined cancer risk in Barrett’s oesophagus. Nat Commun, 7:12158, 2016.
- [152] A Marusyk, V Almendro, and K Polyak. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer, 12(5):323–34, 2012.
- [153] SR McDougall, ARA Anderson, and MAJ Chaplain. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. J Theor Biol, 241(3):564–589, 2006.
- [154] SR McDougall, ARA Anderson, MAJ Chaplain, and JA Sherratt. Mathematical modelling of flow through vascular networks: implications for tumour-induced angiogenesis and chemotherapy strategies. Bull Math Biol, 64(4):673–702, 2002.
- [155] CD McFarland, KS Korolev, GV Kryukov, SR Sunyaev, and LA Mirny. Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci USA, 110(8):2910–2915, 2013.
- [156] CD McFarland, LA Mirny, and KS Korolev. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. Proc Natl Acad Sci USA, 111(42):15138–15143, 2014.
- [157] AJ McGoron, P Nair, and RW Schubert. Michaelis-Menten kinetics model of oxygen consumption by rat brain slices following hypoxia. Ann Biomed Eng, 25(3):565–572, 1997.
- [158] CE Meacham and SJ Morrison. Tumour heterogeneity and cancer cell plasticity. Nature, 501(7467):328–337, 2013.
- [159] LMF Merlo, JW Pepper, BJ Reid, and CC Maley. Cancer as an evolutionary and ecological process. Nat Rev Cancer, 6(12):924–935, 2006.

- [160] A Mir, F Rosselló, and L Rotger. A new balance index for phylogenetic trees. Math Biosci, 241(1):125–136, 2013.
- [161] Portia M Mira, Kristina Crona, Devin Greene, Juan C Meza, Bernd Sturmfels, and Miriam Barlow. Rational design of antibiotic treatment plans: a treatment strategy for managing evolution and reversing resistance. PloS one, 10(5):e0122283, 2015.
- [162] Gary R Mirams, Christopher J Arthurs, Miguel O Bernabeu, Rafel Bordas, Jonathan Cooper, Alberto Corrias, Yohan Davit, Sara-Jane Dunn, Alexander G Fletcher, Daniel G Harvey, et al. Chaste: an open source c++ library for computational physiology and biology. PLoS computational biology, 9(3):e1002970, 2013.
- [163] AR Mitchell and DF Griffiths. The finite difference method in partial differential equations. John Wiley, 1980.
- [164] AO Mooers and SB Heard. Inferring evolutionary process from phylogenetic tree shape. Q Rev Biol, pages 31–54, 1997.
- [165] PAP Moran. Notes on continuous stochastic phenomena. Biometrika, pages 17–23, 1950.
- [166] SJ Morrison and J Kimble. Asymmetric and symmetric stem-cell divisions in development and cancer. Nature, 441(7097):1068–1074, 2006.
- [167] CI Morton, L Hlatky, P Hahnfeldt, and H Enderling. Non-stem cancer cell kinetics modulate solid tumor progression. Theor Biol Med Mod, 8:48, 2011.
- [168] AA Moscona. Tissues from dissociated cells. Sci Am, 200:132–144, 1959.
- [169] N Murugaesu, GA Wilson, NJ Birkbak, TBK Watkins, N McGranahan, S Kumar, N Abbassi-Ghadi, M Salm, R Mitter, S Horswell, A Rowan, B Phillimore, J Biggs, S Begum, N Matthews, D Hochhauser, GB Hanna, and C Swanton. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. Cancer Discov, 5(8):821–831, 2015.
- [170] K Naxerova and RK Jain. Using tumour phylogenetics to identify the roots of metastasis in humans. Nat Rev Clin Oncol, 12(5):258–272, 2015.
- [171] Daniel Nichol, Peter Jeavons, Alexander G Fletcher, Robert A Bonomo, Philip K Maini, Jerome L Paul, Robert A Gatenby, Alexander RA Anderson, and Jacob G Scott. Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. PLoS computational biology, 11(9):e1004493, 2015.
- [172] M Norman, KA Wisniewska, K Lawrenson, P Garcia-Miranda, M Tada, M Kajita, H Mano, S Ishikawa, M Ikegawa, T Shimada, and Y Fujita. Loss of Scribble causes cell competition in mammalian cells. J Cell Sci, 125(1):59–66, 2012.
- [173] PC Nowell. The clonal evolution of tumor cell populations. Science, 194(4260):23–28, 1976.

- [174] LJ Nugent and RK Jain. Extravascular diffusion in normal and neoplastic tissues. Cancer Res, 44(1):238–244, 1984.
- [175] C O’Brien, A Pollett, S Gallinger, and JE Dick. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature, 445(7123):106–110, 2007.
- [176] EB O’Dea and CO Wilke. Contact heterogeneity and phylodynamics: how contact networks shape parasite evolutionary trees. Interdisciplinary Perspectives on Infectious Diseases, 2011, 2010.
- [177] T Ohta and JH Gillespie. Development of neutral and nearly neutral theories. Theor Pop Biol, 49(2):128–142, 1996.
- [178] ME Orme and MAJ Chaplain. A mathematical model of vascular tumour growth and invasion. Math Comp Mod, 23(10):43–60, 1996.
- [179] Markus R Owen, I Johanna Stamper, Munitta Muthana, Giles W Richardson, Jon Dobson, Claire E Lewis, and Helen M Byrne. Mathematical modeling predicts synergistic antitumor effects of combining a macrophage-based, hypoxia-targeted gene therapy with chemotherapy. Cancer research, 71(8):2826–2837, 2011.
- [180] MR Owen, T Alarcón, PK Maini, and HM Byrne. Angiogenesis and vascular remodelling in normal and cancerous tissues. J Math Biol, 58(4-5):689–721, 2009.
- [181] ML O’Connor, D Xiang, S Shigdar, J Macdonald, Y Li, T Wang, C Pu, Z Wang, L Qiao, and W Duan. Cancer stem cells: a contentious hypothesis now moving forward. Cancer Lett, 344(2):180–187, 2014.
- [182] B Palcic and LD Skarsgard. Reduced oxygen enhancement ratio at low doses of ionizing radiation. Radiat Res, 100(2):328–339, 1984.
- [183] R Pang, WL Law, ACY Chu, JT Poon, CSC Lam, AKM Chow, Lui Ng, LWH Cheung, XR Lan, HY Lan, VPY Tan, TC Yau, RT Poon, and BCY Wong. A subpopulation of CD26+ cancer stem cells with metastatic capacity in human colorectal cancer. Cell Stem Cell, 6(6):603–15, 2010.
- [184] AA Patel, ET Gawlinski, SK Lemieux, and RA Gatenby. A cellular automaton model of early tumor growth and invasion. J Theor Biol, 213(3):315–31, 2001.
- [185] H Perfahl, HM Byrne, T Chen, V Estrella, T Alarcón, A Lapin, RA Gatenby, RJ Gillies, MC Lloyd, PK Maini, M Reuss, and MR Owen. Multiscale modelling of vascular tumour growth in 3d: the roles of domain size and boundary conditions. PLOS ONE, 6(4):e14790, 2011.
- [186] Joe Pitt-Francis, Pras Pathmanathan, Miguel O Bernabeu, Rafel Bordas, Jonathan Cooper, Alexander G Fletcher, Gary R Mirams, Philip Murray, James M Osborne, Alex Walter, et al. Chaste: a test-driven approach to software development for biological modelling. Computer Physics Communications, 180(12):2452–2471, 2009.

- [187] J Poleszczuk, P Hahnfeldt, and H Enderling. Biphasic modulation of cancer stem cell-driven solid tumour dynamics in response to reactivated replicative senescence. Cell Prolif, 47(3):267–276, 2014.
- [188] J Poleszczuk, P Hahnfeldt, and H Enderling. Evolution and phenotypic selection of cancer stem cells. PLoS Comput Biol, 11(3):e1004025, 2015.
- [189] K Polyak. Tumor heterogeneity confounds and illuminates: a case for darwinian tumor evolution. Nat Med, 20(4):344–346, 2014.
- [190] G Powathil, M Kohandel, M Milosevic, and S Sivaloganathan. Modeling the spatial distribution of chronic tumor hypoxia: implications for experimental and clinical studies. Comput Math Meth Med, 2012, 2012.
- [191] GG Powathil, DJA Adamson, and MAJ Chaplain. Towards predicting the response of a solid tumour to chemotherapy and radiotherapy treatments: clinical insights from a computational model. PLoS Comput Biol, 9(7):e1003120, 2013.
- [192] GG Powathil, KE Gordon, LA Hill, and MAJ Chaplain. Modelling the effects of cell-cycle heterogeneity on the response of a solid tumour to chemotherapy: biological insights from a hybrid multiscale cellular automaton model. J Theor Biol, 308:1–19, 2012.
- [193] ME Prince, R Sivanandan, A Kaczorowski, GT Wolf, MJ Kaplan, P Dalerba, Weissman IL, MF Clarke, and LE Ailles. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. Proc Natl Acad Sci USA, 104(3):973–978, 2007.
- [194] KA Rejniak and ARA Anderson. Hybrid models of tumor growth. Wiley Interdiscip Rev Syst Biol Med, 3(1):115–125, 2011.
- [195] T Reya, SJ Morrison, MF Clarke, and IL Weissman. Stem cells, cancer, and cancer stem cells. Nature, 414(6859), 2001.
- [196] L Ricci-Vitiani, DG Lombardi, E Pilozzi, M Biffoni, M Todaro, C Peschle, and R De Maria. Identification and expansion of human colon-cancer-initiating cells. Nature, 445(7123), 2007.
- [197] BD Ripley. The second-order analysis of stationary point processes. J Appl Prob, 13:255–266, 1976.
- [198] L Ritsma, SIJ Ellenbroek, A Zomer, HJ Snippert, FJ de Sauvage, BD Simons, H Clevers, and J van Rheenen. Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging. Nature, 507(7492):362–365, 2014.
- [199] R Rockne, JK Rockhill, M Mrugala, AM Spence, I Kalet, K Hendrickson, A Lai, T Cloughesy, EC Alvord, Jr, and KR Swanson. Predicting the efficacy of radiotherapy in individual glioblastoma patients in vivo: a mathematical modeling approach. Phys Med Biol, 55(12):3271–85, 2010.

- [200] IA Rodriguez-Brenes, NL Komarova, and D Wodarz. Evolutionary dynamics of feedback escape and the development of stem-cell-driven cancers. Proc Natl Acad Sci USA, 108(47):18983–18988, 2011.
- [201] Zoë N Rogers, Christopher D McFarland, Ian P Winters, Santiago Naranjo, Chen-Hua Chuang, Dmitri Petrov, and Monte M Winslow. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nature Methods, 2017.
- [202] DE Rosen. Vicariant patterns and historical explanation in biogeography. Syst Biol, 27(2):159–188, 1978.
- [203] MJ Sackin. “Good” and “bad” phenograms. Syst Biol, 21(2):225–226, 1972.
- [204] A Saraste and K Pulkki. Morphologic and biochemical hallmarks of apoptosis. Cardiovasc Res, 45(3):528–537, 2000.
- [205] AG Schepers, HJ Snippert, DE Stange, M van den Born, JH van Es, M van de Wetering, and H Clevers. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science, 337(6095):730–5, 2012.
- [206] Jacob Scott, Peter Kuhn, and Alexander RA Anderson. Unifying metastasis—integrating intravasation, circulation and end-organ colonization. Nature Reviews Cancer, 12(7):445–446, 2012.
- [207] Jacob Scott and Andriy Marusyk. Somatic clonal evolution: A selection-centric perspective. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 1867(2):139–150, 2017.
- [208] Jacob G Scott, David Basanta, Alexander RA Anderson, and Philip Gerlee. A mathematical model of tumour self-seeding reveals secondary metastatic deposits as drivers of primary tumour growth. Journal of The Royal Society Interface, 10(82):20130011, 2013.
- [209] Jacob G Scott, Alexander G Fletcher, Philip K Maini, Alexander RA Anderson, and Philip Gerlee. A filter-flow perspective of haematogenous metastasis offers a non-genetic paradigm for personalised cancer therapy. European journal of cancer, 50(17):3068–3075, 2014.
- [210] Jacob G Scott, Philip Gerlee, David Basanta, Alexander G Fletcher, Philip K Maini, and Alexander RA Anderson. Mathematical modeling of the metastatic process. In Experimental Metastasis: Modeling and Analysis, pages 189–208. Springer, 2013.
- [211] JG Scott, A Berglund, MJ Schell, I Mihaylov, William J Fulp, B Yue, E Welsh, JJ Caudell, K Ahmed, TS Strom, E Mellon, P Venkat, P Johnstone, J Foekens, J Lee, E Moros, WS Dalton, SA Eschrich, H McLeod, LB Harrison, and J Torres-Roca. A genome-based model for adjusting radiotherapy dose (gard): a retrospective, cohort-based study. Lancet Oncol, pages 30648–9, 2016.

- [212] JG Scott, AG Fletcher, ARA Anderson, and PK Maini. Spatial metrics of tumour vascular organisation predict radiation efficacy in a computational model. PLoS Comput Biol, 12(1):e1004712, 2016.
- [213] JG Scott, AB Hjelmeland, P Chinnaiyan, ARA Anderson, and D Basanta. Microenvironmental variables must influence intrinsic phenotypic parameters of cancer stem cells to affect tumourigenicity. PLoS Comput Biol, 10(1):e1003433, 2014.
- [214] CE Shannon. A mathematical theory of communication. ACM SIGMOBILE Mob Comput and Comm Rev, 5(1):3–55, 2001.
- [215] K Shao and RR Sokal. Tree balance. Syst Biol, 39(3):266–276, 1990.
- [216] SK Singh, C Hawkins, ID Clarke, JA Squire, J Bayani, T Hide, RM Henkelman, MD Cusimano, and PB Dirks. Identification of human brain tumour initiating cells. Nature, 432(7015):396–401, 2004.
- [217] K Smallbone, DJ Gavaghan, RA Gatenby, and PK Maini. The role of acidity in solid tumour growth and invasion. J Theor Biol, 235(4):476–484, 2005.
- [218] J Maynard Smith. The theory of games and the evolution of animal conflicts. Journal of theoretical biology, 47(1):209–221, 1974.
- [219] Hugo J Snippert, Laurens G Van Der Flier, Toshiro Sato, Johan H Van Es, Maaïke Van Den Born, Carla Kroon-Veenboer, Nick Barker, Allon M Klein, Jacco Van Rheenen, Benjamin D Simons, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing *lgr5* stem cells. Cell, 143(1):134–144, 2010.
- [220] JA Somarelli, KE Ware, R Kostadinov, JM Robinson, H Amri, M Abu-Asab, N Fourie, R Diogo, D Swofford, and JP Townsend. Phylooncology: Understanding cancer through phylogenetic analysis. Biochim Biophys Acta, 2016.
- [221] A Sottoriva, H Kang, Z Ma, TA Graham, MP Salomon, J Zhao, P Marjoram, K Siegmund, MF Press, D Shibata, and C Curtis. A big bang model of human colorectal tumor growth. Nat Genet, 47(3):209–216, 2015.
- [222] A Sottoriva, I Spiteri, SGM Piccirillo, A Touloumis, V Collins, JC Marioni, C Curtis, C Watts, and S Tavaré. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci USA, 110(10):4009–14, 2013.
- [223] A Sottoriva, Joost J C Verhoeff, Tijana Borovski, Shannon K McWeeney, Lev Naumov, Jan Paul Medema, Peter M A Slood, and Louis Vermeulen. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. Cancer Res, 70(1):46–56, 2010.
- [224] L Sottoriva, Aand Vermeulen and S Tavaré. Modeling evolutionary dynamics of epigenetic mutations in hierarchically organized tumors. PLoS Comput Biol, 7(5):e1001132, 2011.

- [225] Å Søvik, E Malinen, ØS Bruland, SM Bentzen, and DR Olsen. Optimization of tumour control probability in hypoxic tumours by radiation dose redistribution: a modelling study. Phys Med Biol, 52(2):499, 2007.
- [226] Å Søvik, E Malinen, and DR Olsen. Strategies for biologic image-guided dose escalation: a review. Int J Rad Oncol Biol Phys, 73(3):650–658, 2009.
- [227] Å Søvik, E Malinen, HK Skogmo, SM Bentzen, ØS Bruland, and DR Olsen. Radiotherapy adapted to spatial and temporal variability in tumor hypoxia. Int J Rad Biol Phys, 68(5):1496–1504, 2007.
- [228] K Sprouffske, Aktipis AC, JP Radich, M Carroll, AM Nedelcu, and CC Maley. An evolutionary explanation for the presence of cancer nonstem cells in neoplasms. Evol Appl, 6(1):92–101, 2013.
- [229] IJ Stamper, HM Byrne, MR Owen, and PK Maini. Modelling the role of angiogenesis and vasculogenesis in solid tumour growth. Bull Math Biol, 69(8):2737–2772, 2007.
- [230] A Stephanou, SR McDougall, ARA Anderson, and MAJ Chaplain. Mathematical modelling of flow in 2d and 3d vascular networks: applications to anti-angiogenic and chemotherapeutic drug strategies. Math Comput Model, 41(10):1137–1156, 2005.
- [231] CL Stokes and DA Lauffenburger. Analysis of the roles of microvessel endothelial cell random motility and chemotaxis in angiogenesis. J Theor Biol, 152(3):377–403, 1991.
- [232] KR Swanson, RC Rockne, J Claridge, MA Chaplain, C Alvord, and ARA Anderson. Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology. Cancer Res, 71(24):7366–75, 2011.
- [233] K Tamura, M Aoyagi, H Wakimoto, N Ando, T Nariai, M Yamamoto, and K Ohno. Accumulation of CD133-positive glioma cells after high-dose irradiation by Gamma Knife surgery plus external beam radiation. J Neurosurg, 113(2):310–318, 2010.
- [234] MJ Tindall, L Dyson, K Smallbone, and PK Maini. Modelling acidosis and the cell cycle in multicellular tumour spheroids. J Theor Biol, 298:107–115, 2012.
- [235] Leigh Van Valen. A new evolutionary law. Evol Theory, 1:1–30, 1973.
- [236] K Vermeulen, DR Van Bockstaele, and ZN Berneman. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell Prolif, 36(3):131–149, 2003.
- [237] L Vermeulen, F de Sousa e Melo, and DJ Richel. The developing cancer stem-cell model: clinical challenges and opportunities. Lancet, 2012.

- [238] L Vermeulen, FEM DeSousa, M vander Heijden, K Cameron, JH de Jong, T Borovski, JB Tuynman, M Todaro, C Merz, H Rodermond, MR Sprick, K Kemper, DJ Richel, G Stassi, and JP Medema. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol, 12(5):468–76, 2010.
- [239] E Vlashi, C Lagadec, L Vergnes, T Matsutani, K Masui, M Poulou, R Popescu, L Della Donna, P Evers, C Dekmezian, K Reue, H Christofk, PS Mischel, and F Pajonk. Metabolic state of glioma stem cells and nontumorigenic cells. Proc Natl Acad Sci USA, 108(38):16062–7, 2011.
- [240] O Warburg. On the origin of cancer cells. Science, 123(3191):309–314, 1956.
- [241] DJ Watts and SH Strogatz. Collective dynamics of ‘small-world’ networks. Nature, 393(6684):440–442, 1998.
- [242] N Weidner. Intratumor microvessel density as a prognostic factor in cancer. Am J Pathol, 147(1):9, 1995.
- [243] B Werner, D Dingli, T Lenaerts, JM Pacheco, and A Traulsen. Dynamics of mutant cells in hierarchical organized tissues. PLoS Comput Biol, 7(12):e1002290, 2011.
- [244] B Werner, JG Scott, A Sottoriva, ARA Anderson, A Traulsen, and PM Altrock. The cancer stem cell fraction in hierarchically organized tumors can be estimated using mathematical modeling and patient-specific treatment trajectories. Cancer Res, 76(7):1705–1713, 2016.
- [245] JW White, A Rassweiler, JF Samhuri, AC Stier, and C White. Ecologists should not use statistical significance tests to interpret simulation model results. Oikos, 123(4):385–388, 2014.
- [246] MJ Williams, B Werner, CP Barnes, TA Graham, and A Sottoriva. Identification of neutral tumor evolution across cancer types. Nat Genet, 2016.
- [247] MJ Williams, B Werner, C Curtis, C Barnes, A Sottoriva, and TA Graham. Quantification of subclonal selection in cancer from bulk sequencing data. bioRxiv, page 096305, 2016.
- [248] JW Wojtkowiak, KM Sane, M Kleinman, BF Sloane, JJ Reiners, and RR Mattingly. Aborted autophagy and nonapoptotic death induced by farnesyl transferase inhibitor and lovastatin. J Pharm Exp Ther, 337(1):65–74, 2011.
- [249] BG Wouters and JM Brown. Cells at intermediate oxygen levels can be more important than the “hypoxic fraction” in determining tumor response to fractionated radiotherapy. Rad Res, 147(5):541–550, 1997.
- [250] L Yadav, N Puri, V Rastogi, P Satpute, and V Sharma. Tumour angiogenesis and angiogenic inhibitors: a review. J Clin Diagn Res, 9(6):XE01, 2015.
- [251] TM Yeung, SC Gandhi, and WF Bodmer. Hypoxia and lineage specification of cell line-derived colorectal cancer stem cells. Proc Natl Acad Sci USA, 108(11):4382–4387, 2011.

- [252] GU Yule. A mathematical theory of evolution, based on the conclusions of Dr J.C. Willis, FRS. Phil Trans R Soc B, 213:21–87, 1925.
- [253] M Zaider and L Hanin. Tumor control probability in radiation treatment. Med Phys, 38(2):574–583, 2011.
- [254] L Zhang, DF Yankelevitz, CI Henschke, AP Reeves, MF Vazquez, and D Carter. Variation in vascular distribution in small lung cancers. Lung Cancer, 68(3):389–393, 2010.
- [255] Z Zhao, B Zhao, Y Bai, A Iamarino, SG Gaffney, J Schlessinger, RP Lifton, DL Rimm, and JP Townsend. Early and multiple origins of metastatic lineages within primary tumors. Proc Natl Acad Sci USA, 113(8):2140–2145, 2016.
- [256] J Žunić, K Hirota, and PL Rosin. A Hu moment invariant as a shape circularity measure. Pattern Recogn, 43(1):47–57, 2010.